EMBRY-RIDDLE Aeronautical University
SCHOLARLY COMMONS

Doctoral Dissertations and Master's Theses

Summer 2024

# Situation Awareness Assessment of Enhanced Stable Approach Flight Instrument Displays

David J. Hunter
*Embry-Riddle Aeronautical University*, hunted10@my.erau.edu

Follow this and additional works at: https://commons.erau.edu/edt

Part of the Aviation Safety and Security Commons

Scholarly Commons Citation

Hunter, David J., "Situation Awareness Assessment of Enhanced Stable Approach Flight Instrument Displays" (2024). *Doctoral Dissertations and Master's Theses*. 833.
https://commons.erau.edu/edt/833

This Dissertation - Open Access is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in Doctoral Dissertations and Master's Theses by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.

# Situation Awareness Assessment of Enhanced Stable Approach Flight Instrument Displays

David J. Hunter

Dissertation Submitted to the College of Aviation in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Aviation

Embry-Riddle Aeronautical University

Daytona Beach, Florida

June 2024

© 2024 David J. Hunter

All Rights Reserved.

**SITUATION AWARENESS ASSESSMENT OF ENHANCED STABLE APPROACH FLIGHT INSTRUMENT DISPLAYS**

By

David J. Hunter

This dissertation was prepared under the direction of the candidate's Dissertation Committee Chair, Andrew R. Dattel, and has been approved by the members of the dissertation committee. It was submitted to the College of Aviation and was accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Aviation.

Andy Dattel Digitally signed by Andy Dattel Date: 2024.06.09 01:21:11 -04'00'

Andrew R. Dattel, Ph.D.
Committee Chair

David Thirtyacre Digitally signed by David Thirtyacre Date: 2024.06.11 10:11:52 -07'00'

David L. Thirtyacre, Ph.D.
Committee Member

Steven Hampton Digitally signed by Steven Hampton Date: 2024.06.18 09:26:01 -04'00'

Steven Hampton, Ed.D.
Associate Dean, School of Graduate Studies, College of Aviation

John V. Sabel Digitally signed by John V. Sabel Date: 2024.06.11 13:38:09 -04'00'

John V. Sabel, J.D.
Committee Member

Alan J. Stolzer Digitally signed by Alan J. Stolzer Date: 2024.06.18 09:36:58 -04'00'

Alan J. Stolzer, Ph.D.
Dean, College of Aviation

Committee Member

Erik Baker Digitally signed by Erik Baker Date: 2024.06.12 09:38:58 -10'00'

Erik R. Baker, Ph.D.
Committee Member (External)

May 29, 2024

Signature Page Date

**Abstract**

Researcher: David J. Hunter

Title: Situation Awareness Assessment of Enhanced Stable Approach Flight Instrument Displays

Institution: Embry-Riddle Aeronautical University

Degree: Doctor of Philosophy in Aviation

Year: 2024

Through the year 2022, runway-related events, including excursions, abnormal runway contact, and runway undershoot and overshoot, were the third leading causes of fatal aircraft accidents. Principal to these runway-related accidents was the presence of an unstable approach. Pilots currently must depend on the timely recall of stable approach criteria stored in long-term memory, application of those criteria against current flight conditions, perception and cognition of deviation from those criteria, and determination of potential courses of action. Such recall is fallible, and challenges to perception and cognition are ever-present.

This study examined whether the presentation of stable approach criteria boundaries and alerting displays, in isolation and in combination, affected pilot situation awareness as to the presence of an unstable approach. Six hypotheses were presented to evaluate the influence of the two display enhancements on participant response accuracy to situation awareness queries. The data were obtained through experimentation using volunteer participants drawn from university flight training programs, FAA-approved training centers, corporate flight operations, and air carriers.

A 2x2 factorial design was used, with two levels of boundaries, absent or present, and two levels of an alert message, absent or present. This yielded one baseline and three enhanced treatments. A series of 12 video vignettes presented the participant unstable events in the context of completing an Instrument Landing System approach. Data on the dependent variable were gathered using recorded think-aloud protocols to measure response accuracy to unstable conditions. A two-way, within-group, repeated-measures Analysis of Variance was used to test whether the situation awareness of a pilot differed based on specific display treatments applied. The results indicated a statistically significant difference for the main effects of boundaries and the alert. However, statistical significance was not present for the interaction between the two treatments. Missed events, false alarms, and experience level of the participant were also examined.

The study results suggested benefit is derived from providing pilots display enhancements that highlight the presence of an unstable approach condition. These enhancements are worthy of consideration for incorporation into existing primary flight displays currently used in the aviation community. Further research benefit would be derived from examining unstable approach situation awareness when tested within flight simulation devices of greater fidelity and under unusual environmental- and system-related scenarios.

*Keywords:* unstable approach, human factors, perception, cognition, pilot, flight instruments, flight displays, flight training, flight operations.

## Dedication

This dissertation is dedicated to my wonderful wife of many years, Pamela, and to my cherished daughters, Shea and Kara, without whom I would not have sustained the drive to complete this arduous journey. Accomplishment of a Ph.D. program at a point in life when others are making plans for retirement and the concomitant carefree lifestyle was, by all measures, completely illogical. However, doing so fulfilled a long-held dream that was often, and rightfully, put by the wayside for other family matters.

**Acknowledgements**

I wish to first extend credit to three individuals who convinced me that pursuit of a Ph.D. at this late stage in my professional career was indeed a worthy endeavor: Mr. Jason Bialek, J.D; Dr. Barbara Holder; and Raymond R. Roberts (Captain, U.S. Navy, Retired). Mr. Bialek was so kind as to provide a letter of recommendation during the application process and was quick to return the same persistent and uplifting "you can do it" exhortations I bestowed upon him during his legal studies. Dr. Holder, a colleague of high regard in both academia and industry in the field of human factors, also argued on my behalf for selection into the program. Always supportive, she consistently provided an excellent role model as to living up to the technical expertise one assumes of the holder of a doctorate in the field. Her reputation as a flight deck design and human factors researcher at The Boeing Company remains legendary, and it was my privilege to assist her in a number of her studies. And finally, my supervisor and brother-in-arms, Ray Roberts, a Gulf War combat veteran and fighter pilot who understood the magnitude of this endeavor and encouraged me daily to ensure the company tuition funds being used to finance this degree did not go wasted – he pushed me to seek excellence in every paper and examination I faced.

It would be wholly improper not to mention the Embry-Riddle Aeronautical University Ph.D. program itself and provide acknowledgement of a small number of folks who played significantly in my completion of the program. First, a word of thanks to the founders of the Ph.D. in Aviation program, who foresaw both the academic content and delivery method that would allow someone as myself to attend such a program and flourish. Second, my heartfelt appreciation to Dr. Steven Hampton and Dr. Dothang

Truong, who stood by me when a critical family health issue drove me into a reduced course load during the first two years; there were a number of times where completion was thought to be highly unlikely due to the workload at home, work, and school. Without this understanding, I would've become another statistic in the "did not finish" column. My deepest appreciation goes to my current Dissertation Chair, Dr. Andrew Dattel, who supplied his keen insights on the design of experiments when measuring situation awareness and helped formulate an executable methodology. I'd also like to thank my previous Dissertation Chair, Dr. Haydee Cuevas, who not only provided me the "Five Cs – Critic, Coach, Confidant, Cheerleader, and Champion" when needed, but taught two human factors and human-centered design courses that provided me the academic understanding that two test pilot schools of which I attended and then taught did not. And then I'd also like to acknowledge three members of my committee, Dr. David Thirtyacre, Dr. Erik Baker, and John Sabel, J.D., who provided considerable guidance and clarity of thought in helping to ensure the manuscript was of highest quality. I would also like to acknowledge Ms. Cathy Woodyard and Dr. David Hedges, members of Cohorts 7 and 8, who were willing to band together in a support group and provide ceaseless encouragement and, if needed, a properly placed push. And I cannot close without once again thanking the university, which provided the funding for the flight simulator used by Dr. Thirtyacre to undertake his Ph.D. and was subsequently used to develop the vignettes for this research.

Finally, and most importantly, I wish to thank my wife, Pamela, who provides me true purpose in life, and my daughters Shea and Kara, for whom I am always proud. Words cannot convey how important they are in my life.

**Table of Contents**

## List of Tables

## List of Figures

# Chapter I: Introduction

Airplane accidents occur for a number of reasons. One of these reasons is the pilot's failure to discontinue an unstable approach and attempt to land the airplane despite an undesired energy state. The use of instrumentation in developing pilot situation awareness (SA) has been demonstrated in both theory and application in the aviation field. The purpose of this study is to examine how pilot SA when engaged in an unstable approach can be affected by applying enhancements to the flight instrument displays. A series of hypotheses are presented addressing the enhancements individually and in combination. Study delimitations, limitations, and assumptions are addressed.

The mission of the Federal Aviation Administration (FAA) is "to provide the safest, most efficient aerospace system in the world" (FAA, 2018, p. 1). The FAA has enacted rules, promoted technologies, and shaped the national aerospace environment to facilitate this goal. However, commercial air travel still suffers from incidents and accidents that result in serious injury, loss of life, and considerable financial impact (Boeing, 2022). These incidents and accidents can occur during any segment of flight, but the final approach segment of flight is critical. While comprising only 3% of the estimated flight time during a 1.5-hour short domestic flight, and considerably less for intercontinental and international flights, the final approach segment accounts for 15% of commercial jet fatal accidents and 13% of the onboard fatalities (Boeing, 2022). This segment can impact the landing segment as well, which accounts for an additional 31% of fatal accidents and 7% of onboard fatalities. Flight path management, whether manually flown or accomplished through automation, is the responsibility of the pilot. Errors in path management can result from excess or insufficient airspeed, improper airplane configuration, inaccurate thrust management, inadequate course tracking, and other errors or lapses in pilot actions. These errors or lapses, when observed at specified points in the approach, indicate an unstable condition. The presence of poor SA can be instrumental in recognition of these errors and lapses.

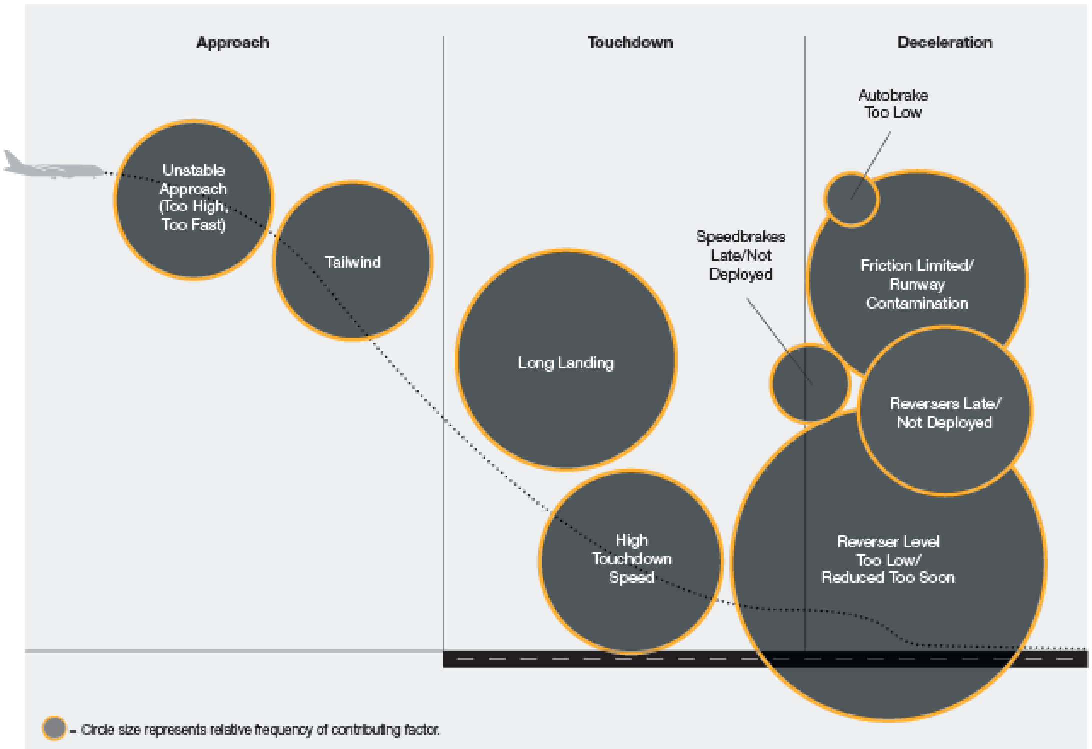
Errors in flight path management have resulted in airplane accidents. From 2013 through 2022, runway-related events, including excursions, abnormal runway contact, and runway undershoot and overshoot, were the third leading causes of fatal accidents (Boeing, 2022). By purpose, air carrier airplanes are operated from taxiway and runway surfaces designed to meet the needs of airplanes with maximum takeoff weights as high as 1.2 million pounds and wingspans in excess of 261 feet (ft), as in the case of the Airbus A380 (Airbus, 2016). Properly completed dispatch planning supports the airplane in arriving at the destination runway with a weight that meets the airplane's capability to stop within the available landing length. However, such planning does not, and cannot, address flight path management errors that invalidate planning assumptions or place the airplane on a trajectory not conducive to a safe landing.

The ensuing damage from failing to maintain the airplane within the longitudinal and lateral confines of the runway surface can be catastrophic, as evidenced by an airplane accident in Halifax/Stanfield International Airport, Nova Scotia, Canada. A Boeing B-747-400 cargo airplane operated by Skylease Cargo attempted to land on Runway 14 in night conditions, with a wet runway and gusty winds approaching 18 nautical miles per hour (kts). The airplane overran the runway, coming to rest approximately 695 ft off the end. In the process, the landing gear collapsed, two of the four engines separated from the airplane, and the remaining two engines were substantially damaged. Subsequent fire broke out, further damaging the airplane (TSB, 2018). This accident did not occur in isolation, nor did it bring the issue to a close. Runway excursions have continued to plague the community since, with recent examples including an Air Niugini B-737-800 flight from Phopei Harbor to Chuuk Island, Micronesia, in 2018 (AIC, 2018); a 2021 corporate Gulfstream G150 flight departing New Smyrna Beach Municipal Airport, Florida, and landing at Ridgeland, South Carolina (NTSB, 2023); and a Miami Air International B-737-800 operating from Guantánamo Naval Air Station, Cuba, to Jacksonville Naval Air Station, Florida, in 2019 (NTSB, 2021).

A number of these excursion events can be traced back to failure of the pilot to maintain stabilized approach criteria. Boeing (2012), in seeking to inform industry of measures being taken to improve training and flight deck displays to support safer flight operations in the terminal phase, published an article addressing the issues of runway overruns and seeking resolution to the problem, as shown in Figure 1.

**Figure 1**

*Identified Causes of Runway Surface Overruns*




*Note.* Excerpted from "Reducing Runway Landing Overruns," by Boeing Commercial Airplanes, 2012, *AERO*, *12*(3), p. 16. Copyright 2012 by The Boeing Company.

The graphic provides a simple but effective presentation of the relationship between various causal factors and presents them in terms of the geolocation of occurrence along the approach profile. The circle diameter serves to depict the relative frequency that each factor contributes to a runway overrun. Although it can be shown that runway overruns are frequently the result of a combination of these factors and others, focus is placed on the first factor, Unstable Approach (Too High, Too Fast), and leaves the other elements for further research. However, two additional identified factors, Long Landing and High Touchdown Speed, may be directly influenced by an unstable approach and therefore cannot stand in isolation (Boeing, 2012). Accidents occur for a number of reasons, one of which is the failure to discontinue an unstable approach. Although several definitions of a stabilized approach exist, current airplane performance instruments do not provide real-time display highlighting the presence of this condition to the pilot. Pilots are left to perceive and interpret the raw flight instrument (airspeed, vertical speed, course guidance) values and compare them against established stable approach criteria.

According to Wickens, “In most circumstances, a pilot’s task involves a continuous stream of activities” (2003b, p. 239). Pilots are reliant upon flight instrumentation for orientation when presented a lack of external visual references while operating under instrument meteorological conditions (IMC). Although the presentation of these displays varies between airplane manufacturers, certain elements remain consistent when following commonly-applied human factors principles and established regulatory guidance. Such display consistency reduces training requirements when pilots are transitioned between various airplane types, as a considerable amount of skill carry-over occurs in monitoring, perceiving, and interpreting displayed information. In addition, display consistency minimizes human factors safety risks evolving from confusion between current and previous model displays.

To support high levels of SA in complex operating systems, Endsley and Jones (2012) developed the SA-Oriented Design (SAOD) process, which consists of three building blocks to consider when developing display presentations. These three blocks include conducting an SA requirements analysis, the application of SA-oriented design principles, and the application of SA measurements. The implementation of a flight instrument display enhancement meant to improve SA can manifest in many ways.

Endsley and Jones recognize the nearly infinite permutations that can arise when applying the various human factors principles. They further note, "as new technologies develop, solid research on the best way to design their features to enhance SA and human performance will generally lag significantly" and "designers may often be surprised to find that certain design features do not work as well as anticipated. The objective evaluation of system design features, therefore, forms the third major building block of SA-Oriented Design" (p. 259).

The Human Factors Analysis and Classification System (HFACS), which evolved from theories on latent and active failures in nuclear safety generated by Reason (1990) and the subsequent aviation-centric research and refinement by Shappell and Wiegmann (2000), provides a taxonomy by which to assess any airplane incident or accident. Referencing this HFACS construct, the genesis of runway excursions could lie within a number of major causal categories and causal factors. Applicable levels of failure for consideration include preconditions for unsafe acts and unsafe acts of operators (Shappell & Wiegmann, 2000). The remaining levels, organizational influences and unsafe supervision, may certainly hold partial culpability in the systematic forensic analysis of these accidents, but such areas are beyond the scope of this research and are assumed to be external to the issue.

Thus, in defining an unstable approach experiment, a decision remains as to the degree to which the levels preconditions for unsafe acts (with its categories personnel factors, conditions of the operators, and environments factors) and unsafe acts of operators (with categories errors and violations) are relevant to the research. Ultimately, the examination of this aviation problem could take one of two paths: (a) either the issue with pilots continuing an unstable approach is rooted in SA insufficiencies resulting from a failure to perceive an unstable condition, comprehend its meaning, and/or predict the possible outcome of continuing the approach, or (b) pilots commit routine or exceptional violations in continuing an unstable approach in light of indications of such a state. It is assumed the pilot places emphasis on the importance of maintaining a stable approach, and companies have the appropriate policies and guidance mandating they do so, and seeks to investigate the impact of perceptual errors, skill-based errors, the technological environment, and physical/mental limitations as they impact SA.

**Statement of the Problem**

As of 2020, the National Aeronautics and Space Administration (NASA) Aviation Safety Reporting System (ASRS) contained 103 flight crew reports submitted since 2009 in which an unstable approach occurred and SA was coded as present (NASA, 2020). These reports reflected flights conducted under various regulations and across a wide range of airplane platforms. The linkage between the inadequate pilot SA and approach instability is intuitive. When flying an approach, pilots make use of flight instrumentation to identify proper airplane configuration, energy state, and flight path maintenance to the desired course.

The use of instrumentation in developing pilot SA has been demonstrated in both theory and application in the aviation field (Mitchell et al., 2009; Mosier, 2010; Whitlow & Dillard, 2019). Much of the historic research focus has been directed at novel methods of presenting airplane performance parameters and airplane system monitoring. Less prevalent is academic research as to the impact of enhanced flight instrument displays in improving pilot SA as to the presence of an unstable approach.

Pilots currently must depend on the timely recall of stable approach criteria stored in long-term memory, application of those criteria against current flight conditions, perception and cognition of deviation from those criteria, and determination of potential courses of action. Such recall is fallible (Wickens & Hollands, 2000). Enhanced flight instrument displays have the potential to assist pilots in assessing the current airplane performance state against proven stable approach criteria and serve as an effective means to support the aeronautical decision-making process as to whether to continue or abort an approach (Lee et al., 2017). Unfortunately, little research has been conducted to validate this potential.

**Purpose Statement**

The purpose of this study was to determine how pilot SA, when engaged in an unstable approach, is affected by applying enhancements to the flight instrument displays. It measured pilot SA while under a representative flight instrument display used on current air carrier airplane systems. The study applied the principles of user-centered design (Endsley & Jones, 2012) to identify the proper application of display presentation. The study then assessed the effectiveness of stabilized approach display treatments on the ability of the pilot to achieve an improved state of SA.

**Significance of the Study**

Human factors practitioners should view pilot SA through the lens of empirical research and engineering application. Research studies directed at the application of flight instrument display design enhancements in the context of specific flight tasks add to the existing body of knowledge. Findings derived from this study furthered an understanding of how employing display enhancements meant to alert pilots to the presence of an unstable approach condition impacted individual SA and, consequently, potentially colors the output of the aeronautical decision-making process. The conclusions serve as a basis to understanding whether such flight instrument enhancements have a positive, negative, or neutral impact on the various levels of pilot SA, and potentially suggest future research as to how these types of enhancement might improve operations and support advancements in flight training.

Analysis of the variation between the modes of criterion presentation to enhance pilot SA provided empirical research data and conclusions that should prove beneficial to the academic community, flight instrument display manufacturers, commercial air carrier operations groups, and pilots undergoing flight training in programs that make use of scenarios to introduce the potential for unstable approaches. Innovations in stabilized approach presentation for flight displays might evolve from data demonstrating that the presentation of such information aids in the development of pilot situation awareness. Additionally, information derived from pilot perceptions on usability might aid in identifying target populations supportive of the introduction of this technology. Study findings may provide a starting point for further research into the application of enhanced approach stability displays in the flight deck.

**Research Question and Hypotheses**

A single research question was posed in warranting a determination as to whether the introduction of the enhanced flight instrument displays pilot affected perception, cognition, and prediction of instability during a final approach to the runway. The research question focused on pilot SA when presented the possibility of an unstable approach:

*Is pilot situation awareness, when engaged in an unstable approach, affected by applying enhancements to the flight instrument displays?*

Pilots may experience an overall improvement in the identification of the existence of factors deemed indicative of an unstable approach when the flight instrument displays incorporate stable approach definition boundaries or warning messages. The proposed displays were expected to reduce the amount of stabilized approach information to be stored in long-term memory by using embedded logic to visually depict the parameters or an alert, thus, providing better SA to the pilot and supporting the proposition that pilot cognitive resources could be better allocated to other relevant tasks. A safety gain must be realized to justify the expense in development of such displays.

The hypotheses in the study were established based on a 2x2 within-group, repeated-measures design with two categorical independent variables (IVs) and a single discrete dependent variable (DV). The first IV manipulated the flight displays by employing stability criteria boundaries, and the second IV manipulated the flight displays by employing an alert message. The state of each of the IVs addressed the presence or lack of a specific display treatment, while the DV addressed the accuracy in responding to an SA assessment when presented an unstable condition. The assumption was that participants possessing higher SA would be able to respond more accurately. A set of hypotheses was posed. The response accuracy (RA)-based hypotheses were as follows:

- Pilots provided an enhanced flight display employing stability criteria boundaries exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented no enhanced flight displays.

$H_01_{RA}$: $\mu_{enhanced\ (boundaries)} = \mu_{unenhanced\ (baseline)}$

$H_A1_{RA}$: $\mu_{enhanced\ (boundaries)} \neq \mu_{unenhanced\ (baseline)}$

- Pilots provided an enhanced flight display employing an alert message exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented no enhanced flight displays.

$H_02_{RA}$: $\mu_{enhanced\ (alert)} = \mu_{unenhanced\ (baseline)}$

$H_A2_{RA}$: $\mu_{enhanced\ (alert)} \neq \mu_{unenhanced\ (baseline)}$

- Pilots provided an enhanced flight display employing stability criteria boundaries exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented an enhanced flight display employing an alert message.

$H_03_{RA}$: $\mu_{enhanced\ (boundaries)} = \mu_{enhanced\ (alert)}$

$H_A3_{RA}$: $\mu_{enhanced\ (boundaries)} \neq \mu_{enhanced\ (alert)}$

- Pilots provided an enhanced flight display employing both stability criteria boundaries and an alert message exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented no enhanced flight displays.

$H_04_{RA}$: $\mu_{enhanced\ (boundaries\ +\ alert)} = \mu_{unenhanced\ (baseline)}$

$H_A4_{RA}$: $\mu_{enhanced\ (boundaries\ +\ alert)} \neq \mu_{unenhanced\ (baseline)}$

- Pilots provided an enhanced flight display employing both stability criteria boundaries and an alert message exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented an enhanced flight display employing an alert message.

$$H_{0}5_{RA}: \mu_{\text{enhanced (boundaries + alert)}} = \mu_{\text{enhanced (alert)}}$$

$$H_{A}5_{RA}: \mu_{\text{enhanced (boundaries + alert)}} \neq \mu_{\text{enhanced (alert)}}$$

- Pilots provided an enhanced flight display employing both stability criteria boundaries and an alert message exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented an enhanced flight display employing stability criteria boundaries.

$$H_{0}6_{RA}: \mu_{\text{enhanced (boundaries + alert)}} = \mu_{\text{enhanced (boundaries)}}$$

$$H_{A}6_{RA}: \mu_{\text{enhanced (boundaries + alert)}} \neq \mu_{\text{enhanced (boundaries)}}$$

**Delimitations**

This study sought to assess the application of treatments to current displays. The study did not seek to examine the interaction of airplane control in conjunction with the task of display monitoring for unstable criteria. Such a scope extension to include functional fidelity would have introduced confounding variables such as the frequency and phasing of attention focus between the displays and the external world, as well as those emanating from differences in how each pilot would respond in terms of control inputs to a given perturbation. Pilots' skills might impact the ability to provide a consistent presentation of the test conditions. Further, access to full flight simulation or flight training devices capable of proper control response, and hence display drivers, could have been costly and restricted. Therefore, no active control of the airplane, or the representative display representing it, occurred. Presentations was limited to developed vignettes capable of presentation outside the training device system.

Even though modern air carrier and corporate airplanes operate with a crew complement of two or more pilots, the research examined the impact of proposed display enhancements on the SA of a single pilot. The presence of an additional crew member, while operationally more realistic, introduces a number of team interactions that can aid or distract from the ability of the individual participant to monitor the flight instrument display and identify an unstable condition. However, it is a primary role of the flying pilot to determine stability of approach.

The definition of stability in an approach context can vary, depending on the regulatory agency or professional body setting the criteria. As such, there is no definitive standard for what constitutes a stabilized approach. However, a number of commonly accepted criteria exist that prove useful in guiding the proposed display enhancements. In order to scope the study to a level that would support the presentation of information without overly-cluttering the display field of regard or introducing confounding variables unrelated to displays, the design focused on the parameters of airspeed, roll angle, vertical path, lateral path, and vertical speed. The study was not intended to examine the breadth of stable approach criteria found within the academic and operational communities, nor validate their appropriateness and efficacy. Hence, other areas such as adherence to power settings and configurations were not addressed. Thus, these parameters were set by the research instrument and briefed as such to the study's participants.

The study application was restricted to those pilots operating airplanes in either an air carrier, corporate, commercial, or university setting, where the likelihood of flying in weather conditions or environments conducive to approach instability was thought to be higher. Focus was placed on pilots holding FAA certificates who were based in the United States. The results from the study would not necessarily generalize to the broader population of pilots.

Pilots operate airplanes in the context of the fidelity of the experience. Noble (2002) provided a basic definition of fidelity that serves well as a starting point for discussion of the types of fidelity: fidelity, then, is "the degree to which a…simulated experience imitates the real world" (p. 33). Beyond the functional fidelity addressed earlier, physical and psychological fidelity were also to be considered (Hays, 1980). Participants were presented such fidelities but limited to the experience provided by the displays themselves. The degree of display realism and cognitive demand were limited by design to allow isolation to the effect being studied.

Study participants may be prone to adapt their normal flight instrument scan techniques and test responses to accommodate the nature of the study. As such, the data can be skewed toward a particular result as the participant provides responses. Any accommodation to the test design was assumed to be consistent among the study participants.

Finally, geolocation of the airplane during an approach can be conducive to improved SA. Issues with the software used to develop experiment vignettes prevented navigation database loading in the Flight Management System (FMS). This precluded selecting a particular approach and the subsequent presentation of the field elevation on the primary display and elevation and course routing on the Navigation Display (ND). This was considered to be a minimal impact to the study as expert reviews indicated reference to the ND was very limited during the approach, where the focus was primarily on course guidance, and would typically become relevant in missed approach scenarios where routing to the missed approach holding point was useful. Further, such presentation was considered specific to certain display designs and likely to go unnoticed by pilots not familiar with these design features.

**Limitations and Assumptions**

The level of fatigue and fitness to participate in the study were addressed as part of the pre-trial briefing. However, it was difficult to ascertain the current state of the participant. Commercial pilots conduct the instrument approach phase of flight at the completion of flight durations routinely in excess of 3-4 hours for domestic flights and 12 hours for international flights. Sometimes these occur following multiple legs of shorter distances. As such, fatigue can be a major factor in the ability of the pilot to remain actively engaged and alert. The study assumed that all participants met a representative state of rest and nourishment emulating a typical air carrier flight routine.

Pilots typically brief the elements of an instrument approach prior to its execution. If serving on a crewed airplane, the interaction between the two pilots establishes an understanding as to the actions to occur and the responsibility of each individual. Pilots routinely brief approaches and airfield diagrams to airports to which they are well accustomed. As such, they can be prone to cognitive complacency in over trusting the events will unfold similar to the last occurrence. Normally, complacency does not become an issue until something fails or the environment drastically changes, and pilots are forced into a high alert state (Wickens & Hollands, 2000). Consequently, pilots must actively engage in a higher state of alert during the final approach phase of flight. It was assumed pilots participating in the study were in an equally high state of alert owing to the experiment process.

As the study will be conducted in English, it was expected that all participants would have an adequate mastery of the language, sufficient to understand directions given, complete pre- and post-trial survey assessments, and communicate verbally with the observer. No evaluation of English language skills was necessary as, though not the first language of some FAA-certified pilots in the study, English is the official language of the global aviation community, and all pilots must meet an International Civil Aviation Organization Language Proficiency Operational Level 4 standard (FAA, 2017d).

The level of experience in the use of advanced flight instrument displays may have an effect on the results of the study. Further, those with more flight time may possess a more highly-developed scan pattern for obtaining critical flight condition parameters, which could bias the data. Design of the research experiment protocol addressed some of the experiential differences between participants. It was assumed that pilots entering would have a minimum level of familiarity with current flight instrument displays, as evidenced by holding an instrument rating.

The lack of motion-related cues to the participant's proprioceptive and kinesthetic, and equilibrioceptic senses limits the ability to realistically present the operational environment experience. Participants were unable to use such cues to foreshadow changing airplane conditions in roll, pitch, and yaw. The degree to which such cues impact a pilot’s particular visual dwell or saccade pattern during an instrument crosscheck was unknown.

Design of the test protocols sought to isolate the participants from influences outside the focus of the study as ambient noises, distractions, and varying light conditions may subject them to visual and aural distractions. Efforts to minimize this effect were addressed. Any residual external influences were expected to be of minimal impact.

## Summary

Unstable approaches have been demonstrated to be a contributing, if not causal, factor in runway excursions. Even as many commercial flight operations and air transport manufacturers appear to have adopted some level of stable approach criteria within their Standard Operating Procedures (SOPs), such events continue to occur. Although the criteria proposed by the Flight Safety Foundation (FSF) in 2010 appear to have served the industry, such stringent criteria may be excessive given the number of unstable approaches that are successfully completed. However, little is known as to the SA levels of pilots when such conditions are encountered. Perhaps, with sufficient levels of perception, cognition, and prediction, these approaches may not have been continued.

Thus, this study examined whether the presentation of select stable approach boundaries and alerts, either in isolation or in combination, enhanced pilot SA in terms of the presentation of an unstable condition. Such SA can be complicated by internal and external influences, to include non-normal airplane configurations, challenging turbulent weather conditions, high-altitude and hot temperature conditions, pilot fatigue, and others. Boundaries and alerts may provide enhanced levels of pilot SA on approach stability and assist pilots in making appropriate decisions to either continue the approach or vacate it and execute a go-around procedure.

**Definitions of Terms**

| | |
|---|---|
| Above Ground Level | The height of the airplane above the plane of the earth immediately below the current or, in some cases, projected position of the airplane. |
| Angle of Attack | The measure of the angle between the direction of air flowing toward the airplane wing, known as the relative wind, and the chord of the wing's surface, as measured by a line from the leading edge to the trailing edge of the wing (Stinton, 1987; Wickens, 2003). |
| Autokinesis | The phenomena in which a fixed, distant, light source appears to wander about the field of view, specifically when observed in dark conditions against an otherwise empty background (Young, 2003). |
| Category I | An ILS approach to the runway in which the weather ceiling is no lower than 200 ft above the runway threshold and the visibility not less than 1,800 ft viewed horizontally (FAA, 2002). |
| Ceiling | The height above the ground of the lowest layer of clouds or obscuring weather that is reported |

| | |
|---|---|
| | as being broken, overcast, or obscured (FAA, 2002). |
| Control Display Unit | A data entry and readout device that interacts with the Flight Management System to provide for the entry and extraction of information pertaining to the route of flight. |
| Equilibrioceptic | The sensation of balance and equilibrium with respect to gravitational forces. |
| Final Approach Fix | The fix from which the final approach to an airport is extended to the runway, identifying the final approach segment of the approach (FAA, 2020). |
| Flight Deck | Otherwise referred to as the cockpit or flight crew station, the location in the airplane in which the pilot is situated when conducting assigned tasks. |
| Flight Director | A command display that informs the pilot as to the direction and magnitude of the lateral and vertical corrections necessary to reestablish the desired flight path (Wickens, 2003a). |
| Flight Management System | A computer system, containing a large navigation database, that allows the pilot to program navigational routes of flight that |

| | |
|---|---|
| | provide real-time guidance to the pilot through primary flight and navigation displays (FAA, 2020). |
| General Aviation | The portion of the civil aviation community that does not include scheduled (air carrier) or unscheduled (on-demand) operations (FAA, 2020). |
| Haptic | The sensation afforded by the combination of both tactile and kinesthetic senses providing information to the brain (Aukstakalnis, 2017). |
| High Definition | Of or invoking a high degree of detail in the imagery presented the viewer; by accepted definition, having a resolution of at least 1280x720 pixels. |
| Inceptor | In this context, a manipulator within the flight deck that affects a change in the flight control positioning (Hess, 2003). |
| Kinesthetic | Of, or having to do with, the brain's awareness of the sensation of motion as derived by receptors in the joints and muscles (Wickens et al, 2004). |
| Navigation Display | Otherwise known as an electronic map, an integrated, dynamic presentation of the two- |

| | |
|---|---|
| | dimension representation of the airplane lateral and along-track location relative to established locations, or waypoints (Wickens, 2003). |
| Oculogravic Illusion | The visual illusion that objects in the visual field are moving relative to subject, when linear accelerations or decelerations are encountered (Young, 2003). |
| Primary Flight Display | An electronic display providing the pilot an egocentric view of the outside world, through which attitude, airspeed, altitude, and navigation system path guidance information are displayed. |
| Proprioceptive | Of, or having to do with, the human brain's awareness of the positioning of the body and associated appendages within a space; a representation of both joint angles and muscle contractions (Wickens et al., 2004). |
| Somatogravic Illusion | The illusion of the sensation of tilt afforded by the vestibular system when the body is subjected to high accelerations or decelerations (Young, 2003). |
| Stabilized Approach | An approach in which all the criteria in company standard operating procedures are met |

| | |
|---|---|
| | before or when reaching the applicable minimum stabilization height (FSF, 2010). |
| Standard Operating Procedures | Written guidelines as to the operational practices for an airplane, meant to ensure standardization between pilots, improve training outcomes, and minimize safety risks. |
| Subject Matter Expert | In this context, an individual holding expertise in the concepts and practices associated with unstable approaches; as such, they prove to be an invaluable resource for determining face validity of the modeled vignettes. |
| Tactile | The sensation of pressure against a touch receptor (Wickens et al., 2004); the detection and perception of external pressures, vibration, flutter, and other felt conditions (Aukstakalnis, 2017). |
| Technically Advanced Airplane | An airplane equipped with an electronic Primary Flight Display, multifunction Navigation Display, and two axis autopilot, as defined by the established regulatory guidance (FAA, 2021). |
| Tracking Task | The work assigned to the pilot requiring application of inceptor inputs, as required, to |

| | |
|---|---|
| | track a designated target; usually accomplished using a closed-loop, negative feedback design (Wickens, 2003b). |
| Velocity Reference | In this context, the reference landing speed ($V_{REF}$) as established by flight manual or onboard systems for the current airplane weight, configuration, and landing conditions (FAA, 2020; FAA, 2017). |
| Vertical Situation Display | A display presentation in which the vertical path of the airplane is presented using an exocentric profile view, often used to provide a third dimension to the navigation presentation (Wickens, 2003). |

**List of Acronyms**

| | |
|---|---|
| AC | Advisory Circular |
| AGL | Above Ground Level |
| ALAR | Approach and Landing Accident Reduction |
| ANOVA | Analysis of Variance |
| ASRS | Aviation Safety Reporting System |
| ATC | Air Traffic Control |
| ATP | Airline Transport Pilot |
| CAT | Category |

| | |
|---|---|
| CDU | Control Display Unit |
| CFR | Code of Federal Regulations |
| DV | Dependent Variable |
| EICAS | Engine Indication and Crew Alerting System |
| EFIS | Electronic Flight Instrumentation System |
| ERAU | Embry-Riddle Aeronautical University |
| FAA | Federal Aviation Administration |
| FAF | Final Approach Fix |
| FAR | Federal Aviation Regulation |
| FD | Flight Director |
| FMS | Flight Management System |
| FOQA | Flight Operations Quality Assurance |
| FPM | Feet Per Minute |
| FSF | Flight Safety Foundation |
| FSTD | Flight Simulation Training Device |
| FT | Feet |
| GA | General Aviation |
| GDTA | Goal-Directed Task Analysis |
| GPWS | Ground Proximity Warning System |
| G/S | Glideslope |
| HAT | Height Above Touchdown |
| HCI | Human-Computer Interaction |
| HD | High Definition |

| | |
|---|---|
| HFACS | Human Factors Analysis and Classification System |
| HRS | Hours |
| HSI | Horizontal Situation Indicator |
| HZ | Hertz |
| IATA | International Air Transport Association |
| IC | Informed Consent |
| ILS | Instrument Landing System |
| IMC | Instrument Meteorological Conditions |
| IQR | Interquartile Range |
| IRB | Institutional Review Board |
| IV | Independent Variable |
| KT | Knots |
| LOC | Localizer |
| LPV | Localizer Performance with Vertical Guidance |
| MAR | Missing at Random |
| MCAR | Missing Completely at Random |
| MFD | Multi-function Display |
| MIN | Minute |
| NASA | National Aeronautics and Space Administration |
| ND | Navigation Display |
| NM | Nautical Mile |
| NTSB | National Transportation Safety Board |
| PF | Pilot Flying |

| | |
|---|---|
| PFD | Primary Flight Display |
| PM | Pilot Monitoring |
| PTS | Practical Test Standards |
| RA | Radio Altitude |
| RA | Response Accuracy |
| RNAV | Area Navigation |
| RNP | Required Navigation Performance |
| S | Second |
| SA | Situation Awareness |
| SAOD | Situation Awareness-Oriented Design |
| SME | Subject Matter Expert |
| SOP | Standard Operating Procedures |
| SPAM | Situation Present Assessment Method |
| TAA | Technically Advanced Airplane |
| TSB | Transportation Safety Board of Canada |
| UK CAA | United Kingdom Civil Aviation Authority |
| VMC | Visual Meteorological Conditions |
| VOR | Very High Frequency Omni-directional Range |
| $V_{REF}$ | Velocity Reference |
| VSD | Vertical Situation Display |

# Chapter II: Review of the Relevant Literature

Failure to execute a stabilized Instrument Landing System (ILS) approach impacts the ability of the pilot to land and decelerate an airplane within the confines of the runway (Boeing, 2022). The resource demands placed upon a pilot executing a stabilized ILS approach can vary with internal and external influences. Curtis et al. (2010) recounted,

> Most of the tasks involved in aviation are contingent on the ability to attend to multiple sources of information efficiently…balancing between tasks that require focus on specific flight critical information to complete a task and monitoring multiple different sources of information of varying relation. (pp. 443-444)

Low visibility at the decision height, the presence of moderate to severe turbulence, diminished ambient lighting conditions, concurrent systems non-normal events, ineffective crew communication, physiological stressors, and reduced pilot SA can impact overall task performance. Although external influences can be significant, their contribution were deemed to be outside the scope of the study. The literature review focused on internal factors and sought understanding of what constitutes a stabilized approach, the contributions of SA and human-centered design, and how long-term memory and working memory affect pilot resource demands.

## The Stabilized Approach

In recognition of the importance of a stabilized approach, in 2007 the FAA published Advisory Circular (AC) 91-79A, *Mitigating the Risks of a Runway Overrun upon Landing*, which emphasized this critical phase of flight (FAA, 2014b). The purpose of the AC was to identify contributory factors common to most runway excursion accidents. In doing so, it addressed hazards associated with runway excursions, proposed risk mitigation strategies, and provided supporting information to develop prevention training. One of the hazards identified was the presence of an unstable approach, which the FAA identified as the "first line of defense in preventing an overrun" (2014b. p. 3). These unstable approach factors included landing configuration, stabilized on profile, descent rate, indicated airspeed, and engine speed (FAA, 2014b). This document, though useful as a foundation for understanding and applying the stable approach concept, failed to address such aspects as the complexities of higher performance aircraft, how to address unique instrument approach designs, and differentiation between approaches where the airport could be seen from those in which it could not. Further, ACs are not binding on the public, unless incorporated into a regulation by means of reference, leaving compliance to the individual (Adamski & Doyle, 2005). Lacking regulatory direction, manufacturers and operators developed a number of well-defined performance measures calculated to minimize the risk of runway excursions. These measures were meant to be monitored and verified by the pilot during final approach using the information provided on flight instrumentation.

Recent innovations in airplane data collection and data mining capabilities support the analysis of a considerable amount of flight condition data. Shelby et al. (2013) noted that enhancements to airplane data collections systems, once limited to basic items such as airspeed, altitude, heading, and vertical acceleration rate, have expanded. Such systems now provide other details, such as tolerances for configuration limitations and compliance with operator-defined stabilized approach criteria (Shelby et al., 2013). Many air carriers have emplaced safety programs that monitor and report unstable approaches, but these reports are not presented to the pilot. As a result, considerable historical data exist to define what constitutes an unstable approach and identify when failure to adhere to stable approach criteria has been an accident factor. However, to understand the impact of a crew failing to stabilize the approach, a foundational understanding of the term *stable approach* must be established.

***Foundational Understanding***

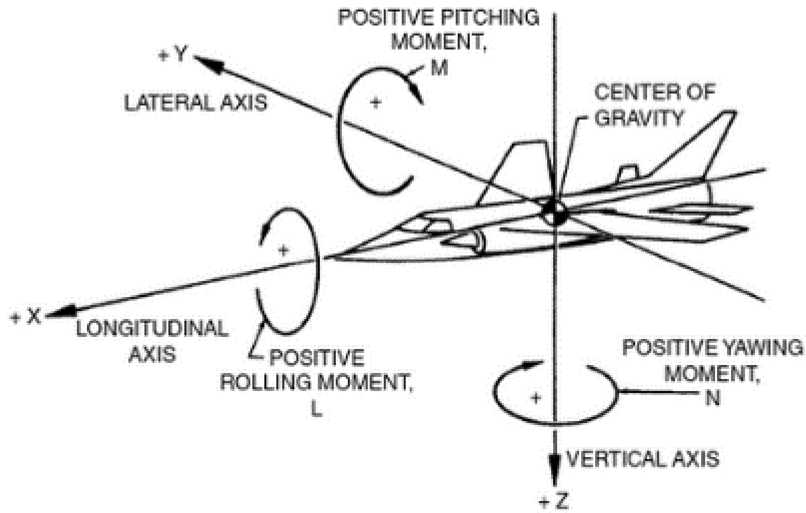
Pilots know and recognize the final approach segment of their journey: the airplane has been configured, the landing checklist completed, the pitch of the flight deck lowered, the engines settings adjusted to control airspeed in a higher drag configuration, the presence of a visual flow field in clear weather suggesting downward movement, and perhaps a sensation of increased minute airplane movement. The pilot, flying the airplane manually or through the automated flight control system, the autopilot, makes subtle path corrections to comply with the commands of the guidance system being used. Very few actions occur with urgency, as proper planning, accommodating flight conditions, and timely performance monitoring and feedback make for few, if any, surprises. Though such an experience is the goal, the reality is that each year a number of approaches prove an exception to this desire.

The stabilized approach concept, then, centers on the premise an airplane should not experience undesired and unplanned perturbations during one of the most critical phases of the flight profile, the final approach segment. These perturbations, when they occur, can be categorized into one of three major areas: those that affect the airplane's flight path, those that impact the airplane's energy, and those that encumber the pilot at an inopportune moment.

Flight path perturbation may occur due to changes in the three airplane body axes of flight – roll, pitch, and yaw – and can result from a number of sources. The three axes are presented in Figure 2. These axes, labeled X, Y, and Z, are mutually perpendicular, extend from the airplane's center of gravity, and are oriented geocentrically (Dole et al., 2017). Deviation from the axes is measured in units of degrees of angle from a defined datum, usually established as the fuselage reference line extending through the length of the airplane (Kumar et al., 2005). As the axes are reference to the ground plane, a nominal 3.00° approach path to an airport runway, as measured upward from the navigation aide transmitter and projecting along an extended runway centerline, would intersect the axes with a 3.00° depression relative to the axis. Positive values for displacement and moments are defined by the use of the + convention. In modern airplane designs, the wing and empennage configuration are such that the airplane will remain stable throughout nominal flight conditions. To deviate from that state, internal or external influences must be applied, be they intentional or unintentional.

**Figure 2**

*Body Axis of Flight*




*Note.* Excerpted from *Flight Theory and Aerodynamics: A Practical Guide for Operational Safety* (3rd ed.) (p. 250) by C. E. Dole, J. E. Lewis, J. R. Badick, and B. A. Johnson, 2017, John Wiley & Sons. Copyright 2017 by John Wiley & Sons.

Flight control inceptors in the form of a control yoke, or stick, and rudder pedals are manipulated to adjust orientation to the axes (Hess, 2003). The airplane is intentionally pitched up and down about the airplane's lateral axis, Y, using control surfaces mounted on the horizontal stabilizer at the empennage of the airplane: elevators, stabilators, or servo tabs (Stinton, 1987). When lateral control is desired, the airplane is rotated about the longitudinal axis, X, to establish a desired roll angle. This is accomplished through the use of ailerons, spoilers, or a combination thereof, positioned aft of the leading edge of the wing. Differential stabilizer systems, allowing each aft horizontal stabilizer to move independently, may also be used for lateral control (Stinton, 1987). Finally, directional control of the airplane, as manifest in yawing motion about the

Z axis, is accomplished through the use of rudder surfaces mounted to the vertical stabilizer or the application of a “flying” vertical stabilizer capable of pivoting about the vertical axis (Stinton, 1987). Pilots flying in a dynamic environment require greater inceptor activity. And though each of these control mechanisms are addressed individually, the reality is potential cross-coupling between the control surface effects (Wickens, 2003b). Pilots are trained to apply appropriate control inputs to correct the observed cross-coupling effect, and such manipulation was beyond the study scope.

The airplane may be intentionally or unintentionally perturbed by internal influences, and pilots themselves may be an internal influence on approach stability. For example, deviation of the airplane from the desired path may result from over control resulting from a lack of currency in manual flying skills, whereas not so much with overall flight experience (Ebattson, 2009); equally, an overreliance on the use of the autopilot may reduce manual flight skills (Haslbeck et al., 2014). Over control may also result from age-related factors (Kennedy et al., 2010). Even the manner in which a pilot holds the flight control inceptor may have an input on the number and magnitude of inputs (Haslbeck et al., 2012). Visual illusions may also be a factor, where the compelling presentation afforded the pilot when first exposed to visual meteorological conditions (VMC) during the approach may cause the pilot to provide errant control inputs to comply with an incorrect perception, and attention switching or attention tunneling on the external visual presentation may momentarily impact compliance with the commanded guidance. Wilson and Binnema (2014) provided background on such illusions, to include sloping runways and surrounding terrain, runway relative size cues, runway light relative brightness cues, the duck-under phenomenon, black hole approach conditions, autokinesis, and other ambiguous visual stimuli. Young (2003) highlighted the impact of oculogravic and somatogravic illusions, as well as vection, on spatial orientation, which can influence the ability of the pilot to correlate perceptions to the reality of the airplane condition.

The airplane may also be unintentionally perturbed from its stable state through external influences. The International Air Transport Association (IATA) suggested environmental threats such as gusty winds associated with mechanical or thermal low-level turbulence, turbulence associated with thunderstorms, or horizontal or vertical windshear may destabilize an approach (IATA, 2016). Abnormal airplane states and flight control anomalies may also result in approach stability perturbations; examples include open access doors, asymmetric airplane flap configurations, imbalanced thrust on multi-engine airplanes following an engine failure, or significantly reduced engine output. Equally, air traffic control (ATC) may hold the airplane at an unusually high altitude or direct the airplane to an excessive intercept angle for the final approach, exacerbating the energy problem facing the pilot. Further, insufficient spacing between airplanes might lead to an unexpected wake turbulence exposure, potentially resulting in an induced rolling moment. Such an encounter can generate a startle response in the pilot driving delayed responsiveness, as well as present a dynamic condition that can exceed the roll capability of the airplane (FAA, 2014b).

Energy-related perturbations to a stabilized approach would involve inaccurate thrust settings or rates of descent for the given approach angle to the runway. Such variance from the desired state might include airspeed that is insufficient, leading to a potential airplane stall or upset condition, as well as landing short of the runway surface or contacting the surface at an excessive rate of descent. Equally, excessive airspeed may lead to difficulty in maintaining the approach path due to increased lift, extended landing flare distances that may lead to runway excursions, and potential over speed conditions for the current flap position.

Finally, perturbations that encumber the pilot at an inopportune moment would include failure to complete a required checklist, setting an inappropriate airplane configuration for landing, or perhaps excessive or inappropriate communications during the approach. Ross (2018) identified a number of human factors-related contributors to unstable approaches: communications breakdowns, distractions, SA, training and qualifications, and pilot workload being the more significant contributors, by frequency of occurrence. Though each of these items has a detrimental effect on the attention of the pilot and can interfere with the ability to manually control the airplane to the desired approach path or monitor and manage the auto flight system as it controls the flight path, it is significant that the modal factor, SA, contributed to unstable approaches in 77.9% of 95 cases examined (Ross, 2018). With an understanding of the fundaments, an examination of accepted and proposed stabilized approach concepts ensues.

***Early Concepts***

As a leading aviation industry safety organization, IATA proposed a simple definition of an unstable approach. Citing the Accident Classification Task Force, an investigative body of IATA, an unstable approach is allocated to an accident when, "it 'has knowledge about vertical, lateral or speed deviations in the portion of the flight close to landing'" (IATA, 2016, p. 4). The FSF commissioned the Approach and Landing Accident Reduction (ALAR) Task Force to study and address issues with runway overruns. The results manifested in the ALAR Toolkit, a collection of guidelines and techniques to reduce the likelihood of an overrun. The FSF ALAR Briefing Note 7.1 (2010) provided a broad-brush definition of a stabilized approach: “An approach is stabilized only if all the criteria in company standard operating procedures (SOPs) are met before or when reaching the applicable minimum stabilization height” (p. 1).

A number of studies highlighted the significance of the problem. Of the 76 approach-related accidents and incidents from 1984-1997 subject to the FSF study, 45% found airplane handling to be a causal factor with poor energy management an associated factor. Further, 36% constituted low-energy approaches, and 31% involved high-energy conditions (FSF, 2010). Sherry et al. (2013) conducted a data mining analysis of aborted approaches and their underlying factors, examining 21 days of radar surveillance track data at Chicago O’Hare International Airport to determine the rate at which such events occurred. They concluded an abort rate, for all possible causal factors, at 7.4 per 1,000 approaches, with a daily variance ranging from 0 to 21 events per 1,000 approaches on a given day (Sherry et al., 2013). Of further significance, their study analyzed 467 voluntary pilot and/or controller reports drawn from the NASA Aviation Safety Reporting System (ASRS) and found that 48% of the aborted approaches were due to airplane issues, of which 4% were the result of high and fast conditions and 3% were due to low-speed and other approach issues (Sherry et al., 2013).

Although conceptually beneficial, this basic definition lacked the concrete parameters that pilots should not exceed to remain stable. As stated in the overarching FSF definition, the determination of stability criteria is left to the operator to develop. However, the ALAR Briefing Note 7.1 provided further guidance to operators in crafting such SOP guidelines. These guidelines have been widely adapted by operators and manufacturers across the entire spectrum of aviation, to include not only air carriers, but business and private aviation throughout the world. Applying a rules-based approach, the FSF (2010) proposed the criteria detailed in Table 1.

**Table 1**

*Early Stabilized Approach Criteria*

| Element | Criteria |
|---|---|
| Profile | The aircraft is on the correct flight path |
| | Small heading/pitch changes to maintain correct flight path profile |
| | Specific types of approaches are stabilized if they also fulfill: |
| | CAT I ILS – 1-dot deviation of glide path and localizer |
| | CAT II/II ILS – within the expanded localize band |
| | Circling – wings level on final when 300 ft above airport elevation |
| Configuration | In the correct landing configuration |
| Energy | Airspeed stabilized $V_{REF}$ +20 kts and not less than $V_{REF}$ |
| | Power appropriate to aircraft configuration and no below minimum power for the approach |
| | Vertical speed no greater than 1,000 fpm |
| | If required sink rate greater than 1,000 fpm, a special briefing conducted |
| General | All briefings and checklists have been conducted |
| | Unique approach procedures or abnormal conditions requiring deviation from above elements require a special briefing |
| | Unstabilized below 1,000 ft above airport elevation in IMC or below 500 ft above airport elevation in VMC requires immediate go-around |

*Note.* CAT I refers to a Category I instrument approach condition, $V_{REF}$ refers to the commanded approach speed for the airplane. Adapted from “Killers in Aviation: FSF Task Force Presents Facts about Approach-and-landing and Controlled Flight into Terrain Accidents,” by the Flight Safety Foundation, 2010, *Flight Safety Digest*, *17*, p. 80. Copyright 2010 by the Flight Safety Foundation.

Of significance to the study were those parameters addressing flight path management, airspeed control, and sink rate. Other factors, such as accomplishment of required briefings, planning consideration for unique non-normal procedures, appropriate engine thrust settings, and landing configuration are critical to flight safety. However, a number of these factors would originate prior to the approach and, once addressed, often would be compartmentalized or verified through checklist action at some point prior to completion of the approach.

Airplane manufacturers may also provide recommended elements for stabilized approach criteria or modify those offered by the FSF. As noted in a customer training manual, The Boeing Company allowed for a slightly different valuation for airspeed on final, suggesting that the airplane should be on the selected approach speed but allowing that deviations no greater than +10 kt or no less than –5 kts are acceptable, provided the current airspeed is trending toward the commanded speed for the approach (TBC, 2007). In addition, Boeing provided additional considerations for the terminal phase of the approach:

> At 100 ft HAT [Height Above Touchdown] for all visual approaches, the airplane should be positioned so the flight deck is within, and tracking to remain within, the lateral confines of the runway edges extended.
>
> As the airplane crosses the runway threshold it should be:
>
> • stabilized on approach airspeed to within +10 kts until arresting descent rate at flare
>
> • on a stabilized flight path using normal maneuvering
>
> • positioned to make a normal landing in the touchdown zone (the first 3,000 ft or first third of the runway, whichever is less). (TBC, 2007, p. 5.5)

However, both sources lacked numeric guidance as to what constitutes roll and power stability on an instrument or visual approach. The United Kingdom Civil Aviation Authority (UK CAA) provided insight derived from studies on accidents:

> To investigate the effect of including Roll and Power in the stable approach algorithm, the instantaneous maximum value of roll and the maximum and minimum of N1 [engine rotational speed as a percentage of maximum revolutions per minute] were obtained. In the absence of detailed performance data, the stable state trigger values were chosen statistically (outside 2 standard deviations) for N1 were 65% and 30%. Nominal roll angles were selected according to altitude: - above 1000ft (10deg), between 1000 and 500ft (8deg), below 500ft (6deg). (UK CAA, 2012, p. 21)

Given the UK CAA research, a possible enhancement for operator stable approach criteria might be the inclusion of specific roll angle limits, whether graduated with elevation above the runway or held at a constant value throughout the approach. Equally, anecdotal evidence suggested roll angle should be included; a number of unstable approaches in which roll perturbations resulted in less-than-desirable roll angles during the approach and landing phase have occurred. When commonly accepted guidance stipulates only minor heading changes, aircraft control theory would support such minor changes cannot be invoked without modifying the roll angle of the airplane to redirect the flight path.

Past studies often did not address the roll parameter following exclusion from the guidance, but recent research supports rethinking this concept. In a case study on hard landings, Bardou and Owens (2014) noted that while normal stabilization criteria vary among specific aircraft types, a pilot callout as to deviation from a stable approach should typically be triggered if the roll angle exceeds 7°, a value in line with the UK CAA (2012) recommendation. Stable approach and go-around studies by Singh et al. (2020) and Campbell et al. (2019) suggested roll angle is a factor in maintaining approach stability. Singh et al. (2020) used accepted unstable approach criteria in their investigation using a sparse variation Gaussian process for real-time unstable approach detection but modified it slightly to include not just minor heading changes but "little changes in…bank" as well. Campbell et al. (2019) asked study participants to conduct successive instrument approaches to measure the effect of presented unstable situations on touchdown performance. A secondary study task was to gather pilot subjective data as to the decision to execute a go-around maneuver. Pilots cited roll angle as the justification in 5.6% of the cases. Follow-on questioning indicated pilots felt the proposed go-around criteria should be adjusted to include roll angle, suggesting values ranging from 5-15°. Further, de Boer et al. (2014) drew conclusions from their study on automatic identification of unstable approaches using flight data analysis. The data supported modification of the FSF criteria to address roll angle excesses as a function of the angular value, the duration of that value, and height above ground.

***Recent Concepts***

In 2011, shortly after the release of the FSF ALAR Briefing Note 7.1 (2010), the FSF commissioned the Go-Around Decision Making and Execution Project to examine an observed lack of go-around compliance when confronted with violated published approach stability criteria (FSF, 2017). The project was conducted over the span of several years with the results published in 2017. The study cited that in examining accident data from 1994-2010, Burian (2011) identified that unstable approaches occurred in approximately 4% of all approaches flown. Yet despite the existence of SOPs and other standing guidance, 95-97% of those pilots found to be in an unstable condition elected to continue the approach to landing. This aligned with the work of Wang et al. (2015), who performed similar analysis of 8,219 approaches conducted by a single air carrier at an unnamed airport and quantified the rate at which the airspeed element of an unstable approach may have been violated. By reviewing surveillance track data for significant groundspeed changes, they found 27.8% of the approaches reviewed exhibited a groundspeed change in excess of 10 kts when examined between 1,000 and 750 ft, 14.1% when between 750 and 500 ft, and 4.4% when below 500 ft above field elevation (Wang et al., 2015). They also examined vertical path sink rate against the 1,000 fpm stipulation suggested by the FSF. Their results, the order of which are commensurate with that of the groundspeed, were 1.9%, 0.7%, and 0.2% respectively (Wang et al., 2015). As groundspeed may not be completely accurate owing to the effect prevailing wind may have on the value, it still serves to indicate the incidence in which the guidelines may have been held in ill regard.

These findings proved antithetical to current understanding of expected pilot responses. The aviation community had assumed that compliance through execution of a go-around would occur, but the FSF evidence placed industry-wide unstable approach policy compliance over a 16-year period at approximately 3% (FSF, 2017). The FSF Go-Around Decision Making and Execution Project then posed two questions: why were pilots intentionally non-compliant when faced with an unstable approach condition, and why was management not enforcing this policy? In response, the study found there existed a collective industry normalization of non-compliance with the unstable approach go-around policy in the category Shappell and Wiegmann (2000) deemed to be routine violation unsafe acts of pilot operators, supported by apparent failure by management to correct the problem (FSF, 2017). Pilots did not see the relevance of the criteria in an operational environment (FSF, 2017). Using a two-segment approach, the study sampled 2,340 pilots from diverse backgrounds, continents of operations, types of operations, and experience levels. The first segment asked respondents to recall a recent incidence of unstable approach and identify the various characteristics and factors affecting that event. In the second segment, the respondents described their tolerance thresholds for initiating a go-around, given the conditions of path deviations, velocity deviations, and sink rate deviations. The results demonstrated that pilots, on average, exhibited a belief they could compensate effectively for approach instabilities to a much lower height above the ground than the original FSF study in 2000 suggested (FSF, 2017). The study concluded that while a number of the stability criteria remained sound and should be retained, a significant number should be modified to align the triggering conditions to match apparent operational practices. The purpose of the new FSF stabilized approach guidelines was to establish a *gate* construct and allow more pilot judgment in the decision to execute a go-around when encountering an unstable condition. The criteria were categorized into profile, configuration, energy, and general elements; this grouping likely assisted the pilot in cataloguing and processing the specific element requirements to be met. The proposed new stabilized approach criteria are presented in Table 2.

**Table 2**

*New Stabilized Approach Criteria*

| Element | Criteria |
| --- | --- |
| Profile | Small heading/pitch changes to maintain correct flight path profile |
| | Specific types of approaches are stabilized if they are within: |
| | CAT I ILS – 1-dot deviation of glide path and localizer |
| | RNAV – ½-scale vertical/lateral deflection and within RNP requirements |
| | LOC/VOR – 1-dot lateral deviation |
| | Visual – 2.75° to 3.25° of visual approach path indicators, lined up with the runway centerline no later than 300 ft. |
| Configuration | Landing configuration – gear and flaps set, speed brakes retracted |
| Energy | Airspeed stabilized $V_{REF}$ +10 kt to $V_{REF}$ without wind adjustment |
| | Thrust stabilized to maintain the target approach airspeed |
| | Vertical speed no greater than 1,000 fpm |
| General | Stabilized approach gates should be observed |
| | Active communication calls made during each approach |
| | Bracketing corrections to maintain stabilized conditions: |
| | Occasional momentary overshoots due to atmospheric conditions are acceptable |
| | Frequent or sustained overshoots are not acceptable |
| | Unique approach procedures or abnormal conditions requiring deviation from above elements require a special briefing |

*Note:* CAT I refers to a Category I instrument approach condition, $V_{REF}$ refers to the commanded approach speed for the airplane. Adapted from “Go-around decision making and execution project”, by the Flight Safety Foundation, 2017, p. 44. Copyright 2017 by the Flight Safety Foundation.

The revised profile criteria incorporated scenarios in which non-precision approaches occurred, an area originally unaddressed in the 2000 FSF study and recommendations. These scenarios included Area Navigation (RNAV), localizer (LOC)/Very High Frequency Omni-directional Range (VOR), and visual. The approach path profiles were considered stabilized if they fulfilled certain criteria, to the tolerance of ½-scale vertical and lateral deflection, 1-dot lateral deviation, and within 2.75 and 3.25° of the visual approach path indicators and lined up with the runway no later than 300 ft above ground level (AGL), respectively. The aircraft configuration requirement remained relatively unchanged, with added clarification of specific airplane systems – gear, flaps, and speed brake – positioning. The energy element comprised three measures: airspeed, thrust, and sink rate.

Rather than continue the use of prescribed altitude stability points that were based upon a given meteorological condition (FSF, 2010), the new criteria applied a consistent gate system using three altitudes. Each altitude was associated with a specific objective. At 1,000 ft AGL, the airplane should be configured in the final landing configuration, although some allowance was granted to establish this objective as early as 1,500 ft AGL and as late as 800 ft AGL, as may be warranted by a specific airplane category. At 500 ft AGL, the airplane should be fully stabilized, as defined in Table 2. At and after 300 ft AGL, the approach should be terminated and a go-around initiated if in an unstable condition (FSF, 2017). There existed within this new framework, however, an avenue for pilot discretion. Per the note provided, "Continuing past the related gate should only occur if meeting the objective of the next gate is achievable; otherwise, go around. Example: If the flight is not configured by 1,000 ft, it could continue if being fully stable by 500 ft is achievable" (FSF, 2017, p. 44). This allowance for discretion required the pilot to conduct a real-time assessment as to the current airplane state – flight path, energy, and duties – and forecast the likelihood of improving the situation within a very short timeframe.

When reviewing the study that led to the newly-proposed criteria, a number of points were highlighted. First, the genesis was initiated in concerns surrounding non-compliance of existing go-around directives for unstable approaches. The study noted that although considerable research has been done on factors contributing to approach and landing accidents, there existed no known study addressing the psychological aspect of the phenomenon of pilot non-compliance with company policies and the election to continue an unstable approach (FSF, 2017). Second, the methodology used was survey, in which 2,340 pilots were asked to recall the recent instances in which they found themselves in an unstable approach *below* the 2010 criteria and provide subjective detail on their thoughts and attitudes, as well as recalled objective detail on the airplane and environment condition (FSF, 2017). Pilots were then provided a hypothetical situation in which they were presented randomly assigned risk conditions, and then asked the lowest altitude in which a safe go-around could be accomplished. Pilots were assigned into one of four groups in terms of analysis, based on individual flight experience recall from the past five years. The groups were divided into (a) those who could recall only continuing to land from an unstable approach or approaches, (b) those who could recall only having flown go-arounds without an unstable approach present, (c) those who could recall executing a go-around from an unstable approach and were randomly assigned to recall their most recent go-around, and (d) those who could recall executing a go-around from an unstable approach and were subsequently randomly assigned to recall the most recent unstable approach (FSF, 2017).

The FSF research included an SA assessment, drawn from the psychological and social factors surveyed, using a proprietary model. The results showed pilots who recalled continuing unstable approaches were situationally less aware than those who conducted a go-around. Further, pilots who perceived a lower level of threat posed by the instability of the approach scored much lower on the SA assessment. Mental models, also referred to as schema, constitute an internal expectancy as to a particular system – its components, its functionality, and its employment – and are developed through repeated practice to elicit the appropriate procedures and actions to follow when presented an applicable scenario (Wickens, 2003b). But these models can be of varying accuracy in terms of completeness and correctness (Wickens et al., 2004). Reduced competencies in developing SA may manifest in pilots who are prone to continuing an unstable approach, thereby adopting mental models that further minimize the risk, aggravate the ability to employ safety threat mitigations, and feel greater comfort when operating at the edge of the safe flight margins (FSF, 2017).

Even as the FSF study developed a thorough framework as to the perceptions and attitudes of a sampling of pilots, some issues remained. The effect of pilot recall on the accuracy of recalled details may impact the results. Subjective, direct measures of SA as employed in the FSF study may be confounded by memory lapses, misinterpretation of the experience, lack of self-awareness, and other threats to study validity (Endsley & Jones, 2012). In retrospect, pilots can become subject to a false memory mechanism. Mosier et al. (1998) described an experiment in which 25 pilot participants were presented an engine fire message on the airplane Engine Indication and Crew Alerting System (EICAS) but lacked corroborating indications such as fluctuating engine parameters and a number of other indicators. A full 67% of the study participants displayed a false memory of having been presented at least one other cue, and many indicated memory of two or more cues – none of which were present (Mosier et al., 1998). Equally, those encountering a more severe unstable approach condition may be more prone to recall specific data due to the intensity of the experience. The lack of detail as to the severity and duration of deviation detracted from the significance of the findings, preventing analysis of diverse levels of variance from a stable approach. Even though the study gathered a broad palette of flight operation types, operating locations, and pilot experience levels, the diversity of the airplanes flown and accompanying flight instrument configurations prevented ascertaining the impact of the display design on the outcome.

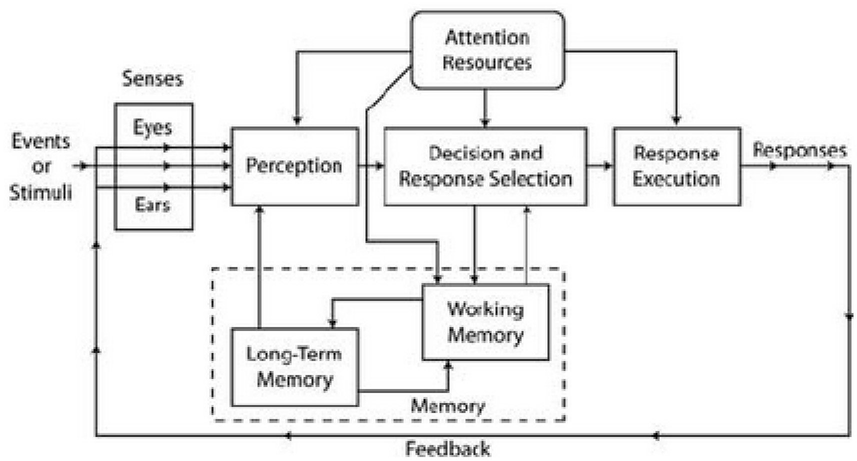
The 2017 FSF study appeared to approach the unstable approach issue using valid survey data but fell short of validating its findings in terms of outcomes resulting from application of the new criteria. Thus, these findings served as recommendations to industry to be assessed and either implemented or rejected. Without additional research, it was impossible to ascertain whether implementation would place pilots at elevated risk.

**Cognitive Theories Relevant to Display Design**

In a study examining the manner in which pilots handled a side stick-configured airplane during a manually flown approach, Haslbeck et al. proposed “an operator’s active task consists of three major processes: sensory information acquisition and perception, cognitive processing, and response execution” (2012, p. 2.1). Mosier et al. (1998) demonstrated the significant ramifications of a poorly-designed display system. Of note, 100% of the pilots engaged in commission errors by shutting down the engine in which a fire was indicated on the EICAS display. This shutdown was accomplished without confirming the accuracy of this presentation, despite having identified the need to do so in the post-experiment survey and having been trained on the verification items during pre-experiment training. A review of the more prevalent Aeronautical Decision Making models highlighted the presence of cue processing, detection, problem identification, and stimuli blocks at the onset of the respective process (O'Hare, 2003). One such model, the Human Information Processing model proposed by Wickens, is shown in Figure 3 (Wickens et al., 2015).

**Figure 3**

*Wickens' Human Information Processing Model*




*Note.* Excerpted from *Engineering Psychology & Human Performance* (4th ed.) (p. 17) by C. D. Wickens, J. G. Hollands, S. Branbury, and R. Parasuraman, 2015, Psychology Press. Copyright 2015 by Psychology Press.

An event or stimuli provides a sensation to the eyes and ears to be perceived. It should be observed that other cues, such a tactile and haptic, proprioceptive and kinesthetic, and equilibrioceptic serve as stimuli feeding the process. Such cues, however, are generally less relevant when the display system is taken in isolation. However, the design and placement of system controls would certainly be contributory if critical information necessary to the perception process were presented in that manner.

Visual information from the display gains access to the brain and is placed in short-term sensory storage, which helps prolong the stimulation (Wickens et al., 2015). Perception then may serve a critical role in the decision-making process, processing raw sensory data to derive meaning from it. Without it, a pilot cannot be expected to avoid the possibility of acts of omission – failing to recognize and address the presence of an undesired state. Wickens et al. (2015) explained two key features of perception. First, it tends to occur automatically, without intention on the part of the pilot, and does so rapidly. Second, it involves both top-down and bottom-up processing, and as such is fed by both long-term and working memory. When the pilot observes an alert message on a display, it is quickly perceived and perhaps expected. However, its meaning must be understood in order for the pilot to derive sense from it. Cognitive operations such as reasoning and image transformation are carried out in working memory, where the pilot understands the meaning of the stimulation (Wickens et al., 2015).

Decision and response selection and execution of that response follow the perception and cognition. When a pilot understands the airplane flight path may not be as desired, corrections are determined and inceptor changes made to correct the path. Feedback, in the form of a change in the flight path display presentation, provides visual evidence the response executed achieved the desired result. Key to this model is understanding that stimuli may come from either an input external to the process, for example the first glance at the display, or from on-going input, such as error indication when tracking a commanded path.

Perception, memory, decision and response selection, and response execution all draw from limited attention resources available to the pilot. Even in benign situations, attention tunneling may occur. To that end, Wickens (2003a) provided a series of display principles to be followed in developing graphic information presentations of significance and relevance to pilots.

**Design Principles Relevant to Display Design**

Wickens' (2003a) display principles included consideration of information need, legibility, display integration/proximity compatibility, pictorial realism, moving part, predictive aiding, and discriminability. Wickens (2004) later expanded on such principles, adding consideration as to the avoidance of absolute judgment, application of top-down processing, the benefits of redundancy gain, minimization of the resource cost to information access, the use of multiple resources, and the value of consistent presentation. Such principles guide the manner in which the display presents a given airplane state.

Flying is a task involving vigilance and action. Pilots execute vigilance in the form of monitoring displays, systems operation, and the environment, and then undertake action in the form of decision-making execution and airplane control management. Dillard et al. (2014) characterized the nature of vigilance, noting "Vigilance or sustained attention tasks require observers to monitor displays for extended periods and detect the appearance of critical signals…vigilance tasks can be described as 'go/no-go attention assignments in which the frequency of 'no-go' events outweighs the 'go' events" (p. 1364). Although the duration of the final approach segment is short in actual terms, it often follows an extended segment of enroute time in which continuous monitoring of the airplane system, inclusive of displays, must occur. As such, it represents the tail end of a much greater period of vigilance. To ensure attention capture and focus, boundaries should incorporate legible symbol constructs that provide only the information necessary to improve SA, integrate within the existing display framework such that access is achieved with minimal resource cost, discriminable in not easily confused with values acceptable to the operation, and provide an immediate context when first observed (Wickens, 2004). Proper adherence to such design principles furthers effective top-down processing, where the information presented is as the user expects it to be based upon previous experience.

In a study on pilots' visual scan patterns and SA in flight operations, Yu et al. (2014) identified key considerations relevant to flight instrument scanning. They highlighted the majority of the information retrieved and processed by pilots is obtained through the visual channel by means of scanning; of this visual scanning, the vast majority of the pilot errors that occur result from "poor perceptual encoding," and that "attention plays a central role in cognitive processing" (p. 708). Of greater significance, they commented that "due to the limited capacity of a human's working memory, it is necessary to focus attention on the most critical task at hand and ignore stimuli from the environment when selecting the visual channel to be attentive to" (p. 708).

Thus, in the determination of the channels to which to attend, one must consider the four factors enumerated by Wickens (2004): salience, effort, expectancy, and value. The balance of these factors contributes to a useable, effective display enhancement. Captain Bechara Mallouk, Boeing Technical Fellow, detailed a mnemonic applied in aviation system design to facilitate graceful degradation into a failed condition: hint, alert, nudge, and distract, or HAND (personal communication, December 4, 2019). Hint modalities provide low-level intimation to the pilot that it may be beneficial to monitor the particular parameter. Examples of hints include the use of navigation performance scales and position trend vectors – predictive aids – to suggest that the airplane may be approaching a particular limit, be it system operation or airplane flight path. The downward presentation of an airspeed trend vector employed on some attitude presentations would fall under this category. Alert modalities raise the level of notification; communications such as EICAS message or FMS message act to draw the attention of the pilot to the issue. An EICAS indication of an airspeed discrepancy between the two primary indications in an airplane crewed by two pilots serves as an example. When a nudge modality is invoked, a prompt is provided to encourage the pilot to make the appropriate entry or take a particular action in an attempt to alter current behavior. Such a feature might be a flashing indicator or message when a parameter is exceeded, such as a roll angle indicator turning amber when exceeding a prescribed angle of roll. By design, the undertaking of that nudge is voluntary (Thaler & Sunstein, 2021). A distraction modality, then, is the presentation of a noise, an aural cue or warning, or tactile and kinesthetic sensations such as the application of a stick shaker or stick pusher in the case of an approach-to-stall scenario.

***Usability Principles***

The introduction of any display enhancement poses a risk of introducing excessive clutter to the display and reduces the perception of the other parameters being presented. Displays should consider usability principles and human-computer interaction (HCI). As this display was automatically rendered upon completion of both the procedural steps to select a particular approach and guidance path capture, no additional flight crew actions were required. Thus, many of the general interface design principles of Nielson (1994), as they addressed switch application, tracking, user interface, control and freedom, error prevention and recovery, and flexibility of use were generally inapplicable to the proposed enhancement. The one exception was aesthetic integrity. Usability principles, however, are an integral part of a good design.

Data on the usefulness of the display, satisfaction with the presentation, readability, placement location, clutter, frequency of reference to the features, frustrations in interpretation, caution display latency, and suggested improvements (Wickens et al., 2004) should be gathered when assessing a given display. Yeh and Chandra (2004) identified four questions useful in assessing a display presentation, and these form the foundation of the pattern of inquiry. They note:

> The design and selection of symbols should consider the range of functions for which the display will be used. Symbols can be tested prior to use in order to determine their usability by measuring the performance impact against each of the following criteria:
>
> - Is the symbol easy to find?
> - Is the symbol distinctive from other symbols?
> - Is the on-screen symbol size appropriate?
> - Can all encoded attributes of the symbol be decoded quickly and accurately? (p. 5, C.3-2).

Wickens et al. (2004) commented there exists a long history of human factors evaluation and may be more inclusive in its assessment of a system than a usability evaluation, owing to the complexity of systems and organization design. Usability test has evolved into the realm of HCI and pertains more to the *interaction* between the user, the pilot, and the system itself.

**Human-Centered Design**

In developing display enhancements, system designers must balance the inclusion of SA-enhancing information against the possibility of SA degradation due to oversaturation of the user in the form of excessive information to process. In examining workload and SA, Vidulich (2003) differentiated between the presentation of new information and the reformatting of existing information.

> If the new information must be processed in addition to all of the other information, then the processing resources required should be increased, and a workload increase should be incurred. But it is also possible that the new information could allow a change of strategy that would eliminate the impact of mental workload or even allow a reduction to occur. (p. 135)

One could infer difficulty predicting the impact of any design change to user SA on the basis of theory alone. Empirical studies must be undertaken to determine the net change to the user's SA. Vidulich (2000) conducted a meta-analysis of 18 studies in which SA measurement was examined for sensitivity to changes in user interfaces, which would be inclusive of display change permutations. Of the studies, nine employed the use of added information, whereas the other nine applied reformatted information. In the case of added information, seven demonstrated statistically significant increased SA outcomes, two showed no change, while none resulted in loss of SA. For the application of reformatted information, all nine resulted in a statistically significant increase in SA (Vidulich, 2000).

This survey suggested that a properly-designed display addition or reformatting would result in a gain in user SA.

***Situation Awareness Oriented Design***

Endsley and Jones (2012) introduced the concept of Situation Awareness Oriented Design (SAOD) in the development of displays provided pilots. They provided 50 SA design principles to be taken into consideration when developing systems. These principles were grouped to address qualitative aspects of display and automation systems as well as those of multi-operator scenarios and SA training constructs. Of particular application were those addressing general SA assistance and cues, the certainty and salience of the data provided within the displays, the complexity and density of the data presentation, and the benefits and pitfalls of alarm systems.

Even though many of the SAOD principles appear to deal with specific design elements supportive of enhanced SA, a particular few addressed the broader objective of supporting a stabilized approach by focusing on the core data presentation issue. Endsley and Jones (2012) highlighted the underlying objective of this principle by stating, “requiring the operator to learn and remember a complex series of actions in order to perform a task not only adds to the operator’s cognitive load, but also leaves room for error if one or more of the steps is forgotten or performed incorrectly” (p. 145). Further, they noted that “reducing the number of steps needed to achieve a particular system state lessens the likelihood that an error will be committed and reduces the complexity of the mental model the operator must develop in order to interact with the system” (Endsley & Jones, 2012, p. 145).

The application of memorized actions to various tasks is a common theme within the aviation community (Dismukes, 2010; Jones & Endsley, 1996). A prime example includes accomplishment of the flight deck preflight, where pilots execute memorized flow patterns to ensure the various system controls and display settings are set in a safe, standardized, and appropriate position (Boeing, 2007). These actions tend to be thought of in terms of sequential, independent activities whereby the steps are completed in an ordered list until the final item is accomplished. Airplane manufacturers vet these steps to ensure no unintended consequences arise from their completion.

Accomplishment of dynamic tasks such as tracking a path within a set of parameter values requires a continuous, time-sequenced series of evaluations (Allsop & Gray, 2014; Wickens et al., 2004). Though not a strict procedure in terms of exhibiting a formed listing of steps to be accomplished, they nonetheless add to the task complexity and should be reduced. These evaluations would include the perception and comprehension of key parameter states and projection to future conditions given the current rate of deviation, achievement of Level 1, 2, and 3 SA (Endsley, 1995b; Wickens, 2016; Wickens & Carswell, 1997). These data to be perceived and comprehended include adherence of the airplane reference symbol to the ILS flight vertical and lateral path guidance, the difference between the commanded airspeed and current airspeed, the roll angle, the rate of descent, and the current altitude.

***Projection and Alarms***

With SAOD Principle 25, Endsley and Jones suggested a better method to presenting an unstable approach might be to provide the pilot a projection of an approaching unstable state, rather than alarm them to the existence of one. They reflected on alarms in general, noting "By their very nature, alarms put people in the position of being reactive. When an alarm sounds, they must act to develop an understanding of why it alarmed and what they should do. The alarm itself adds stress to this process" (Endsley & Jones, 2012, p. 161). The better strategy to follow, they suggested, is to provide the pilot with the SA-supporting information to project and subsequently make corrective control input decisions necessary to maintain a stable approach (Endsley & Jones, 2012). Supporting information has been incorporated in the evolutionary progress of modern flight instrument displays. Commonplace are command values to draw attention to the pilot, such as a targeted airspeed to fly or an altitude to be captured. Similarly, boundaries calling out excessive bank angle, aerodynamic stall, minimum maneuver speeds, secondary controls (landing gear and flaps) airspeed restrictions, and runway elevation are found useful to the pilot.

One example of supporting projection is the use of displays that show parameter trends. Appendix A presents a typical modern, electronic flight instrument display, a Primary Flight Display (PFD) used in current variants of the Boeing B-737. This presentation meets the criterion of a technically advanced airplane (TAA), as defined by the FAA, in presenting an airspeed indicator, turn coordinator, attitude indicator, heading indicator, altimeter, and vertical speed indicator (FAA, 2011). With minor exception, the display parallels those of other manufacturers in presenting data in relative locations reminiscent of earlier mechanically-based instrument suites. The PFD item at Callout 8 provides not only the target airspeed and current airspeed value, useful information in support of SA Levels 1 and 2, but also the airspeed trend vector. By presenting the predicted airspeed as derived from the current state of acceleration or deceleration, the pilot can easily project the future energy state of the airplane. Conformation to the FSF (2010) stable approach criteria of airspeed not to exceed $V_{REF}$ +20 kts or reduce to less than $V_{REF}$ is somewhat simplified, as the presentation of the future state is instantaneous, freeing the pilot of the need to conduct several samples of a static parameter to determine the airspeed rate of change. However, without proper marking of the acceptable criteria, the pilot is still left to overlay the boundaries to assess whether the airplane will soon be outside the criteria.

Principle 25 did not preclude the use of alarms (Endsley & Jones, 2012). Properly used, an alarm system can aid in automating the decision to discontinue the approach. This can prove useful when the pilot experiences a high degree of task saturation and fails to recognize presentation of the exceedance. However, Endsley and Jones (2012) cautioned alarms present a reactive, rather than predictive, level of understanding for the pilot. This can add stress as the pilot seeks additional information to understand the root cause of the alarm. They recommend providing supporting data on the parameters used to trigger such an alarm.

***Validation and Tradeoffs***

Principle 26 stated that crews will seek to confirm or deny the validity of the alarm, and displays should provide the underlying raw data. Endsley and Jones (2012) remarked, “At least partially because false alarms are such a problem, people do not immediately respond to alarms but often seek confirming evidence to determine whether the alarm is indicative of a real problem" (p. 152). For example, when a pilot conducts a takeoff into IMC with a flight instrument suite that employs an FD system, a tendency exists to blindly follow the guidance under the assumption of system accuracy and integrity. However, such an approach can be fatal when the guidance is faulty. Consequently, pilots are directed to "look through" the guidance to the underlying pitch attitude reference and confirm the FD-commanded parameters are sound.

Such confirmation is not without cost. Wickens (2003a) reminded that the amount of effort entailed in this verification is minimized when the confirmatory information destination is of greater proximity and the degree of comparison and integration of the information necessary is reduced. In the early implementation of Ground Proximity Warning Systems (GPWS), a flight instrumentation tool developed to prevent pilots from undesired encounters with closing terrain, studies by DeCelles (1991) demonstrated the presence of pilot delays when responding to the system. System designers assumed almost immediate response to system alarms, yet nearly 73% of the time the pilot delayed response to the command, seeking internal, instrument-based, or external, visual –based, verification. Pritchett and Hansman (1997) identified similar issues during simulated flight operations at closely spaced, aligned runways with another avoidance system meant to prevent mid-air collisions. Hesitancy of the response was affected by the manner in which the criteria for the display was presented.

Principle 29 invoked judicious application of tradeoffs between missed alarms and false alarm rates and avoiding what Wickens et al. (2004) suggested might result from excessive false alarms: flight crew attempts to disable the alarms and the issue of mistrust of the alarm itself. Endsley and Jones (2012) pointed out that vigilance tasks, where the pilot is actively focused on the task at hand, may be more forgiving of the presence of false alarms in order to minimize the possibility of a failure to present an alarm when one is warranted. They cautioned, however, that during periods of moderate or higher workload, the presence of high false alarm rates should be avoided. Such frequency can detract from the pilot executing the task at hand, degrading pilot responsiveness to the alarm and nurturing a general distrust in the alarm system. In instances where the number of alerts is high, assistance may be required to notify the pilot of the presence of a hazard.

**Elements of Memory**

Wickens and Hollands (2000) and Tsang (2003) presented an overview of predominant memory types, mainly working memory and long-term memory. Short-term working memory was defined as “the relatively small amount of information that one can hold in mind, attend to, or technically speaking, maintain in a rapidly accessible state, at one time” (Cowan, 2005, p. 1). Wickens and Hollands (2000) likened short-term memory to the “‘workbench’ of consciousness where we examine, evaluate, transform, and compare mental representations” (p. 241). It serves the pilot as a repository for information drawn from either long-term stores or newly obtained from perception and cognition processes, to be applied to a current situation or newly encoded into long-term stores. Working memory is comprised of three components: the verbal component consisting of the phonological store and an articulatory loop for rehearsal, the spatial component of the visuospatial sketchpad, and the central executive. Pilots apply the verbal component when they identify particular words or sounds, such as directions from

ATC or the excitation of an aural stall warning alert. The spatial component supports analog information retained in the form of visual imagery. An example would be the presentation of a visual alert to the pilot indicating conflicting traffic and its vertical path trend. Such encoded graphics would be recognized and interpreted in a spatial context. The central executive, finally, moderates working member activity and resource allocation (Wickens & Hollands, 2000). It is the pacemaker of the system. Working memory can be subject to interference, impacting the ability of a pilot to apply the information in a timely, or appropriate, fashion. Measures to reduce such interference include rehearsal, both verbal and spatial, and avoidance of proactive and retroactive interference (Wickens & Hollands, 2000). Working memory can also fall prey to confusion when items of similar content are present.

Long-term memory, on the other hand, provides for more permanent storage. It is identified by several attributes to include long storage periods often measured in years, a meaningful system of data organization, and a repository for expertise and working models (Tsang, 2003). Long-term memory is critical to the learning process, is stored in associative networks, draws from the strength and associations of the material retained, and can be distinguished as either semantic- or event-based (Wickens et al., 2004). Other than associative networks, information in long-term memory is stored in schema, mental models, and cognitive maps. Schema constitute the entirety of knowledge about a particular topic, such as a sequence of activities for depowering an airplane after flight. Mental models, on the other hand, detail dynamic systems – their design, operation, and application to the environment – and are framed in a set of expectancies (Wickens et al., 2004). Cognitive maps tend to represent spatial information and are analogous to the visuospatial component of working memory. During the critical approach phase of flight, pilots must draw upon their long-term semantic memory to accomplish a number of tasks, such as defined instrument approach and go-around procedures, as well as unexpected procedures that may be essential in an inadvertent approach-to-stall condition. Pilots may also draw from long-term memory stores to apply a mental model to a particular situation. Long term memory is not without its fallibilities, with the potential for retrieval difficulties, tendencies toward strength and association decays, interfering associations (Wickens et al., 2004).

An additional form of memory, long-term working memory, was postulated by Ericsson and Kintsch (1995) to address an apparent skilled memory phenomenon. Activities such as flight display interpretation or route planning require the use of working memory, but also must draw on long-term memory. Skilled tasks are often interrupted, well past the time in which one might expect short-term working memory to fade, and yet demonstrate little degradation in performance. Additionally, skilled tasks may require copious amounts of information, beyond the limits associated with short-term working memory, to be recalled and applied quickly (Wickens & Holland, 2000). Long-term working memory retains the durability of long-term memory but is recalled by temporary retrieval cues that support quicker access and transfer to the working memory register. However, such memory is event- or domain-specific; a pilot able to access long-term working memory using a retrieval structure for one particular skilled task likely will not be able to do so for other tasks, reverting to the limitations associated with normal short- and long-term memory. Tsang (2003) differentiated types of working memory by their duration: short-term and long-term. Short-term working memory duration can be measured in minutes (min) or seconds (s), whereas long-term working memory ranges from hours to minutes.

The use and allocation of memory is closely linked with the creation and maintenance of SA. Pew (as cited in Endsley, 2000, p. 30) provided the salience in commenting, “SA requires immediate access to the procedures required to accomplish a task as well as the information required. Skilled performers will carry the procedures in long-term memory and bring them into working memory when they are specifically needed.” Current compliance with unstable approach mitigation measures is left to the pilot, where they must draw the criteria from long-term memory into working memory. The task of the pilot is to observe and perceive the data presented by the flight instruments, assess the data against the mitigation criteria to understand the current state, and predict the future airplane state (Endsley & Jones, 2012).

Studies have shown that humans will experience reduced levels of cognitive function when faced with demanding situations: engaged in a stressful activity, exposed to a threat to their well-being, or overloaded with an excess amount of information. The occurrence of a surprise event can impair the pilot’s working memory as attention is focused on elements of greater salience (Martin et al., 2015); such attention channeling and memory impairment may encumber the pilot’s ability to perceive, understand, and act upon an unstable condition. As Martin et al. (2015) concluded, “Narrowed attention, decreased search behavior, longer reaction time to peripheral cues, decreased vigilance, degraded problem solving, performance rigidity, and degraded working memory function are just some of the cognitive impairments noted under the effects of stress” (p. 100). They noted that stress events result in a period of significant cognitive disruption and moderate startle can lead to degraded information processing for as much as 30 s, a significant portion of the approach segment of flight (Martin et al., 2015).

According to Burian et al. (2005) the length of retention time and quantity of information that can be kept in the pilot's working memory are inversely proportional, as the stress levels increase, the capacity and retention of working memory decreases. They cite working memory as critical in retaining and manipulating information in the cognitive arena. As a result, high stress and workload levels present not only the potential to miss relevant cues but impact the ability to pull together disparate information within working memory. Wickens et al. (2004) reinforced this concept, identifying working memory loss as an outcome of stress exposure. A pilot facing high stress situations, such as encountered in gusty wind conditions, low visibility, or at night, will be hampered in cataloguing and processing subsequent information input such as checklist completions, missed approach instructions, approach airspeed, and altitude callouts. Human et al. (2018) cited previous studies in which elevated cortisol levels negatively affected both the speed and accuracy of working memory tasks and did so to a greater extent when the participant was placed under higher cognitive loads. This relationship has been demonstrated in laboratory and real-world human performance studies, although the magnitude of the relationship is not absolute. Nor is this absolute, as a small number of studies showed working memory tasks may be unaffected or even improved with increased cortisol (Human et al., 2018).

O'Hare (2003) noted that the ability of an individual pilot to rapidly and efficiently draw from long-term memory may vary with experience. According to O'Hare (2003), "Expert pilots seem to utilize a long-term memory strategy based on the identification of situationally relevant cues. Their performance appears to be more resistant to stress effects" (p. 224). However, when pilots are inundated with the presentation of excessive data, they can equally experience difficulty in assimilating it into useful information, especially if the individual elements that comprise the data are contextually incongruent. They may focus on one cockpit indicator light and miss other relevant cues as occurred in the crash of Eastern Air Lines Flight 401, in which an unsafe gear indication captured the attention of all three flight crewmembers who failed to perceive a slow descent toward ground impact (NTSB, 1973).

As an example of the impact of workload on task performance, Morris and Leung (2006) conducted a study using 37 male and five female students from a Melbourne university. The participants were enrolled in aviation courses and had some previous flying experience, with the average amount of flight time recorded as 134 hours. Using random assignment, the students were placed in one of three groups for comparison purposes. The study was conducted using a joystick to control the "airplane" and a headset-equipped computer that served as the host for the Microsoft Flight Simulator® software. The three groups were assigned to low, medium, or high workload task levels to ascertain the impact of load variation as measured in specified metrics. According to Morris and Leung (2006), a low workload consisted of merely manipulating the flight control inceptors and communicating. A medium workload meant the participants were manipulating the flight control inceptors, performing rule-based tasks, and communicating. Finally, a high workload consisted of manipulating the flight control inceptors, performing rule-based tasks, problem solving/high cognitive demand, and communicating (Morris & Leung, 2006).

The Morris and Leung (2006) study demonstrated the limited capacity of short-term auditory memory in an aviation context. Each group was provided an ATC message to be observed, processed, and repeated back to the researchers. The study found that when the information was provided in one chunk, the comprehension errors observed in all three groups were insignificant. When the amount of information was increased to five chunks, the low and medium workload groups demonstrated comprehension with only minor errors. The error rate increased with additional information chunks, peaking at 86% when presented chain lengths of five and seven chunks. Conversely, the high workload group experienced 11% comprehension errors when presented with only three chunks of air traffic information. When presented with seven chunks of information, the error rate rose to 93% (Morris & Leung, 2006). The work demonstrated how memory applied to assess compliance with stability criteria could be impacted, or even inhibited, when presented with a high-workload environment, such as an instrument approach flown in demanding conditions.

However, it should be noted that Endsley (2015) emphatically stated that SA is exclusively contained within memory. In clarification, working memory is rather viewed as a bottleneck for those in unique situations for which previous modeling lacks. The existence of developed mental models can obviate the limitations of working memory (Endsley, 2015). Endsley (2015) cited studies in which working memory application in SA assessment was investigated. These studies showed that the impact of memory may be more complex than originally thought. In an early study as to whether responses to SA queries during multiple freezes of a simulation varied over time, Endsley found equivalent responses with 5-6 min delays as when questioned immediately after the freeze, suggesting that SA information may be held in other forms of memory and may be a subset of long-term memory (Endsley, 2015). Other studies have indicated that experts may rely more heavily on long-term memory than working memory, that those with lower levels of working memory were just as accurate in their responses to queries as those possessing higher levels, and that working memory abilities did not accurately predict levels of SA for experienced pilots (Endsley, 2015).

**Attention Management**

Attention management is a higher order, cognitive skill, subject to executive control (Tsang & Vidulich, 2003). In the information-rich flight environment, a pilot must determine those data elements considered to be most relevant to the situation and manage the attention rendered between those and competing elements. Attention can be characterized as an attribute supporting acquisition of information from various environment sources (Wickens et al., 2004). Pilot ability to filter information is a critical aspect of human information processing, and the mechanism for doing so is through the application of attention. Under resource theory, attention has been viewed as a limited commodity for which the resources may be allocated along a number of dimensions to include visual and auditory (Wickens et al., 2004). Harrivel et al. (2016) identified attention-related human performance limiting states as present in 13 of 18 international loss of control inflight airplane accidents that occurred between 2001 and 2010. Distraction, in the form of channelized and diverted attention, was present in all 18.

Wickens and Carswell (1997) specified attention as having three different possible states – selective, focused, and divided. Selective attention filters the environment for information to process and on which to focus. Focused attention, or attention capture, addresses the sustained processing of the target information while excluding other potentially distracting or influencing information. Divided attention, then, is the ability to process more than one element of information at any particular time (Wickens & Carswell, 1997). Failure to properly attend all elements of information and continuously vet those most critical to the current task has been identified as causal to a number of controlled flights into terrain events (Wickens et al., 2004).

Flight operations necessitate routine use of selective attention. This attention selection is heavily driven by internal goals and pilot expectancies and operates in serial fashion: perception and cognition for differing tasks cannot be attended in parallel (Vidulich et al., 2010). The selection as to which element to attend is driven by the salience of the information, the efforts required to obtain the information, the expectancy of the information, and the value of it (Wickens et al., 2004). Pilots interpret information that is sensed and perceived using one or both of two often concurrent methods: bottom-up processing by feature analysis and top-down processing drawing from long-term memory. These processes influence where the pilot places attention and what the pilot holds as expectancies (Wickens & Carswell, 1997).

In the context of human information processing, bottom-up processing uses a data-driven focus whereby salient cues capture the attention of the pilot, the cues are subsequently interpreted, a number of options are generated, the chosen option is selected, and actions are taken (Endsley, 2013; Wickens & Carswell, 1997). Top-down processing of information takes a goal-driven approach, recognizing that the goals direct attention, the goals determine the development of SA, and the goals determine selection of a mental model for interpreting the information (Endsley, 2013). It is driven by the expectations of the pilot based on knowledge of the world and the value of the information to the pilot. Wickens (2004) commented that it is the interplay between what the pilot observes to be the state (bottom-up) and what the pilot believes should be there (top-down) that facilities much of the processing of perceived information. Top-down information processing bias may guide the pilot into an incorrect mental model application. This potential supports the importance of including bottom-up processing in the SA process (Wickens, 2003b). Such was the case in the 1977 collision between two Boeing 747s in Tenerife, Spain, where non-standard and garbled ATC directions were misinterpreted by the eager pilot as clearance to depart. The expectation of receiving a clearance overpowered the incomplete information being heard (Wickens, 2004).

Focused attention serves to filter information elements from compelling and sometime competitive sources. Its value in attending to critical information is clear, but it is not without cost as the pilot expends attention resources to the target information, often at the expense of other information elements (Wickens & Hollands, 2000). The transition from focused to selective attention and back incurs cognitive friction. Additive flight instrument display information might pose a solution but must be done with caution, as it can actually detract from the task at hand and imposed unnecessary clutter (Curtis et al., 2010). In the extreme, pilots can fall prey to attention tunneling. Wickens (2005) defined attention tunneling as a state in which (a) attention resources are allocated to a particular source of information, (b) the attention dwell time is longer than optimal, and (c) there is a failure to divide or switch attention resources to other task-relevant sources. This tunneling can affect both spatial and subjective priority (Wickens & Hollands, 2000). Pilots of all experience levels, when presented a modern flight instrument display employing a split-cue flight director (FD) system that uses separate command axis for pitch and roll, will tend to channel their attention toward the task of ensuring the airplane pitch reference is *exactly* placed in the middle of those commands, and do even more so when the external environment results in airplane movements that deviate from the path. In these tunneling instances, the field of regard the pilot attends shrinks dramatically as pilots focus on a minute arc of the display.

Divided attention characterizes pilot ability to balance resource allocation between one information element and other. Managing this division is critical to safe and effective flight operations. To be successful, internal guidelines must exist to specify the need to redirect attention and to which element to redirect it. When two or more tasks are undertaken, the pilot must tier activity in order of priority. Attention placed on an emerging higher prior task can result in a dropped critical element (Vidulich et al., 2010), although with strong attention management and quick attention switching, the process can approach parallel processing (Wickens et al., 2004). Evidence suggests some parallel processing can occur, particularly when the task is a routine closed-loop action and different processing requirements (visual and aural, for example) are experienced (Vidulich et al., 2010). However, interference between the two or more tasks may confound attention management. Such a scenario can occur when airplane flight path management under challenging conditions is interrupted by presentation of a systems non-normal event. This event can capture pilot attention and interfere with precise guidance tracking.

Ziv (2016) conducted a metasynthesis of current literature on pilot gaze patterns and visual attention, finding that gaze behavior appears an important variable in task performance and corresponds closely with visual attention. The conclusions drawn from the review indicated attention can be affected by the phase of flight and airplane maneuvering requirements, and are individuated among study participants (Ziv, 2016). Experts tend to better apply attention management, applying more refined scan patterns with more fixations between saccades and shorter dwell times on a particular data element. Further, the presence of anxiety can result in inefficiencies in attention allocation and increased attention entropy for pilots at all levels of experience (Allsop & Gray, 2014; Janelle, 2002; Vine et al., 2015).

A number of strategies exist to assist the pilot in the allocation of attention. Training on attention management in conjunction with deliberate practice in specific tasks can improve the automaticity of pilot actions and reduce pilot demands (Wickens et al., 2004). Equally, display redesign and system automation can reduce attention resource depletion by providing cues as to when the flight parameters approach an unstable state. Such redesigns should reduce the cost of acquiring the information: provide the stable approach cues in close proximity to reduce the transition time and effort to reallocate focused attention from one display element to another, or from one task to another.

**Gaps in the Literature**

Considerable research has been conducted as to the human factors and user-centered design as they related to the display of information. A general understanding as to how pilots perceive, comprehend, and project within the context of an unstable approach is much less defined. Given the continued incidence of runway excursions owing to unstable approaches, it strikes as surprising that few recent studies have been conducted on the topic of unstable approaches from the perspective of SA development through the use of display enhancements.

The aviation industry has been seeking a viable solution to the problem of unstable approaches. However, such efforts have little empirical research to support their proposed solutions. Shish et al. (2015) conducted a study in which they assessed the inclusion of trajectory prediction and alerting for airplane energy state. Instrument displays were modified to provide information as to airspeed, roll, and vertical speed limitations as calculated by predictive algorithms to extrapolate the current energy state using a number of system inputs. Research participants were exposed to a series of scenarios in which they were given low energy states, high energy states, icing, and a stabilizer system non-normal event. Predictive alerting of an unstable approach caution message was annunciated on the EICAS and provided on the ND and vertical situation display (VSD); however, no indication was provided on the PFD, nor were the stable approach bounds displayed (Shish et al., 2015).

Wang (2016) developed a methodology for both identification and subsequent nowcasting of an unstable approach condition based on radar surveillance track metadata for a particular instrument approach procedure. Analysis demonstrated that at a position 3.5 nm from the runway, 75.3% of unstable approaches could be predicted correctly. Using parameters trained from the prediction models for each approach, an unstable condition could be predicted through real-time source data for potential presentation on the PFD. Although Wang furthered the science of performance prediction and suggested the use of unstable approach probabilistic alerts for airspeed or rate of descent, no assessment of such display enhancements on flight crew SA was considered.

Sembiring, Liu, Koppitz, and Holzapfel (2018) examined unstable approaches from the perspective of general energy management. Their research indicated that various air carrier standard operating procedures used different measures for detection of unstable approaches. They proposed that unstable approaches were the by-product of the inability to manage airplane kinetic and potential energy. In establishing upper and lower bounds for energy excess and insufficiency, they analyzed unstable approaches identified in a survey of 2,000 flights to a particular airport. Of those identified as unstable using a rule-based approach such as that presented by the Flight Safety Foundation, a full 26% were detected using the energy management model. However, no assessment was made as to the validity of the pilots in determining an unstable condition.

Rao and Puranik (2018) focused their research on the GA community, accomplishing a retrospective analysis of historical National Transportation Safety Board (NTSB) accident reports. Culling the reports using a keyword search for variants of the term *stable* and searching on the NTSB code for such events, 24042, there were 205 unstable cases between 1982 and 2017 were identified. A failure to maintain airspeed, ranked as the primary element of the sample unstable approach accidents, was present in 42.4% of the cases. Glidepath, descent rate, and roll angle were frequently present at 28.8%, 17.6%, and 15.6%, respectively. Pitch attitude, correlated with airspeed and descent rate, was present in 14.6% of the cases. Of critical importance to the generalizability of the proposed research, they found “Our results show certain similarities between commercial and GA operations…some of the top causes are similar to (if not the same as) the key elements to a stabilized approach in commercial operations” (Rao & Puranik, 2018, p. 7). As they point out, all the causes identified can be considered "triggers" to an impending or current hazardous state. The study made appropriate assessment of the data, yet it did not delve into the details of each particular report, avoiding the question of pilot SA; equally, it limited its examination to those cases in which an accident occurred, ignoring that a large number of unstable approaches result not in an accident, but in an incident or no reporting at all.

A systems engineering approach was used by Moriarty and Jarvis (2014) to examine the results of semi-structured interviews of 25 pilots from a selected air carrier on the topic of approach speed and airplane configuration. Using grounded theory, data were collected on questions pertaining to instrumentation used in the flight deck to prompt airplane configuration changes, techniques used to predict instability when a deviation from the company configuration profile occurs, experiential factors contributory to an unstable approach, the impact of ATC direction, and any suggestions for safety improvements. The data were coded and a qualitative analysis using word analysis was conducted. Most telling in the study was a particular discussion statement:

> The continuation of an unstable approach to a landing is a more significant system failure than an unstable approach that ends in a go-around. In this case, the limits that should have prevented an unstable approach or should have directed the pilot perform a go-around have failed. (Moriarty & Jarvis, 2014, p. 201)

They postulate that plan continuation bias may be suspect in these cases, where pilots elect to continue the approach in an unstable speed or configuration condition and determine the ability to successfully complete the approach by the point of minimum stabilization height. However, pilots must balance confliction between their inherent goal-oriented approach to task and the demands of ATC controllers. Distinctive was the application of a systems approach to the unstable approach condition and the identification of real-life issues that impact the pilot, as the operator of that system (Moriarty & Jarvis, 2014).

Despite the development and application of rules-based criteria for pilots to be applied during flight operations and the extant research addressing the root factors, displays, and energy aspects of the issue, undesired unstable approaches continue to occur. Accordingly, a number of these unstable approaches have contributed to runway excursions. Within the academic body, attention must be drawn to the need for better understanding of the SA element of this flight operations issue. The general lack of research pertaining to the application of display enhancements to support pilot situation awareness in the face of an unstable approach is a striking commentary. This is especially so given that pilot failure to maintain a stable approach and the subsequent risks to runway excursions have been in existence for a considerable period, coupled with the aviation community's affinity for quickly adapting advancements in technologies, where doing so is merited.

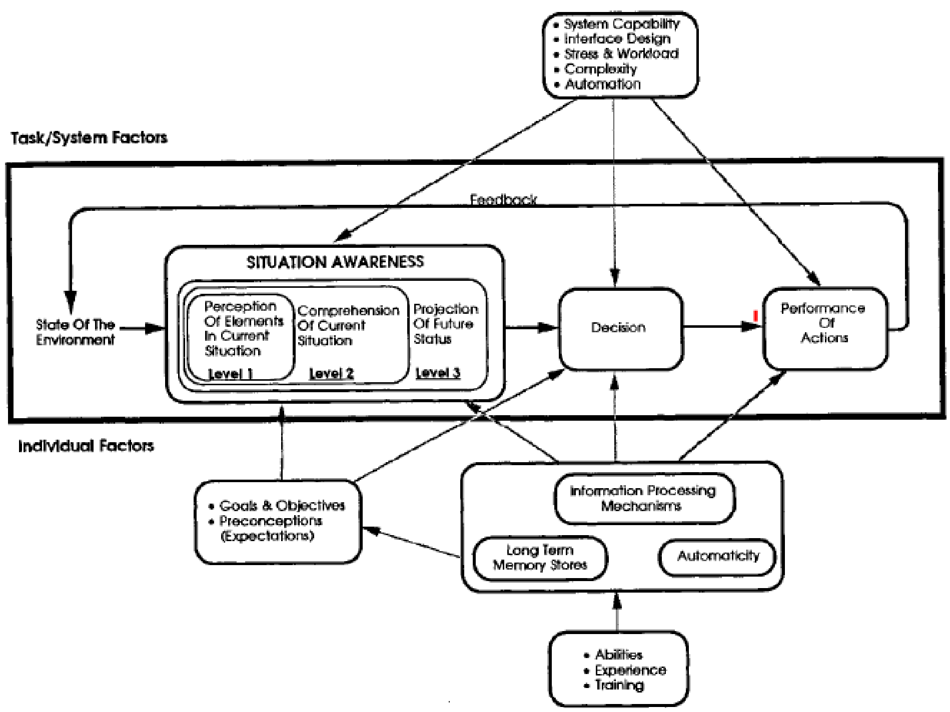
**Theoretical Framework**

When pilots execute specific flight maneuvers and procedures, they are exposed to a number of task actions and airplane system issues that impact not only the development of the various levels of SA, but also the actual decision-making step and the performance of the actions to be completed as a consequence of that decision. In a ground-breaking work, Endsley (1995b) proposed a dynamic feedback model, underpinned in human information processing constructs, to explain how individuated SA fits within the broader context of dynamic decision making. This model consists of three components: SA, the decision, and the performance of actions, all as moderated by task/system and individual factors. This model is presented in Figure 4.

According to the model, pilot perception of the environmental elements, as drawn from sensory input from flight displays or indigenous sensors, are foundational to pilot SA. Pilot SA serves as a major contributor to the underlying basis for decision making, as affected by individual goals, objectives, and expectations. Pilot SA may, in fact, influence the process of decision making itself (Endsley, 1995b). Decisions then drive actions which, when executed, result in a given level of performance. This performance outcome then serves not only as a stand-alone outcome, but also as a feedback loop to reinitiate the process in a changed state of the environment.

**Figure 4**

*A Structural Model for Information Processing*




*Note.* Excerpted from "Toward a Theory of Situation Awareness in Dynamic Systems," by M. R. Endsley, 1995, *Human Factors* 37(1), pp. 32-63. Copyright 1995 by Human Factors Journal.

Task/systems factors such as system capability, interface design, stress and workload, system complexity, and the application of automation can influence all three components of the model. Individual factors, in contrast, address not only the individual goals and objectives but also information processing, long term memory, and automaticity of response. Again, each of these individual factors impacts all three components of the dynamic decision-making process. Endsley (1995b) noted that information processing, long term memory, and automaticity variation among pilots can account for variation in the ability of individuals to acquire SA from a given environment. Attention, memory, and experience play heavily in the ability of the pilot to develop a high state of SA. The types of flight instrument display enhancements proposed – the application of boundaries or alerts – may improve pilot SA in comparison to the current presentation of flight parameters. Further, the use of duplicate cueing modalities would afford the pilot even greater opportunity to develop the necessary SA to address an unstable approach condition.

***Situation Awareness***

Endsley (1990) defined SA as “The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of the status in the future” (p. 1-3). As such, the concept identified three main elements: perception (Level 1), comprehension (Level 2), and projection (Level 3) (Endsley, 1995b). The combination of these elements into an SA “picture” is integral in the safe execution of flight operations. Endsley (2015) clarified that although the naming convention of each level might imply as such, the SA model is not linear; holding Level 3 SA does not necessitate prerequisite Levels 1 and 2 SA in linear, discrete steps. They are ascending levels of SA, by definition, but Level 3 SA can inform the other two levels (Endsley, 1995b). The model is as much driven by forward-facing progression through the levels as it is goal-driven processing to inform the data needed (Endsley, 2015). As such, “This top-down processing will operate in tandem with the bottom-up processing in which salient cues will activate appropriate goals and models” (Endsley, 1995b). Endsley was quick to point out the differentiation between the process of SA development and the state of SA knowledge, noting that situation *assessment* are the processes used in deriving situation *awareness*, and the two should not be confused.

Acknowledging the existence of limitations to the individual approach to SA, a number of assumptions, when accepted, inform the researcher as to the nature of this approach, in that “(1) it is a cognitive phenomenon residing in the heads of human operators; (2) there is a ground-truth available to be known; and (3) good SA can be derived from reference to expert or normative performance standards” (Stanton et al., 2017, p. 455).

As Vidulich (2003) postulated, SA serves as the substrate connecting the external environment, the world, to the decision making and action processes. It serves a systematic approach to ingesting a chaotic, rich flux of otherwise confounding data, making sense of it, and then using it to forecast the most likely future state. This projection is then incorporated into the decision-making process, which inevitably leads to an action on the part of the pilot.

**Level 1 Situation Awareness.** Endsley and Jones (2012) noted that Level 1 SA, perception, constitutes the greatest problem area for SA in the aviation domain. In an earlier study, they identified that failure to perceive needed data was evident in 76% of pilot SA errors; further, the majority of those cases were due to either failure to detect the desired data or the lack of data presentation (Jones & Endsley, 1996). Pilots must invoke the use of visual, auditory, and tactile senses to collect the requisite data. Even as the capture of such data might seem rather straight-forward, perception can be impacted by numerous competing environmental data inputs. In the flight domain, the presence of airplane buffet, turbulence, air traffic communications, precipitation on the windscreen, intra-crew verbal and non-verbal communication, fluctuating ambient light conditions, and other factors present the pilot with a considerable challenge in staving off undesired attention narrowing. As much as pilots seek to assiduously place their attention on the needed flight instrument display elements, inevitably failures occur. Of particular note, Vidulich (2003) suggested that SA is more concerned with the quality of the data perceived by the pilot, and not directly with the attention resource demands placed in the comprehension of that data.

Endsley (1995a) developed a taxonomy of general factors that impact SA at various levels. Lack of data availability, data being difficult to detect and/or perceive, or a basic failure to scan or observe the data presented – either through omission, attention narrowing or distraction, or a high task workload – can be found causal to an error in SA at Level 1 (Endsley, 1995a). For example, a pilot might visually sense an airspeed deviation from the commanded value due to turbulent wind conditions but become so channelized in that element that the roll angle of the airplane is allowed to deviate significantly from a nominal wings-level state. Misperception of the data being presented may also prove causal, as would a memory failure (Endsley, 1995a). Such a case might manifest when the pilot, noting an increased rate of descent during the approach, is suddenly distracted by an EICAS caution or warning, and fails to attend to the deteriorating approach condition.

**Level 2 Situation Awareness.** In obtaining Level 2 SA, comprehension, pilots attempt to make sense of the data received in the context of the task being performed. The rather disjointed data obtained in reaching Level 1 SA must be processed to garner something of use to the pilot. As Endsley and Jones (2012) stated, “It involves integrating many pieces of data to form *information* [emphasis added], and prioritizing that combined information’s importance and meaning as it relates to achieving the present goals” (p. 17). In conducting an instrument approach in turbulent conditions, the pilot notices the airspeed rapidly decreasing below the desired approach reference speed. The pilot must evaluate the degree to which the airspeed has decreased and the rate at which it is occurring; the proximity of the current airspeed to that of an unstable approach condition; and the time remaining before landing to correct the deficiency before comprehending the picture presented by the data. Level 2 SA provides meaning, and 19% of SA errors have occurred owing to problems in this element (Jones & Endsley, 1996).

Jones and Endsley (2000) emphasized the significance representational errors play in the development of Level 2 SA. Such errors occur when the information presented to the pilot is correctly perceived, reaching a Level 1 state, but the significance of that information is improperly understood. They introduce the nomenclature *schema-bizarre* and *schema-irrelevant* information, where the information obtained is inconsistent with the schema being employed, and thus stands out, or is unimportant to the schema construct. A study of ATC controllers, in which misidentification of airplane type, flight path errors, and communication errors were employed, concluded that cues that were schema-bizarre had higher probability of detection than those considered schema-irrelevant. Individuals may also be equally alert to cue expectations that go unfulfilled as to the presentation of cues that are unexpected (Jones & Endsley, 2000). In a flight scenario, pilots have expectations as to the manner in which an approach will unfold. When challenged by outlandish cue presentation against the expectation, a higher probability exists the cue will be perceived and identified. Thus, a sudden roll angle or airspeed change during an expected calm-day approach to a runway will stand out. More so, the same scenario executed on a turbulent day may cause as much alert when the expected variations in pitch, roll, or airspeed do not occur.

In Endsley's taxonomy (1995a), Level 2 SA error causal factors weigh heavily on the application of mental models. Examples of such error impacts would include a poor mental model or the complete lack of a model, application of an inappropriate mental model for the situation, overreliance on default values in the mental models, or, as with Level 1 SA, the presence of memory failure. In the approach scenario, a pilot might fail to comprehend the presence of an incorrect landing configuration when transitioning from one airplane to another: one in which a reduced flap setting is allowed and one in which it is not. The resulting reduced aerodynamic drag would render the expected thrust or power settings for the approach ineffectual, increasing pilot workload to maintain the commanded airspeed for a stabilized approach condition.

**Level 3 Situation Awareness.** In Level 3, projection, the pilot uses the perceived and meaningful data to predict their ramifications in the proximate future. Using the previous example of the instrument approach, Level 3 SA would be present when the pilot is able to project that the rate of the decreasing airspeed will lead to an unstable approach and, if uncorrected, could result in stalling the airplane at low altitude, experiencing a hard landing on touchdown, or experiencing a runway excursion. Endsley and Jones (2012) proffered a failure of the pilot to develop an accurate Level 3 SA projection may result from one of two possible causes: insufficient mental resources to process the information or insufficient knowledge of the aviation domain. Mental resource levels will vary among individuals as will individual pilot levels of knowledge within the aviation domain. For example, a senior air carrier captain will likely enjoy greater breadth and depth of aviation knowledge than a newly-minted instrument-rated pilot. Of the totality of SA problems, a mere 6% are within Level 3, which may reflect not an ease of developing such high-level SA, but rather the degree of difficulty in obtaining Level 1 or Level 2 SA in the aviation domain (Jones & Endsley, 1996). Failure to obtain the lower levels of SA may preclude achieving the projection of Level 3. Thus, any system that would aid in the perception and comprehension of data will further progress toward projection, a useful condition in averting safety risks.

Humans are subject to a limited amount of attention and cognitive resources. Studies have shown that while the approach phase may constitute a relatively small segment of the overall flight profile, considerable mental and physical demands are placed on the pilot during that time:

> The final approach and landing constitute only about 2% of the average total flight time, yet almost 50% of all aviation accidents and incidents occur during this phase. The reasons for this are many, but the most significant ones are that both the pilot workload and pilot fatigue are at the highest level, coinciding with the narrowest margin of safety. Operating close to the ground in a fast-moving

aircraft is the most hazardous part of the flight and requires a maximum of the pilot's skills. (Daidzic & Shrestha, 2008, p. 2131)

Sumwalt et al. (2015) examined 110 flight path deviation cases submitted between December 1973 and July 2013 to the NASA ASRS. Using a unique coding and collection form, their research highlighted the nature of such events. Of significance was the finding that 20% of the cases occurred during the final approach phase. The predominate activities being undertaken at the time included radio communications, traffic or ground reference point acquisition, and dealing with abnormal conditions. Contributing factors included distractions, fatigue, high workload, and complacency. A key revelation from their study was that in 104 of the 110 reports, "the primary means of detection was someone or something other than the flight crew," meaning each of the pilots failed to detect a compromise of the flight path adherence (2015, p. 11). Of the 110 cases, 25 were classified as an accident, meaning either a loss of life or serious injury, airplane substantial damage, or both during the period of embarkation for flight through disembark of the airplane. In these cases, 68% were in the approach phase, and airspeed, descent rate, and roll angle deviations were present (Sumwalt et al., 2015).

Pilot workload can be greatly reduced, SA enhanced, and safety proportionately increased through proper presentation of critical data on the flight instrument displays. During the approach phase, the crew must make multiple changes to the airplane configuration, flight trajectory, and energy state, as well as attend to the presence of hazards such as weather and conflicting traffic. These operational demands present a continuous challenge in the development of SA.

For Level 3, Endsley (1995a) lists two items, a poor mental model or a lack thereof as causal factors to an SA Error. Recurrent training for most air carrier pilots is conducted in a flight simulation training device (FSTD) and not the actual airplane. Such training is often, although not always, accomplished at a training-relevant airplane gross weight reflective of a particular load of passengers and fuel. Too often, insufficient variation of that weight during training sessions leads to a mental model as to the deceleration rate, pitch attitude, and thrust setting for the approach. Pilots experiencing an early return to the airport for critical emergency may have insufficient time to reduce their landing weight. Application of a model meant for one condition, a training weight, might result in the pilot predicting inaccurate anthropometric and proprioceptive cues for flight control and thrust inceptor placement to achieve the desired airplane performance.

**Summary**

Industry and academia provide pilots guidelines on airplane configurations and energy state conditions that are indicative of a stable approach. These guidelines serve as a useful tool but are dependent on the pilot obtaining a level of SA that will support appropriate decision making to prevent unintended runway excursions, upsets, or controlled flight into terrain. Indeed, the information processing phases of sensation of the stimulus, perception, and the SA phases of cognition and prediction all affect the pilot's ability to make a timely and accurate decision for the scenario in which the airplane is found. Development of a suitable display enhancement should look to the principles of Human-Centered Design to achieve optimum benefit. Accordingly, the objective of this research study was to determine whether pilot SA, when engaged in an unstable approach, was affected by applying enhancements to the flight instrument displays.

# Chapter III: Methodology

This chapter introduces the research methodology by explaining the experimental design, defining the population and sample selection process, discussing the data collection instrument, and finally, the data analysis process. It establishes that a within-group, repeated-measures, quantitative design was the appropriate method to inform changes in a participant's performance. Purposive convenience sampling was used to select participants from the target population of qualified pilots within the United States.

A pilot study was conducted to gather early and valuable insights into the relationships between the IVs and DV and validate the overall experiment design and delivery. Study participants were asked to observe a series of video vignettes showing the PFD and ND displays during an instrument approach. Data were collected using think-aloud protocols captured through video and audio recording. The data were assessed to determine the RA from real-time SA responses. A two-way, within-group, repeated-measures Analysis of Variance (ANOVA) tested whether pilot SA differs between the control and various experimental enhanced display treatments.

## Research Method Selection

A within-group, repeated-measures, quantitative experiment (Creswell, 2014) was used to gather the necessary data to investigate the research question and hypotheses. Vogt et al. (2012) define a repeated-measures design as one in which participants are measured two or more times on the DV. The within-group approach allowed participants to serve as their own control, with each participant being exposed to the same combinations of the IV (Vogt et al., 2012). As such, both systematic variation from unintended experimenter actions and unsystematic variation due to random factors affecting the experiment were controlled (Field, 2013). Field (2013) notes, "When we look at the effect of our experimental manipulation, it is always against a background of 'noise' caused by random, uncontrollable differences in our conditions. In a repeated-measures design this 'noise' is kept to a minimum and so the effect of the experiment is more likely to show up" (p. 17).

The focus of the study was placed not on the overall individual participant performance, but rather the differences in each participant's performance resulting from each treatment. The question – whether the application of display enhancements affected pilot SA – was thus assessed for each participant. The within-group, repeated-measures, quantitative experiment was structured as depicted in Figure 5.

**Figure 5**

*Single Group Series Design*

| |
|---|
| Participant X(A):O($A_{1-n}$)__X(B): O($B_{1-n}$)__X(C): O($C_{1-n}$) …X(L): O($L_{1-n}$) |

*Note*. Series shown using sequential presentation of vignettes for the first participant. Subsequent vignette presentation to participants was randomly delivered.

The "X" denotes participant exposure to a specific treatment to be measured. The effect on SA was measured immediately during unstable approach conditions that arose during presentation of the treatment and is indicated by the "O" immediately following the treatment. The number of unstable conditions varied with each vignette. As such, within any particular vignette, there were one or more measures. The study design did not require a pretest or post-test data collection, as the analysis compared disparate display treatments within the presentation of each successive vignette (Creswell, 2014).

A cursory review of previous studies on the subject of stabilized approaches in the context of flight instrument display enhancements has shown to be unsupportive in terms of archival data. Although survey instruments may capture pilot attitudes concerning display proposals, application outside the laboratory environment would be limited to perceptions of acceptance and non-inclusive of exposure to the technology itself. The data collected provided empirical evaluation as to the value of the proposed enhanced flight displays to pilot SA.

## Population/Sample

The experiment data were obtained using volunteer participants. Participants were drawn from a reduced sample frame by use of purposive convenience sampling. Each was deliberately chosen to facilitate a representative sample of flight experience and background. Participants were selected from a target population of flight training programs, corporate flight operations, and air carriers.

### *Population and Sampling Frame*

The research population comprised all pilots currently holding FAA-issued certificates or ratings allowing operation under IMC, as they had the potential to encounter conditions conducive to an unstable approach. Instrument rated pilots are permitted to fly in IMC when an instrument flight plan is submitted and granted. Participants holding an instrument rating would have a greater likelihood of experience with current flight instrument displays incorporating a PFD display presentation.

Participants were required to hold a number of airplane qualifications and experiences, have recent flight experience and exposure to electronic displays, and meet key vision and hearing standards. The specific criteria to be met by each participant included the following:

- Hold an instrument qualification, as demonstrated through possession of either a Private or Commercial certificate with an Instrument rating or an FAA-issued Airline Transport Pilot (ATP) certificate.
- Hold an airplane single- or multi-engine land category and class rating.
- Possess instrument currency within the past five years.
- Have experience with current electronic flight instrument displays.
- Demonstrate adequate visual and auditory acuity through possession of a current FAA medical certificate, or a current driver license with demonstrated responsiveness to visual and aural cues.
- Be of 18 years of age or older.

The determination of instrument currency was established using the requirements of Title 14 Code of Federal Regulations (CFR) Part 61. This, along with the other listed criteria, ensured the pilot held the prerequisite level of instrument aeronautical skills to understand the research and complete the trials, and possessed the visual and auditory capabilities necessary to successfully operate an airplane. It was assumed the highly-regulated and standardized certification processes for obtaining such qualifications were properly executed.

Multistage sampling (Creswell, 2014) was undertaken so as to ensure a participant homogeneity and maintain a balance of pilot experience based on the chosen measure, total flight time. Drawing from various regions and operational domains, the influence of any one region or type of operation was mitigated for factors such as operating environment or equipment flown. Thus, participants were randomly selected from a reduced sampling frame comprised of universities, training centers, and corporate and air carrier volunteer pools.

Participants with lower-level flight experience were drawn from the flight training operation at the Prescott, Arizona, Embry-Riddle Aeronautical University (ERAU) campus. Individuals responded to notices placed at key campus locations and messages sent directly to students. Boeing-operated 14 CFR Part 142 training centers at Miami, Florida, and Seattle, Washington, provided access to more experienced pilots to allow the capture of highly-experienced air carrier pilot samples. Pilots were provided notice of the study, and volunteers were selected from those who expressed interest. Corporate pilots situated at Centennial Airport, Denver, Colorado, provided further access to corporate pilots of varying flight experience levels.

***Sample Size***

G*Power software (Faul et al., 2014) was used to determine the minimum sample size for the study. The ANOVA repeated measures, within-factors formula was applied. Sample size was determined based on an a priori, between-subject, experiment construct using ANOVA main effects and interactions, with a medium effect size (f) = .25, an error probability (α) = .05, a power (1-β) = .80, a repetitive measures correlation of 0, a nonsphericity correction (ε) = .75, with four treatment groups and 20 measures. Using these input parameters, G*Power software version 3.1.9.2 recommended an a priori calculated minimum sample size of approximately 24 total participants to provide sufficient power (Faul et al., 2007).

Hair et al. (2010) support this assessment. They note that maintaining adequate statistical power is essential, establishing a minimum acceptable 1-β in the .80 range. However, they emphasize the need to balance strict values of α against setting criteria that would prevent differences from being found. Field (2013) discusses the importance of f in being able to identify mean differences that may exist between each group against the observed variance in values, in effect detecting the signal from the noise. The goal is to not overpower the statistical test such that differences cannot be found yet control the likelihood of Type I or Type II errors occurring. Cohen (1992) establishes f values as a function of the test index to be applied but cautions that such benchmarks must be considered in the context of the instruments being used to collect the data and the nature of the effect in the environment (Field, 2014). Cohen (1992) notes when using one-way ANOVA testing these values would range from .10 to .40 for small to large, respectively, but does not address interactions that may occur in factorial designs. The goal is to measure effects that are of such significance that display designers would opt to consider the proposed enhancements. Previous research by Brams et al. (2018) and Stelzer and Wickens (2006) used or expected a large effect size when examining questions pertaining to effective gaze behavior in complex error-detection cockpit tasks and pilot strategic compensation for display enlargements in surveillance and flight control tasks. However, Thirtyacre (2021) applied a medium value for f in a study of remotely piloted aircraft command and control latency during within-visual-range air-to-air combat. Harbour (2015) used a similar value when evaluating predictors of situation awareness and display usability with United States Air Force pilots while performing complex tasks using both head up and head down displays. Lovakov and Agadullina (2021) indicate where thematic subgroups are not covered in their study on empirically derived guidelines for effect size in social psychology, researchers can use the median effect size. Hence, a medium value for f was chosen. Analysis using G*Power shows the effects of varying the values of f and 1-β on the sample size. The G*Power calculation basis and sample size analysis outputs are presented at Appendix B.

***Sampling Strategy***

Notification of the study was sent electronically to targeted industry and academic cluster groups explaining its general purpose and the demographics of participants sought. Directions on how to contact the researcher for participation were provided. Pilots wishing to participate were initially vetted to ensure they met eligibility requirements.

Vogt et al. (2012) remark that when the research question is mostly focused on the variables and their interrelationship to each other, generalization to a broader population becomes less significant. They introduce the concept of purposive sampling:

> A purposive sample is one in which the researchers choose participants deliberately (with a purpose in mind), usually in order to make the sample more representative. Purposive sampling is typically done when random statistical methods would be difficult or impossible. (Vogt et al., 2012, p. 348)

Given the limited study resources and the experimental nature of the research, the approach to sampling frame reduction while still meeting the randomized, controlled trial requirements included the pre-treatment identification of participant selection variables based on a desired breadth of characteristics (Vogt et al., 2014). In this case, clusters were created using purposive sampling to meet variable desires (Vogt et al., 2012) and individuals were randomly assigned to the study.

There were more volunteers to participate in the study than were required to meet the chosen statistical parameters. The key parameter was flight experience. Previous research by Ho et al. (2016) examined the concept of *experience* as it pertains to the flight domain. Their research concluded flight hours to be a common metric in determining varying levels of experience. Initially, participants were randomly selected from the pool and designated as holding either low or high experience level on the basis of a total flight time. Criteria were set at less than 1,500 hours (hrs) and greater than 1,500 hrs total flight time, respectively. The participants were subsequently stratified and clustered accordingly (Creswell, 2014). Sampling bias was accordingly avoided.

## Data Collection Process

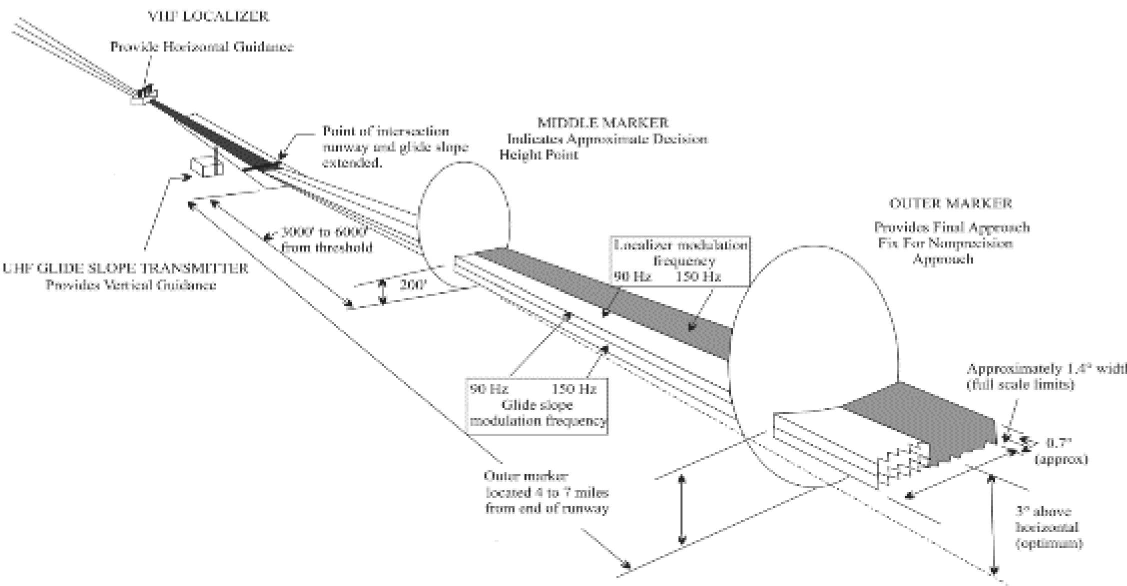
### *Design and Procedures*

Participants were asked to observe a series of ILS approach video vignettes of the PFD and ND displays with primary emphasis placed on the PFD presentation. The vignettes were designed to highlight a particular element of an unstable approach to be observed by the participant. Each participant was assessed as to their level of SA during unstable events encountered in each trial. The vignette, as opposed to static presentation, was chosen for two reasons. First, SA development occurs over time. Pilots are rarely placed in a position where they lack awareness as to how they arrived in the situation in which they find themselves. A factor in understanding the current state and predicting the future state is the data culled from the environment immediately preceding the determination or decision point. A single presentation alone allows for multiple interpretations that could be acceptable. Second, making an SA comparison when assessing disparate displays is possible when the format of each is materially different and subject to meticulous inspection. Baker (2017) made use of single frame captures when studying SA on airplane energy states as indicated by standard instrument displays and the Oz display. In the Baker study, considerable variation between the displays was present. For the current study, the core presentation remains consistent with only the enhancements applied.

Figure 6 shows the nominal design of an FAA ILS final approach; particular note should be taken of the presence of the localizer and glideslope features, which provide the electronic signal allowing adherence to a defined path through the sky (FAA, 2017d). The expanse of runway excursion causal factors was constrained to unstable approaches in order to limit the scope of the design study. Consequently, the segment prior to the instrument approach final approach fix, which marks the beginning of the final descent to runway, was deemed outside the scope. Equally, actions following the initiation of the airplane flare maneuver, nominally accomplished 50 ft above the surface at the approach end of the runway, were not considered. Experiencing a stable approach so as to arrive at the point of flare initiation (a) within the appropriate landing zone, and (b) devoid of excess energy or roll angles, was the targeted task. Thus, data collection focused on pilot cognitive goals during the flight period bracketed by those action points.

**Figure 6**

*FAA ILS Final Approach Nominal Design*

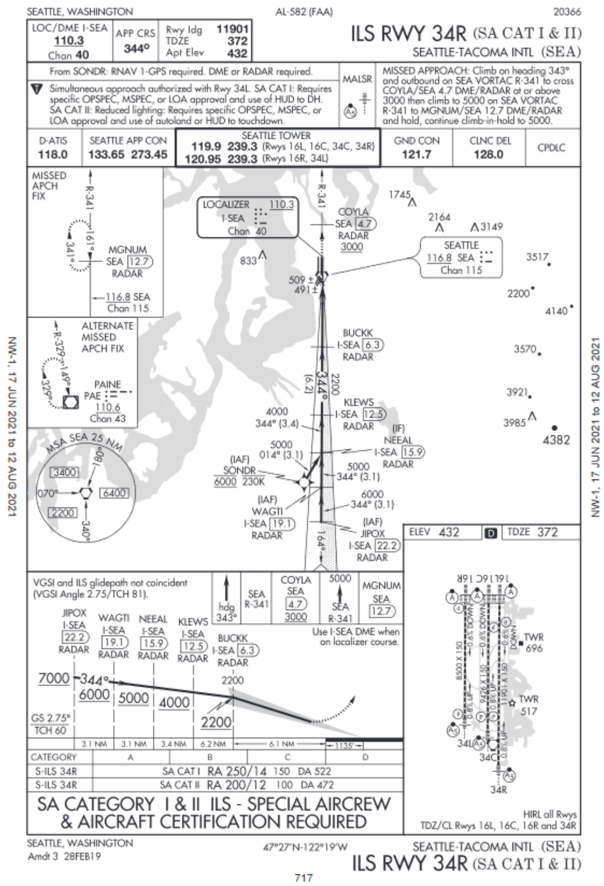



*Note.* Excerpted from "Aeronautical Information Manual," by Department of Transportation, October 12, 2017, *Aeronautical Information Manual*, p. 1-1-14. Copyright 2017 by Department of Transportation.

All vignettes were constructed depicting an approach to a single airport, Seattle-Tacoma International Airport (International Civil Aviation Organization airport code KSEA). This airport facility was chosen due to the simplicity of both the Terminal Instrument Procedures design and National Aeronautical Charting Office terminal procedure chart presentation. This facilitated quick chart study and review during the pre-trial and within-trial phases. Participants were afforded a reference copy of the chart during the trial period. The KSEA ILS 34R terminal chart (FAA, 2021b) is presented in Figure 7.

**Figure 7**

*KSEA ILS 34R Terminal Chart*




*Note.* Excerpted from "ILS RWY 34R, Seattle-Tacoma INTL (SEA)" by Federal Aviation Administration, June 17, 2021, *U.S. Terminal Procedures, Northwest (NW) Vol 1 of 1*, p. 717. Copyright 2021 by Federal Aviation Administration, Aeronautical Information Services.

Using an iterative approach, subject matter expert (SME) recommendations and ASRS report analysis provided potential profiles and timelines for the content of the vignettes. The SMEs were drawn from a cadre of individuals possessing high flight experience and safety background qualifications. The output of their effort aided in the credibility of the vignette content and minimized the risk of the vignette instrument failing to present the intended effect. To ensure independence of the expert feedback, the

SME identities were withheld from each other, as were the recommendations provided by them. As the research was limited to examining airspeed, vertical speed, roll angle, vertical path, and lateral path stable approach bounds, the scripting focused on those elements.

Participants were afforded training on the individual display elements of the PFD and exposed to an orientation vignette to allow accommodation to the testing instrument and the delivery process. Participants observed the interaction between the display elements, experienced the pacing associated with an airplane operating in the particular performance regime, and became familiarized with the SA inquiry method. The stable approach bounds applied to the experiment and the expected pilot response were discussed. Participants were queried as to the level of comfort with the airspeed, roll, vertical speed, and vertical and lateral path values used, and the resulting questions answered by the researcher. To ensure training consistency and ease of replication, the researcher used a scripted presentation, shown at Appendix H.

Wickens et al. (2004) caution that task saturation can occur as early as completion of five or six trials. This value depends on the type and duration of trial to be conducted. Previous flight-related research efforts have successfully employed constructs involving greater numbers of trials (Baker, 2017; Harbour, 2015; Thirtyacre, 2021). Covelli, Rolland, Proctor, Kincaid, and Hancock (2010) limited helicopter flight simulation time to approximately 30 min when examining field of view effects on pilot performance in flight to avoid undesired pilot fatigue. The core trial profiles for this study are presented in Table 3. These core profiles were used for the control and subsequently modified to overlay each treatment. This ensured each core profile was presented to the participant four times, once in a baseline control presentation and three more times using the study treatments. Thus, participants were presented 12 trial vignettes, each following an orientation vignette and optional practice vignette. Scoring for each core profile was the same, regardless of the treatment applied. Vignette duration was approximately 150 s inclusive of a 7 s title presentation. The total duration for all trials was approximately 40 m, depending on individual actions and readiness to continue. The total participant involvement in the study did not exceed 70 m.

**Table 3**

*Situation Awareness Trial Profiles*

| Trial | Exceedance Type/ Number of Presentations | | Description |
|---|---|---|---|
| | 1,000-500 | 500 - End | |
| Orientation | None | RAE/1 | Baseline event for participant familiarization. Presentation of an approach using the existing airplane PFD and ND displays to orient the participant to the information provided, the movement of the display scales, the timing of the sequence, and any applied treatment. |
| A/D/G/J | VSE/2<br>RAE/3 | AE/1<br>VSE/1<br>RAE/2<br>PE/1 | The roll angle oscillates to a point at which it exceeds the roll angle stable approach bound and is trending further away from the bound at vignette conclusion. |
| B/E/H/K | RAE/2<br>VSE/1 | VSE/1 | The roll angle oscillates to a point at which it exceeds the roll angle stable approach bound but is trending back toward the bound at vignette conclusion. |
| C/F/I/L | VSE/1 | VSE/1<br>PE/1 | The vertical speed decreases from the target value to a condition in which it exceeds the stable approach bound and is trending toward the bound at vignette conclusion. |

*Note*. Exceedance type/number of presentation measures in feet radio altitude. Exceedance conditions were defined as follows: (AE) airspeed excess, (RAE) roll angle excess, (VSE) vertical speed excess, and (PE) path excess. All approaches were initially presented depicting the airplane at airspeed and vertical rate of descent appropriate for common landing weights and using a normal landing configuration.

The chosen trials ensured a representative sampling of at least one of the bounds and reduced the potential for bias resulting from repetitive presentation to the participant. Vignette presentation used a Latin Squares counterbalance application (Anderson & McClean, 2018) to negate the order-of-treatment effects (Vogt et al., 2012). Failure to account for learning effect could affect the participant response behavior and potentially introduce undesired false alarms. A true counterbalance can become rather unwieldy in design due to the number of permutations involved. Latin Squares, an incomplete counterbalance method, avoided undesired complexity and provided an equal opportunity for each measure to be observed, but in a different and unique order to the prior observer.

Blocking was avoided as collecting the vignettes into treatment blocks, and randomizing those blocks, raised concerns with participant familiarization, especially with the more dramatic profiles. Consequently, a 12x12 Latin Square as shown in Figure 8 was employed. As the number of measurements and the sample size did not match, adjustments were necessary to accommodate additional participants. Each vignette was labeled using a Latin designator by block, labelled A through L, in the following order: Treatment 1, Treatment 2, Treatment 3, and finally Treatment 4. The trial delivery orders were repeated for those participants in excess of the 24. This repetition was viewed to have minor effect of the study.

**Figure 8**

*Latin Square Table*

| Participant | Vignette Order | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 and 13 | A | B | L | C | K | D | J | E | I | F | H | G |
| 2 and 14 | B | C | A | D | L | E | K | F | J | G | I | H |
| 3 and 15 | C | D | B | E | A | F | L | G | K | H | J | I |
| 4 and 16 | D | E | C | F | B | G | A | H | L | I | K | J |
| 5 and 17 | E | F | D | G | C | H | B | I | A | J | L | K |
| 6 and 18 | F | G | E | H | D | I | C | J | B | K | A | L |
| 7 and 19 | G | H | F | I | E | J | D | K | C | L | B | A |
| 8 and 20 | H | I | G | J | F | K | E | L | D | A | C | B |
| 9 and 21 | I | J | H | K | G | L | F | A | E | B | D | C |
| 10 and 22 | J | K | I | L | H | A | G | B | F | C | E | D |
| 11 and 23 | K | L | J | A | I | B | H | C | G | D | F | E |
| 12 and 24 | L | A | K | B | J | C | I | D | H | E | G | F |

*Note*. Vignette order of assignment by participant. Adapted from "Balanced Latin Square Generator" by Damien Masson, 2023 (https://cs.uwaterloo.ca/~dmasson/tools/latin_square/).

All participants were presented an orientation trial to familiarize with the presentation of the PFD and ND and then, provided an additional practice session was not desired, were run through the series of randomly sequenced scored trials. Profile geolocation initiated at a 0.5-1.0 nautical mile (nm) straight-in segment prior to the glideslope intercept point at the BUCKK waypoint, as would be experienced when being provided vectors to the final approach course by ATC in accordance with established regulatory guidance (FAA, 2017c). This allowed a 6.6 nm run distance from the runway threshold. Initial altitude was set to 2,500 ft, slightly above the glideslope intercept altitude depicted for the KSEA ILS RWY 34R approach. This allowed participants sufficient time to segue into the vignette.

A B-737-800 series flight simulation was used in creating the vignettes. Airplane fuel loads, passenger and crew weights and placement, and cargo compartment loads were frozen across the vignettes to allow consistency of the commanded airspeed and the resulting nominal pitch attitudes and thrust settings. Key fuel and payload values were

127,553 lbs gross weight and center of gravity set 25% aft of the forward limit. The physical configuration was gear extended with the flaps set at 30 units. This established a targeted pitch attitude of 0.9° and 57% N1 engine thrust setting. Model winds for all vignettes were set at light, variable values. Approach minimums were set at 250 ft radio altitude (RA) to meet the terminal chart value for the approach.

All profiles were developed with a reference speed of 140 kts and the command airspeed set at 145 kts. This value was representative of moderate performance turbojet airplanes and supported by the modeling software used for the experiment. In conjunction with the 2.75° glideslope of the KSEA ILS RWY 34R approach, this airspeed resulted in a targeted vertical speed of approximately 700 fpm in the light wind condition. At the nominal targeted vertical speed, the time from the BUCKK waypoint and glideslope intercept to 1,000 ft AGL and from 1,000 ft AGL to a nominal 250 ft AGL decision height were 73 and 54 s, respectively. This window reflected the timeliness of pilot decision-making in the approach environment. Participants could follow their vertical and longitudinal progress toward the runway using the PFD. The ND scale was set at 10 nm throughout the final approach segment to provide optimum resolution.

Participants were directed to assess the approach using novel stable approach criteria. These criteria incorporated several parameters: bank angle, airspeed, vertical speed, course, and vertical and horizontal path. The study intentionally used criteria that did not perfectly align with existing industry stable approach guidance to avoid bias that could result from individuals having differing levels of experience using them. Air carrier pilots possess an abundance of experience under such criteria, whereas most pilots with less experience do not. By selecting somewhat unique criteria, low- and high-experience pilots would have similar adjustments to their mental models.

Participants were asked to assess and respond on the basis of the unique bounds shown in Table 4. In all vignettes, the required briefings, checklists, and airplane configuration were briefed as complete and verified. The RA system, as equipped in the B-737-800, provided two audio annunciations: “1,000” and “500” when sensed. The airspeed, vertical speed, roll angle, course, and path criteria were to be considered upon annunciation of “1,000” to the participant. When these criteria were exceeded, the airplane was considered to be in an unstable approach state. After hearing the “1,000” but before hearing the “500” annunciation, the expected pilot response would be to correct the airplane path or energy state to achieve a stable condition. Below the “500” annunciation, the presence of an unstable approach, as defined by the bounds, would direct immediate go-around procedure accomplishment. Participants were briefed on the topic, were subsequently given a card with the criteria, point of application, and response actions, and were asked to take a moment to study and memorize them.

**Table 4**

*Experiment Stabilized Approach Criteria*

| Element | Criteria |
|---|---|
| Airspeed | Airspeed not more than $V_{REF}$ +10 kts and not less than $V_{REF}$ |
| Vertical Speed | Vertical speed no greater than 1,000 fpm |
| Roll Angle | Roll angle not in excess of 6° |
| Path | Less than one dot deviation from localizer and glideslope |

The same profiles were used for the control treatment and all three treatments, with the difference between the treatment groups being the presentation of added system indications overlays to the existing instrument presentation. No material changes occurred in the ND presentation provided the participant, as ILS path adherence errors did not manifest to the point that the ND, alone, indicated instability. The presence of the ND was considered useful to the study design in that pilots have established monitoring behavior patterns that include occasional crosschecking of approach progress and projection of future actions.

In the experiment design, one concern faced was the ability to sufficiently replicate the cognitive rigor when flying an ILS approach and monitoring the FD for unstable conditions. Simply monitoring the FD for unstable conditions would not simulate the shared cognitive burden associated when controlling the aircraft during an approach and monitoring for unstable conditions. Therefore, a secondary task was added. Given study participants were required to divert attention resources to the control of the airplane flight path, differentiation among participants could have been masked by the completion of a trite task. That is, the task effort gradient might have been insufficient to expose the variation between the treatments, indicating the need for reconsideration of the task workload. In that instance, the resource model of attention as applied by Vidulich (2003) would suggest an inverse approach to the under-tasked participant: the introduction of a secondary task whose purpose is to elevate the workload level placed upon the participant, and thereby more affect the reserve capacity available.

Such a secondary task furthered the psychological-cognitive fidelity of the experiment design, placing the participant in an elevated level of stress commensurate with the operational environment. Wickens and Liu (1988) completed an experiment in which they examined the relevance of codes and modalities in the prediction of task interference, using either verbal or key press actions that were solicited during a second-order tracking task. They followed this with a review of the extant literature on codes and modalities in multiple resources. They found that interference with tracking was consistently greater when a spatial decision task was paired with a manual response. As such, the application of a visual secondary task was supportable as a means to impact the task workload.

Pilots flying an airplane employ the use of both hands to provide input to the flight control and throttle inceptors, the typical configuration being the placement of the left hand on a yoke or stick and the right on a throttle(s) or thrust lever(s). The study secondary task included a physical response to cueing by the software. Two distinct visual cues were displayed outward of the FD for task triggering. A red cue was displayed for a simulated flight control inceptor response and a green cue displayed for a simulated throttle/thrust inceptor response. Cueing was configured to the prevalent inceptor design, and no consideration was given for recency in either particular seat of a crewed airplane. The look and placement of the visual cues is presented in Figure 9. These cues were located so as to allow detection within peripheral vision and not impact the participant’s instrument scan. For the first 15-20 s of each vignette, these cues were inhibited to allow the participant to settle into the profile.

Cue presentation frequency varied. Red cues, Callout 1, were shown at a varied 0.25-0.5 cycles per second (Hz) rate, whereas green cues, Callout 2, were shown at a varied 0.16-0.2 Hz. These values were established using information drawn from aircraft flight control data traces and SME input and feedback. Dwell time for each cue display was 13 frames (0.52 s). This value allowed sufficient dwell time to be perceived by a participant conducting an efficient instrument crosscheck, but not so long as to decrease the participant response gain. Pilots were asked to respond in a timely fashion using separate handheld switches. When each cue displayed, participants were to depress the inline handheld switch corresponding to the side the cue was displayed. Participants were informed that only a momentary press was needed for successful recording, even though no actual tracking of performance was measured. Consistent compliance with the secondary task was critical to accuracy of the data. Verification of participant response was monitored by means of red light emitting diode activation on a connected control box, one for each channel, removed from the direct field of view of the participant to prevent an experimenter effect. When the participant began to lapse in response, the researcher provided a “Cue Response” reminder.

**Figure 9**

*Secondary Task Visual Cue Presentation*


*Note*. Decisions and SA requirements for executing the major goal, *Maintain Aircraft Control.* Adapted from "Prepar3D®" by Lockheed Martin, 2017 (http://www.prepar3d.com/prepar3d-store/). Copyright by Lockheed Martin.

Within each vignette, a number of unstable conditions were presented. When an unstable approach condition was recognized, participants were tasked to speak aloud (a) the presence of an unstable condition, (b) the element exceeded, and (c) the action to be executed. Participants were asked to do so as *timely and accurately* as possible whenever observed, no matter how short lived in duration. Further, they were informed there may be multiple criteria being exceeded at a given time, so the pre-trial briefing emphasized the need for each to be identified. This was critical when presented the alert treatment, where it would be easy to rely solely on its presence and fail to recognize more than one criterion being exceeded. Participants were reminded that in time-critical flight phases, accurate flight parameter perception and quick decision-making and action are necessary. Each participant was notified that responses outside five seconds would not be counted, so they were to avoid excessive deliberation and not delay response submission.

The accuracy of the responses was measured against values determined by the SMEs and individually recorded. A point was allocated for each query element – the presence of the condition, the element exceeded, and the action to execute – when the participant correctly responded. In total, there were 51 possible points allocated for each treatment. Incorrect responses were not counted against the participant score, but data were collected on false alarm instances. The discrete IV valuation was used for the statistical analysis.

***Pilot Study***

Prior to formal experiment data collection, a pilot study was undertaken. Ritter et al. (2013) present the pilot study as an integral step in the preparation for running an experiment. Cherulnik (2001) notes that researchers must pay regard to the degree by which the IV selection and design, as well as the presentation of those treatments, provide a sufficiently realistic impact on the participant. Cherulnik cites *mundane realism*, the degree to which it mirrors the real-world experience. Pilot studies can be conducted with small groups and lessons garnered by the researcher from that experience; appropriate adjustments may then be made to refine the treatments used to ensure the hypotheses are adequately operationalized in the IVs (Cherulnik, 2001).

The purpose of this pilot study, therefore, was twofold: (1) to gather early and valuable insights into the relationships between the IVs and DV, as well as their initial support to the various hypotheses under test; and (2) to validate the overall experiment design and delivery, to include all pre-briefings, software deployment, data collection, and post-event debriefings.

Four individuals were selectively drawn from the range of study participant demographics to serve as SMEs. These pilot study participants were asked to observe and respond to the vignettes to gain insight on the reliability and validity of the instrument. Pilot study participants were then asked to comment on the appropriateness of the vignette design and provide suggestions for improvements.

***Display Enhancements***

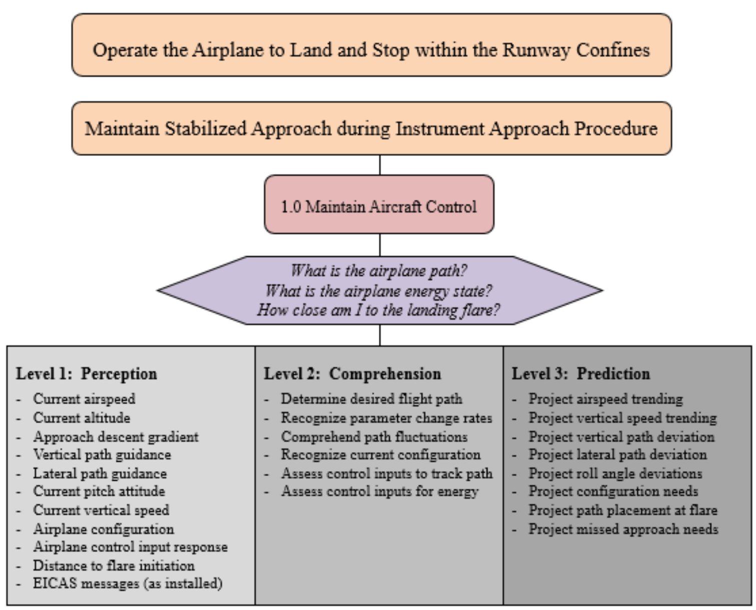
The initial instrument design step involved determination of the SA requirements for a pilot to accomplish an assigned task. The completion of a well-researched and applied Goal-Directed Task Analysis (GDTA) was a key factor in that determination. The GDTA seeks to document the cognitive demands required of a job rather than the physical activities that must be accomplished (Endsley & Jones, 2012).

In examining the purpose of the GDTA process, Endsley et al. (2003) note, “The problem of determining what aspects of the situation are important for a particular operator’s SA has frequently been approached using a form of cognitive task analysis called a goal-directed task analysis” (p. 268). In doing so, the major goals of the particular task are identified, and the major subgoals necessary for meeting each of these goals determined.

To support this study, a previously-developed GDTA structure derived from three SME interviews was applied (Greer et al., 2018). The GDTA methodology is presented at Appendix C. Applying the GDTA, the major goals, subgoals, and decisions were identified. The major goal was the overall task objective the pilot must achieve to successfully perform the job (Endsley & Jones, 2012). Subgoals formed individual subordinate jobs that must be completed in order for the main goal to be achieved. Decisions in the GDTA drew upon higher order evaluation and synthesis, and the output was determined necessary to achieve the superior goal. These decisions were presented in the form of an open-ended question and called upon the pilot to use the SA information available to successfully respond (Endsley & Jones, 2012). The results of the major goal, *Maintain Aircraft Control*, are presented in Figure 10. The SA requirements for each of the levels were considered in the design of the study.

**Figure 10**

*Goal-Directed Task Analysis – Maintain Aircraft Control*




*Note*. Decisions and SA requirements for executing the major goal, *Maintain Aircraft Control.*

The baseline PFD, the control for the study, is presented at Appendix A, while the proposed enhancements are presented at Appendix D. The PFD chosen was that of a

Boeing 737, but equally could have been drawn from a number of manufacturer offerings. Altitude, airspeed, roll angle, vertical speed, and path display differences among manufacturers were immaterial as a collection of international regulatory documents, technical standards, and industry positions provided electronic flight display standardization for the placement of the data (FAA, 2014a). Display enhancements were developed by amalgamation of current and proposed stable approach criteria. Each proposed enhancement was presented as an adjunct to that of the control display, which was devoid of any criteria or alerting content pertaining to unstable approaches. Thus, the proposed flight instrument display enhancement built upon the existing presentation of the airplane performance, path guidance, and path adherence flight data, acting as an overlay of additional information. The proposed displays were expected to reduce the amount of stabilized approach information to be stored in long-term memory by using embedded logic to visually depict the parameters or an alert. Thus, they were anticipated to provide better SA to the pilot and support the proposition that pilot resources could be better allocated to other relevant tasks.

### *Apparatus and Materials*

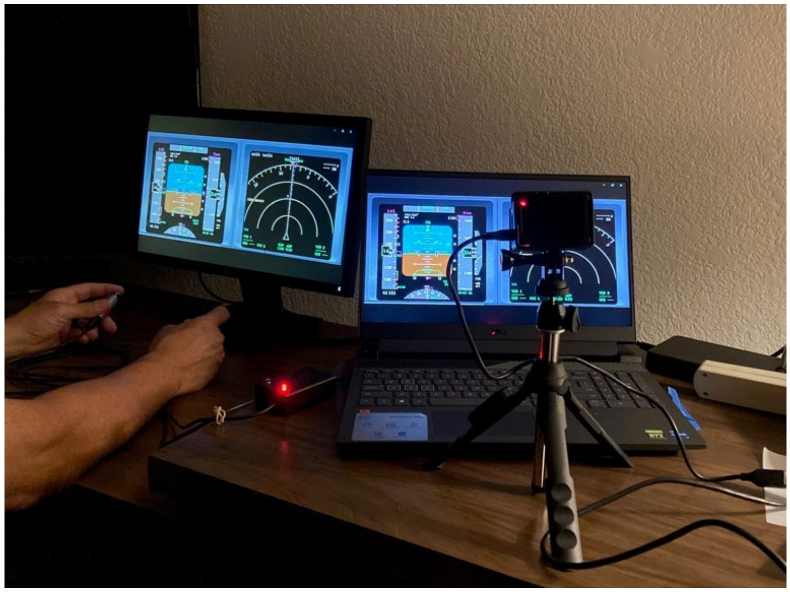
The study was conducted in a controlled environment in which issues affecting participant comfort and safety could be managed. Participants engaged from a stable, seated position, as would be experienced in a modern flight deck, and were placed clear of any obstructions that may interfere with movement or pose harm. However, participants were expected to remain relatively motionless during the study trials.

The presentation was isolated to the data observed on a representative PFD and ND during an ILS approach. No other airplane-related presentation was made, as addressed in the Delimitations section. Therefore, the test apparatus had to demonstrate the ability to provide an accurate presentation of the display parameters and the dynamics of each as the approach progresses. As ambient facility noise and other extraneous distractions could impact the results of the study, trials were conducted in isolated areas. Flight deck audio was used to complement the participant experience. These served to aid in participant immersion while restricting the influence of surrounding audio distractors.

Participant think aloud audio output was recorded commensurate with the video vignette presented. Vignette video was recorded simultaneously to synchronize participant output data with each vignette. A primary (Vivitar DVR4K-BLK) and secondary camera (GoPro HERO4) were used to improve the likelihood of complete data capture. Participants wore two clip-on Weishan-brand corded lavalier omnidirectional condenser microphones to capture audio output. The microphones were attached to the participant's clothing, proximate to the neckline. The microphone output was fed directly into the Vivitar and GoPro cameras. The video cameras were placed in front of the researcher's laptop to provide a direct capture of each vignette. The data were subsequently downloaded onto the study research laptop for later analysis and backed up on a Western Digital My Passport portable hard disk drive. The entirety of the data collection system was portable to allow for easy movement between data collection sites and placement in a number of orientations and locations. The data collection apparatus is shown in Figure 11.

**Figure 11**

*Data Collection Apparatus*



*Note*. Set up of the data collection apparatus, showing the participant and researcher stations. Recording camera placement is shown. The secondary task switches and switch activation indicator lights can be seen. Lavalier microphone connection is out of view.

Accurate representation of the Boeing 737 displays was achieved using the Lockheed-Martin Prepar3D® v4 Professional Plus flight simulation software employing the Flight1 Software iFly Jets 737NG add-on application. This combination software package was capable of accurate presentation of a number of flight instrument display configurations and supported the fidelity required by the study. Audio information including turbine engine noise, pitch trim system operation, radio altitude system altitude callouts, approach minimum altitude callouts, and chimes and alerts were captured to provide an elevated sense of realism to the participant.

To properly present the information, participants were positioned relative to a flat panel monitor to achieve the anthropometric design eye reference point – the fore-aft, lateral, and vertical placement of the pilot's eyes within the flight deck. The positioning emulated viewing a pair of two 8.5 in. by 8.5 in. side-by-side presentations situated perpendicular to the design eye reference point from 32 in at a -40° to -45° decrement from the horizon. This placement afforded the pilot a field of view that supports full field of regard search of flight deck instrumentation and eternal windscreen presentations (Wang et al., 2018).

A high definition (HD) display was used and proved beneficial for a number of reasons. First, HD displays provide flexibility in screen size selection, allowing for adequate visual field presentation for the study participant. Such systems are able to do so at a moderate cost. Second, when properly sized and located within the research space, the device aids in occlusion of external visual disturbances. Distractors such as ambient lighting changes and extraneous nearby movement could be removed. Third, such systems afford a degree of portability, allowing the study to be conducted at various locations and support a diverse sampling method. Finally, required fidelity can be met at the lowest system level. The HD displays allowed for display resolution and size matching that of an actual airplane. More complex systems such as virtual reality, while likely to provide a greater degree of immersion for the participant and add a sense of uniqueness, were determined to be more difficult to develop, imposed a requirement for additional equipment, risked participant exposure to potential nausea-inducing and disorienting presentations, and posed ethical issues with experimentation on human participants (Aukstakalnis, 2017).

Vignettes and individual treatments were created using Adobe Premier Pro® video editing software. The raw Lockheed-Martin files were imported into Premier Pro®, where title slides and fading of video and audio tracks were applied. Each of the core vignettes were reviewed in frame-by-frame fashion. Depending on the applicable treatment addressed, boundaries or alerts were added as video tracks. For boundary presentation, all elements were fixed in space with one exception, the airspeed boundaries. Key framing techniques were used to ensure a smooth tracking of the amber marks relevant to the airspeed tape. For the alert treatment, overlapping exceedance conditions were combined so as to present only when parameters, even in combination, were out of bounds.

***Sources of the Data***

Data were drawn from the pre-trial, within-trial, and post-trial phases. Instruments used to gather this data are presented at Appendix E. These sources included a pre-trial demographic survey instrument, within-trial SA measurements, video and audio recordings, hand-written notes, and measured numeric data values. Each source of data originated from a specific phase of the study.

The pre-trial survey contained questions to gather the background and experience demographics of each participant. The demographics included education level, flight training history, current position, pilot certificates and ratings held, instrument time, total flight time, experience with electronic displays, previous experience with an unstable approach, and knowledge of the FAA and FSF stable approach guidance. The information obtained from this survey formed the basis for characterization and analysis of the sample used to obtain the data. As the study employed a within-group, repeated-measures, independent design, verification of adequate experience representation supported the validity of the research and extension to the broader pilot community.

Data obtained within-trial were collected automatically and manually. Recording allowed not only the capture of participant detection of actual unstable conditions, but commentary surrounding the conditions and instances of missed and false alarm events. False alarm events provided valuable insights on individual SA at times where no actual unstable condition was present. Handwritten notes were taken during each participant's session to track the order of vignettes presented to the participant and document experiment execution issues that arose.

The post-trial survey consisted of both forced choice and open-ended questions addressing (a) ease with which the display could be read, (b) potential interference with completion of the monitoring task, (c) enhancement of perception and understanding of the current airplane stable approach state, (d) enhancement of prediction of the future airplane stable approach state, (e) benefit in flight operations, (f) benefit in training operations, and (g) order of preference of enhancements. Scores were gathered using a 5-point Likert scale. Response options included *strongly disagree, disagree, neutral, agree, and strongly agree,* where a score of 1 indicated *strongly disagree* and a 5 indicated *strongly agree*. At the completion of the forced choice questions, comments were solicited concerning the value and use of the proposed display enhancements and potential study conduct improvements.

**Ethical Consideration**

The use of human participants necessitated the completion of an Institutional Review Board (IRB) assessment of the research protocols. The completion of a pilot study prior to scored data collection further required IRB approval. The applicable ERAU IRB forms were completed along with an informed consent form to be signed by each participant. Proposed advertising and participant recruitment messaging were also submitted for review and approval.

In the experiment design type, Vogt et al. (2012) identify four key research ethics issues: limiting deceptive practices, curbing physical or psychological harm, informed consent, and debriefing. Each of these was addressed in the IRB application. Per ERAU policy, a subsequent redirection in the study method necessitated an amendment to the IRB. Key IRB-related documents are presented in Appendix F. Based on the proposed research design and subsequent modification involving no more than minimal risk, expedited review was determined and approval to proceed granted.

***Gaining Access to Participants***

When required, notification of the study was sent electronically to targeted industry and academic groups explaining its general purpose and the demographics of participants sought. Directions on how to contact the researcher for participation were provided, as well as the study title, purpose, participation criteria, risks and discomforts, and method for contacting the researcher for data collection dates and locations. Pilots wishing to participate were initially vetted to ensure they met the eligibility requirements. Chosen participants were categorized into levels of flight experience to ensure a balance within the group.

Agreements were reached at the various participant locations on the terms of researcher access and addressed facility security, privacy guidance, resource use, and recording requirements. Where study participants were engaged as part of their normal employment or participation in an academic program, permission to do so was obtained from supervisory personnel. The researcher coordinated follow-up participant engagement and negotiated commitments to provide post-research presentations.

***Informed Consent and Expected Risk***

An Informed Consent (IC) form was developed and included in the IRB package. It was noted that participation would be voluntary, and participants could elect to terminate participation at any point. The mental and physical risk to the study participants was expected to be minimal and no higher than that during other training or practical flight experiences in which visual displays of flight instrumentation are presented. The informed consent document is presented at Appendix G.

As with any study, slight levels of mental and cognitive stress and fatigue may be present. If the participant experienced any discomfort, the participant was allowed to bring the trial to a stop. Centers for Disease Control and Embry-Riddle Aeronautical University sanitation, masking, and safe distancing protocols were observed. When data collection sites required more stringent practices, they took precedence.

During the briefing, the participants were informed that the purpose of the study was to observe and gather data on participant SA during the final approach segment of a flight operation. The ultimate purpose of the study was withheld to ensure the participant was not disposed to hyper-alertness to approach instability conditions, which could confound data collection efforts and the research results. Withholding of some information relevant to the study was unavoidable, but the benefits to society outweighed any risks that could have arisen from doing so. Each participant was debriefed upon completion of the final trial as to the objective of the study and the manner in which the data would be used.

The results of the study were expected to have no detrimental effect on the participant's career. As an additional safety measure, protocols were established to ensure the results of individual trials were kept confidential and participants would be identified only by assigned number.

***Data Protection***

The study kept all participants anonymous. Appropriate measures were put in place to ensure privacy and confidentiality and provide for protections against traceability. This was particularly critical for the population being sampled, as pilots take considerable pride in their professional capabilities and have shown a tendency to fear the release of information that may reflect negatively on their flight-related abilities among their peers and, where applicable, school administration and employers.

Neither the pre-trial demographic questionnaire, experiment data collection, nor post-trial questionnaire gathered personal identifiable information. Confidential participant information such as name, certifications held, experiences, trial responses, and perceived enhancement usability were only accessible to the researchers and used for the purpose of (a) determining eligibility to be a member of the experiment sample group and (b) supporting study results and conclusions. Otherwise, names or any other identifying demographics were not able to be matched. Publication of the experiment results referred to the data in a generic manner, not inclusive of any identifying information.

To ensure confidentiality, each participant was assigned a random participant identification number attached to the data files. Each participant was subsequently referred to by their participant identification number. Retention of specific name identification was only necessary for possible follow-up inquiries. The key code for specific names and their correlation to an identification number was kept under strict control and locked in a separate location.

Video was collected of the screen display and did not include capture of the participant image or other identifiable information. All data collected in the study were kept under strict control during study execution. The laptop used to collect and store the data was physically retained in the possession of the principal investigator during trial sessions and stored in a controlled access locked room between collection periods. The laptop and data files were password protected.

Data were secured using the current Embry-Riddle Aeronautical University policy. Information collected as part of this research could be used for future research studies and was retained for a period of three years. Only the principal investigator and the listed other investigator retained access. Paper record of participant data was destroyed using common shredding methods following approval of the dissertation.

***Avoidance of Bias***

Practices were established to avoid researcher bias which can impact study internal validity. Data collection was automatic and objective, removing the researcher from primary measurement activities. The literature review guided the factors to be measured by the instrument. Further, SME input was obtained when developing the experiment vignettes to improve construct validity. Finally, SMEs were also employed when interpreting participant responses provided to open-ended survey questions.

***Benefits***

While there were no benefits to the participants, findings from this research could provide empirical research data and conclusions that might prove beneficial to the academic community, flight instrument display manufacturers, commercial air carrier operations groups, and pilots undergoing flight training in programs that make use of scenarios exposing the potential for unstable approaches. In consideration of their time and effort and to serve as an incentive, study participants were afforded a $20.00 gift card for their participation. All participants who initiated the research process received the gift card, regardless of whether they completed the research trials or not. Participants who showed to participate but did not consent were not provided the incentive. The nature of this reward was determined in consultation with ERAU faculty and was set at a level that is customary within the research community.

**Measurement Instrument**

A unique SA measure was devised to collect discrete data during the within-trial phase, in support of the DV. Endsley & Jones (2012) note that SA, as an internalized mental construct, can be difficult to measure. They further note that SA is either inferred through observable behaviors, comments, or outcomes that are easier to assess, or it is directly measured using various verbal or written instruments. The decision as to which measure to be used in SA research is dependent upon a number of factors such as the type of study being conducted, the duration of the data collection period, the ability for the participant to respond during or after a trial, and the volatility of memory. Elements of two different indirect and direct measures were used to gather participant SA. The advantages and disadvantages of each measure is shown in Table 5.

**Table 5**

*Situation Awareness Measures – Advantages and Disadvantages*

| Method | Advantages | Disadvantages |
|---|---|---|
| Indirect Measures<br>Verbal Protocols<br>(Think Aloud) | Provides information on SA strategies or processes<br>Provide insight into information integration and use | Does not provide a complete representation of what is processed<br>Participant verbal skills can impact the results<br>Can slow performance while responding<br>Information in memory isn't captured |
| Direct Measures<br>Online Queries<br>(SPAM) | Overcomes memory issues from post-test queries<br>Queries are embedded in the on-going task | May intrude on task performance<br>May alter SA by shifting attention<br>May increase workload<br>Potential for increased data point requirements |

*Note:* Adapted from *Designing for Situation Awareness: An Approach to User-Centered Design* (2nd ed.) (pp. 261-276), by M. R Endsley and D. G. Jones, 2012, New York, NY: CRC Press. Copyright by CRC Press.

***Indirect Measures***

Indirect measures attempt to assess the participant's SA by inference, measuring the cognitive processes participants express in SA development (Endsley & Jones, 2012). One such method is the application of *think aloud* verbal protocols, whereby the participant vocalizes their thought processes while performing a task. Roth and Mavin (2014) successfully employed think aloud protocol in a study of how expert pilots assess their peer experts in the aviation domain. Participant pilots were asked to critique peer performance in the pilot flying (PF) role as they underwent constructed simulator profiles. In an analysis of driver behavior during emergency situations, Banks et al. (2014) asked drivers to provide verbal reports when completing a difficult driving task involving an emergency situation. They noted two types of protocols: concurrent and retrospective. Concurrent protocols require participants to verbalize their thoughts while performing the desired task, whereas retrospective protocols are accomplished and measured after the fact. They caution that just because a participant didn't vocalize a response to a particular trigger event does not mean the participant wasn't aware of the presence of the event. They further comment that retrospective tasks should be limited to short duration tasks or risk omission. As such, concurrent responses are more likely to be accurate.

Thus, the think aloud verbal protocols typically provide a running commentary as to the thoughts and decisions being made by the individual, which is later dissected and analyzed to determine the participant's SA. When recorded, they provide an archival source of data for analyses of not only response accuracy, but response timeliness as well.

***Direct Measures***

Direct objective measures attempt to assess the participant's SA by making a comparison between the SA as reported by the participant to that of the reality of the case presented the participant, and can be divided into those in which the operator provides direct response to either real-time online probes or post-trial questionnaires. When applying direct objective measures, Endsley and Jones (2012) comment, "The comparison is often made by querying operators about aspects of the environment and then assessing the accuracy of the responses by comparing them with reality" (p. 270).

One such direct objective measure is the Situation Present Assessment Measure (SPAM). Lau and Boring (2018) address the advantages and disadvantages of the SPAM probe technique. Previous studies by Loft et al. (2015) and Durso et al. (2006) applied SPAM to predict incremental variance in SA associated with submarine track management and ATC tasks, respectively. Shelton et al. (2013) successfully applied

SPAM principles in the medical field when assessing real-time SA assessment in critical illness patients, gathering data using in-person, telephonic, and personal digital assistant queries. Loft et al. (2015) found this method to be successful in predicting such variance, while Pierce (2012) conducted a study in which workload and performance of ATC controllers during accomplishment of the Air Traffic Scenarios Test were assessed using SPAM, non-SPAM, and no-probe conditions. Pierce found that the administration of such assessments posed little impact on the performance of more experienced participants, but novice participants demonstrated some effect on their task accomplishment.

The SPAM advantages and considerations supported application within this study. The SPAM measures the RA from real-time SA probes and correlates the results to the degree of SA held by the participant. The assumption behind the SPAM is that those participants possessing higher SA will be able to respond to the inquiries in a more rapid fashion and do so more accurately. It does so in real-time and does not require the cessation of the profile in order to gather the data (Durso & Alexander, 2010). Further, the data are collected with the instrument present, so as to focus on the participant's ability to respond to the data presentation (Shelton et al., 2013). Thus, memory does not factor into the determination of SA. The accuracy of the response to a query is of importance in determining participant SA, as has been identified in a number of previous studies by Shelton et al. (2013) and Baker (2017). Baker successfully applied SPAM in conducting a comparative analysis of conventional and OZ display concepts in the assessment of airplane energy management.

The within-trial phase incorporated some elements of each method in that it measured in response accuracy real-time while immersed in the on-going display presentation (SPAM) and asked the participant to vocalize the presence of an unstable approach condition (think aloud). Participants were not queried directly by the researcher, but rather were briefed and expected to initiate the desire response when they observed every instance in which the approach was unstable. They were not expected to provide a running narrative as to their thought processes throughout each trial vignette. Further, they were not expected to retain their SA-specific data for any post-trial assessment of each vignette. Interaction with the system was limited to cue responses replicating flight path management inceptor frequency, but not directly controlling the simulated path.

***Variables***

The study employed a 2x2 factorial within-group, repeated-measures design using two separate categorical IVs. Both IVs addressed the PFD stabilized approach treatment presented to the study participants. The structure is presented in Table 6. The two IVs, labeled Boundaries and Alert, are depicted along the rows and columns respectively. This factorial structure generated four treatments to be assessed by each participant.

**Table 6**

*2x2 Factorial Design*

| | | Alert | |
|---|---|---|---|
| | | Absent | Present |
| Boundaries | Absent | Treatment 1 | Treatment 3 |
| | Present | Treatment 2 | Treatment 4 |

The first IV manipulated the presence or absence of the stabilized approach boundaries, and the second IV manipulated the presence or absence of the alert message. Thus, one control and three treatment conditions were present. Each of the categorical IV treatments examined a variation in the presentation of an unstable approach condition. In the first treatment, Treatment 1, the IV combination presented the PFD as currently designed, devoid of any boundaries or alerts. This treatment served as the control. A depiction and discussion of the symbology used for the control PFD is presented at Appendix A. The subsequent three treatments provided boundaries and alerts applied to the control display, allowing analysis of the main effects. Treatment 2 presented stability information in terms of amber-colored boundaries for airspeed, roll, vertical speed, and vertical and lateral path. These criteria manifested as an overlay of the existing PFD presentation. This IV treatment is presented in Figure D1 in Appendix D. In Treatment 3, a singular amber instability alert message, UNSTABLE, was provided to the pilot. Presentation of the cautionary alert message on the PFD occurred whenever the airspeed, roll, vertical speed, or vertical or lateral path bounds were exceeded. This IV treatment is presented in Figure D2 in Appendix D. In the final treatment, Treatment 4, both boundaries and an alert were presented on the PFD. This IV treatment is presented in Figure D3 in Appendix D.

Therefore, the lower bound for the IV display treatments was that of no change to the existing design, whereas the upper bound was limited to an information presentation point in which the user realized saturation and a net decrease in SA. Information manipulation in each IV treatment was applied so as to avoid exceeding the upper bound. A single DV was evaluated using repeated-measures factorial ANOVA. The DV measured the accuracy with which the participant responded to SA queries presented during each trial.

## Data Analysis Approach

### *Participant Demographics*

Study participant demographic collection can prove useful in analyzing what part, if any, particular participant traits may play a role in the study results. They can serve to identify cofactors when undertaking data analysis and deriving conclusions. Further, they can assist in identifying participants who fail to meet the entry criteria and thus risk the introduction of data bias.

Typical demographic participant areas include self-identified characteristics of age, gender, ethnicity, education level, field of study, employment status, marital status, and the like. Though some of these demographic values were relevant, several were not. Reasoning suggested that age may be correlated to flight experience, given a defined starting point for entry into the flying profession. However, a younger pilot entering professional flight training while attending a university program may actually possess considerably more instrument exposure at an early age than a pilot pursuing flight as a second career. Equally, education was thought to impact the response of the individual, with those choosing an aviation-related field of study possibly demonstrating better training in, and deeper understanding of, flight operations in instrument conditions and the interpretation of various instrument displays. Employment status in the aviation career field was believed to assist in flagging those pilots whose current proficiency may be lacking, or those whose abilities may be exceptional. As such, a pilot holding an air carrier pilot position may be more likely to demonstrate higher levels of SA than a university student who recently obtained an instrument rating.

Of greater importance to the study were those demographic values that relate to exposure and qualification in the aviation domain. These included the level of pilot certificate currently held, to include any additional qualifications such as instructor. The number of flight hours were thought to also impact the results of the study; properly crafted, self-identification of flight hours within defined bands facilitated data analysis. Exposure to Electronic Flight Instrumentation Systems (EFIS), specifically PFD/ND formats used in the study, was felt to be helpful in determining a possible level of preexisting comfort with this presentation method for airplane parameters. Of other interest was the amount of experience the pilot may have in particular flight environments: time in actual or simulated IMC and numbers of ILS approaches flown. Finally, and critically, experience with defined stable approach criteria proved of interest in determining if previous exposure to such parameters impacted the results.

To gather this information, a pre-event questionnaire was administered to pilots participating in the study. Participants were briefed to provide as accurate a response as possible but were also informed that response to these demographic questions was optional.

***Reliability Assessment Method***

Measurement of pilot behavior was accomplished through application of an empirical-analytical approach (Drost, 2011). Reliability addresses the ability of the instrument to consistently measure the values being examined and can be influenced through random errors of the measurement. Drost (2011) identifies these errors as systemic, owing to inaccurate calibration, or random, resulting from actions such as misreading the scales. Given the discrete DV values, scoring did not require measurement to high levels of granularity. A single source of trial vignette delivery and recording avoided systemic reliability error potential. Random errors were avoided through the use of a sufficient number of trials, ensuring the participant was appropriately rested, and other means.

***Validity Assessment Method***

In the experiment construct, two forms of validity were addressed: internal validity and external validity. According to Creswell (2012):

> Internal validity refers to the extent to which a study's results can be correctly attributed to the treatment of the independent variable. A study is internally valid when one can rightly draw accurate conclusions about the causal effects of the treatment on the outcome in the sample studied. (p. 53)

When examining the design of an experiment, the fidelity of implementation must be considered if the research seeks to understand not just how a particular IV treatment affects the DV, but also how it does so. Vogt et al. (2012) succinctly present the beauty of the experiment design in supporting internal validity: "In short, experiments are especially good at ensuring *internal* validity, that is, drawing correct conclusions about your sample, especially regarding causal effects" (p. 55). The conduct of the experiment in a laboratory setting inoculates the research from confounding variables that may arise when working in a naturalistic setting.

Internal validity challenges have been addressed. The participant selection process, drawing from a number of sources at various locations, minimized the potential for selection bias. In grouping by flight time totals, an equal balance of low- and high-experience level was represented. Conducting a single sitting for each participant minimized potential history threats resulting from external events. Little opportunity was afforded for external factors to impact the study outcome. Timely data collection completion across the entire sample negated maturation and mortality threats.

With application of a Latin Squares sequence, testing validity threats were reduced. The likelihood of a participant identifying a particular core profile, recalling the experience, and biasing the response was small. Participants were escorted in and out of the facility to ensure no contact took place between individuals prior to their participation. This practice avoided potential diffusion of the study results through information exchange between the group members.

Each participant was provided a short orientation to the use of the simulation equipment and the PFD and ND display format. They were made aware of all treatments being applied and that the vignettes and treatments did not change. This approach minimized the influence of potential confounding variables and instrument decay and ensured that each pilot had proximate exposure to the display environment prior to data collection trials. Equally, the within-group design controlled for threats owing to resentment and rivalry, as all participants saw every treatment. Additionally, nominal compensation for participation was equal in value regardless of background and experience or level of contribution to the study.

External validity, in contrast, is concerned with the degree to which the results of the study can be generalized to the broader population (Vogt et al, 2014). The controls placed on an experiment to facilitate the mechanism of causality bind it to a smaller, more restrictive application. The sample frame sought to capture pilot participants who held instrument flight certification and focused on the use of the PFD and ND as the framework on which the experiment rested. Generalizing the results beyond the research sample frame was not recommended since very specific training is required to become qualified to operate an airplane in instrument flight conditions.

Vogt et al. (2014) identify construct validity as an important consideration when coding and measuring under experimentation. This validity seeks to ensure that the manner in which the concepts are coded and measured support the study being conducted. Content validity was addressed through the application of pilot testing of the instruments prior to application in the study. Subject matter experts in flight operations, human factors, and aviation psychology assessed the instrument and provided feedback for improvement.

Special care was given to the issues of systemic error in the presentation of each core scenario and random error in the accuracy of the data collection by the researcher. Internal documentation and process reviews ensured high internal consistency reliability and dry runs of the tasks undertaken to ensure complete understanding of the form, content, and delivery of the tasks and the measures.

Individual participant exposure to the study battery of trials did not exceed 70 min duration. Application of a repeated-measures structure supported the objective of gathering all data from each participant in a single setting, thus reducing the potential for experimental attrition. This design allowed up to five participants to complete the study every day.

### *Data Analysis Process/Hypothesis Testing*

Data analysis was conducted using the current version of the International Business Machines Statistical Package for the Social Sciences (IBM SPSS®) (2023)

software. A two-way, within-group, repeated-measures factorial ANOVA was used to test whether the SA of a pilot differs based on specific display treatments applied. Field (2013) notes that the application of factorial ANOVA is appropriate when two or more IVs exist with a commensurate desire to examine the interactions between different IVs and contrast the differences between the associated treatments. Whereas it might appear reasonable to have completed the analysis using several *t*-test calculations, Field (2013) cautions that when conducting a number of statistical tests on common data, the impact of committing Type I errors, in which the decision is made to reject the null hypothesis when it is true, begin to accumulate.

As an omnibus test, ANOVA provided indication as to whether the means of each treatment DV were equal or not but would not isolate the manner in which they are. According to Field (2013), planned contrasts may be used to identify which of the treatments differ without inflating the Type I family-wise error rate that would occur with multiple *t*-test application. Such contrasts are appropriate when there are specific hypotheses to be tested.

Field (2013) suggests a procedural approach to analysis, beginning with an exploration of the data to identify potential issues with outliers, lost or missing data, normality, and other issues. The application of boxplots, histograms, descriptive statistics, and other measures may prove useful in these determinations. Once identified, outlier corrections and normality problems should be dealt with to ensure the integrity of both the data and the statistical testing method. Once the data are in proper form and content, a two-way, repeated-measures ANOVA should be run. The results of this

ANOVA were to be used to ascertain if there is a statistically significant difference among the treatment combinations exhibited by the two IVs.

Data were analyzed for the DV of RA accuracy. Baker (2017) established a protocol for treating inaccurate responses. In his study, RA was critical as a representation of SA. Therefore, RA values were assessed against a correct standard developed by SMEs. This study pursued the data analysis in alignment with Baker's protocol.

***Assumptions for Statistical Analysis Techniques***

The proper application of parametric statistical tests such as the two-way, repeated-measures ANOVA is dependent upon certain research design and data assumptions being met. Research design structure addresses the nature of the model in terms of the type and number of IVs and DVs. Once met, Field (2013) identifies four key assumptions, common to most all parametric statistical models, which are extended to the ANOVA case. Field (2013) then addresses the assumption of sphericity, present when conducting analysis in repeated-measures designs. These assumptions include the following:

- Execution of the appropriate design
- Absence of significant outliers
- Independence of the residuals
- Normality of the DV data distribution
- Homogeneity and sphericity of the DV variance

**Execution of the Appropriate Design.** In order to employ a two-way, within-group, repeated-measures analysis, the research design must incorporate a single outcome

DV, with two or more IVs. If these IVs are categorical and make use of the same participants, the factorial ANOVA is suited, provided the other assumptions of parametric tests are met (Field, 2013). This study met these requirements.

**Absence of Significant Outliers.** Hair et al. (2010) address the importance of linearity of the data and the impact of outliers on the validity of statistical results. These conditions are critical to any parametric analysis, and attention must be made to ensure the associated issues are addressed so as to avoid biasing the results. The requirement for linearity of the data can be verified a number of ways. The first method examines scatterplot outputs of each variable for nonlinearity patterns. Data demonstrating linearity will characteristically follow a reference depiction of linearity, and comparison to that reference will highlight any variation. Other options for detecting nonlinearity include conducting a linear regression of the DV to explore the residuals and modeling the nonlinear relationship through curve fitting (Hair et al., 2010).

The presence of outlier data can impact the outcome of the research, and care must be taken to ensure such cases are identified and treated. Hair et al. (2010) suggest multiple methods for examining the presence of outliers. One graphic method for assessing outliers is through the use of boxplots, which provide outlier indications of data points falling outside 1.5 times the interquartile range (IQR) above the 75th percentile or below the 25th percentile. Extreme value indications exceed three times the IQR above the upper or lower quartile. Outliers may also be identified through the univariate, bivariate, and multivariate means, using standardized values, confidence intervals, and Mahalanobis $D^2$ measures, respectively. As this study involved categorical variables for the IV, use of Mahalanobis $D^2$ measure was determined inappropriate, and graphical analysis was performed. Studentized residual values for each RA score were assessed against the accepted standard of ±3 standard deviations (Field, 2013).

Equally, the issue of missing data and its ability to significantly bias the results must be considered. Hair et al. (2010) address a four-step process for identifying and remedying missing data. Using this approach, missing data were first assessed to determine the type of missing data, be they ignorable or not ignorable. If it cannot be determined, the next step is to examine the extent and pattern of the missing data and consider the potential for case and/or variable deletion. Next, the randomness of the missing data should be diagnosed, either as Missing at Random (MAR) or Missing Completely at Random (MCAR). The final step in the process entails selection of the data imputation method, should the necessity to retain the particular case(s) be present. Depending on whether the data are MAR or MCAR, a number of imputation options are available.

**Independence of the Residuals.** Hair et al. (2010) highlight independence as “the most basic, yet most serious, violation of an assumption” (p. 684). The residuals should be statistically independent, meaning that the responses provided by any particular treatment group for the IV are made independently of those made for the other treatment groups. Hair et al. (2010) offer that “any number of extraneous and unmeasured effects can affect the results by creating dependence between the groups” (p. 684). They then address the two most common, those being serial correlation and group impacts from extrinsic environmental factors, but note that there exist no tests that can, with absolute certainty, detect the infinite potential for the presence of interdependence and it is left to the researcher to identify and explore all the possible ways and seek out mitigations.

Participants were engaged independently, with protections to ensure that those having been exposed to the research design had no likelihood of contact with those about to be engaged. Further, the experimental design used Latin Square design to avoid time-ordered effects that could bias the results when within-groups, repeated-measures studies are conducted.

**Normality of the DV Data Distribution.** Field (2013) notes the residuals should be normally distributed. The tools available to assess univariate normality include visual assessment of both individual DVs using histograms and standardized Q-Q plots. Data demonstrating univariate normality will be depicted as a normal distribution in each histogram, failing to display any indication of leptokurtosis, platykurtosis, or any unusual skewness of the data. Examining the Q-Q plots, univariate normality will be evident if the plot of expected cumulative output to observed cumulative output follows an upward, diagonal line. In instances where the data plot bulges or sags from the diagonal, kurtosis is present. If the data takes an "S" shape about the diagonal, skewness will likely be the cause. An additional tool for numeric evaluation of univariate normality is the examination of the descriptive statistic output values for skewness and kurtosis. Commonly used statistical software often adjust such that values near zero approximate a normal distribution.

Beyond the application of graphical assessment, Field (2013) suggests application of either the Kolmogorov-Smirnov test or the Shapiro-Wilk test as a tool to assess the DV scores of each measure against a normally distributed set of scores that possess the same mean and standard deviation as the DV scores. Whereas both tests provide a similar function, Field notes "The Shapiro-Wilk test does much the same thing, but it has more power to detect differences from normality (so this test might be significant when the K-S test is not)" (2013, p. 188). Each test uses a model with a normal data value distribution developed from the sample mean and standard deviation. This is then assessed against the actual sample measured to test if the sample is significantly different from the model. Should the test determine insignificance – a $p > .05$ –, the sample data is not significantly different from normal, and the sample and normality can be assumed (Field, 2013).

**Homogeneity and Sphericity of the Variance.** Field (2013) points out the application of ANOVA assumes that the variances for each treatment of the IV are approximately equal or homogeneous. This determination is made through thoughtful examination as to whether the variances of each of the groups are approximately equal. Such homogeneity can be assessed using visual methods. Plots of standardized residuals against predicted values in which the data presents a funneled shape would suggest the presence of heteroscedasticity. Field (2013) offers the application of Levene's Test of Equality of Variances to assess the null hypothesis that the variances of each of the independent groups are the same, with a result of non-significance if the groups are deemed similar. This test accomplishes a simple one-way ANOVA between the absolute score of each value and the mean from the group of which it is a member. Steps would be necessary to correct the heterogeneity, should the variances of each group prove to be different.

The application of repeated-measures design introduces the assumption of sphericity, a general condition of compound symmetry (Field, 2013). Sphericity is a measure of the equality of the differences between the levels of treatments, as opposed to those within each level. Sphericity can be tested using Mauchly's test for two-way, repeated-measures ANOVA, which tests the equality of the variances of the differences between the treatment conditions (Field, 2013). The test hypothesizes the variances are equal and if significant, concludes a meaningful difference exists and the assumption of sphericity violated. The corollary, that the test result is non-significant, if demonstrated, would indicate the differences are roughly equal and sphericity is present, providing added confidence in the *F* ratio results.

For sphericity to be an issue, at least three treatment conditions must be present (Field, 2013). In this study, two treatment conditions existed and disallowed a comparison between at least two separate sets of data for each IV. Thus, sphericity was not considered an issue. This was affirmed in SPSS data analysis outputs where the Mauchly's test will indicate a value of 1.0, denoting perfect sphericity.

**Options if Assumptions Violated.** Hair et al. (2010) offer that data transformations can be used to correct for the lack of normality. Such transformations modify the DV and serve as the principle means of correcting for nonlinearity and heteroscedasticity of the data (Hair et al., 2010). Field (2013) closes in stating there exists no non-parametric ANOVA counterpart; if the countermeasures for overcoming assumption violations don't remedy the situation, there are no options for data analysis.

## Summary

A within-group, repeated-measures, quantitative, experiment using a posttest-only, control-group design was used (Creswell, 2014). The research population comprised all pilots currently exercising the privileges of FAA-issued certificates to operate under IMC, although a reduced sample frame was employed in obtaining the study sample. Based on an a priori, between-subject, experiment construct using ANOVA main effects and interactions, with accepted levels of effect size, error probability, and power, a minimum sample size of approximately 24 total participants was necessary to obtain a statistically supportable result. Participants were asked to view and then assess a battery of ILS approach video vignettes of PFD and ND displays. The design of each of the vignettes highlighted particular element(s) of an unstable approach to be observed by the participant.

# Chapter IV: Results

This chapter presents the participant demographics and the results from the study. The purpose of this research was to determine whether pilot SA was affected by applying enhancements to the flight instrument displays during an unstable approach. Purposive sampling was employed to approximate a representative segment of low- and high-time pilots holding an instrument qualification. Purpose-built vignettes were used in a series of trials. Each participant experienced a single orientation vignette, with the option of an additional practice vignette, followed by 12 data collection vignettes. Data analysis employed the process suggested by Truong (2017) whereby a data analysis plan was derived, the statistical and conceptual assumptions for the underlying technique were met, the model was estimated and assessed for fit, and the DVs were interpreted.

## Pilot Study

A pilot study was conducted prior to formal data collection. Four SMEs were exposed to the experiment protocol concepts, and data were collected over a period of several months in which a number of concept iterations were undergone and assessed. Delphi methods (West, 2011) supported anonymous SME response, allowed iterative and controlled SME feedback, and developed an aggregate pilot study response. The purpose of the study was to:

- verify continuity and reliability in presentation of the vignettes;
- assess methods and reliability for capturing participant think aloud output;
- conduct, evaluate, and refine the administrative actions for each trial;
- test the predicted timeline for administrative and data collection activities;
- validate the structure of each vignette – timing, start point, lead-in period; and

- obtain feedback on areas of the study requiring improvement or deletion.

Pilot study SMEs were polled individually for constructive commentary concerning elements of the experiment design and the data collection tool employed. In addition, SME responses were catalogued as ideal and set the standard against which participant response accuracy was measured. Feedback from the pilot study formed the briefing provided to each participant prior to commencing data collection. Several issues were identified and addressed during the pilot study. In addition, SME input provided validation of the instrument used.

First, the original design presented 20 vignettes to each participant as historic evidence suggested this number would not induce participant fatigue. In practice, however, SME feedback indicated that both attentiveness and cue responsiveness began to wane after completing 14-15 vignettes. This was understandable, as the recording time to collect data for the 20 vignettes exceeded 60 m, resulting in a participant total study commitment approaching 90 m. To reduce potential participant fatigue, the number of vignettes was reduced to 12. Initial data collection with this number proved beneficial, as study participants did not comment on undesired study duration, and cue response accuracy was acceptable. With the 12 vignettes, recorded data collection time was reduced to 34-40 m for all participants, inclusive of a short period for capturing pertinent comments.

Second, data collection-related technical difficulties were encountered. Participant think aloud protocol audio was captured commensurate with the vignette video to aid scoring. As previously noted, each vignette included various sounds heard in the flight deck deemed essential to the model fidelity. Initially, the video recording system was placed near the participant and slightly off to the side to capture the participant's monitor, under the assumption it would provide a strong audio signal above the vignette noise. While it did so, SMEs indicated the recording assembly proved to be distracting, as it was at eye level and only slightly outside direct line-of-sight. Review of the video showed placement to the side skewed the presentation, making scoring difficult. Subsequently, recording was attempted using the researcher's display and the vignette audio volume reduced to avoid interference with think aloud output. This approach, while workable, resulted in less audio fidelity presented to the participant and made subsequent data scoring difficult. The ultimate solution employed used a lavalier microphone described in the Methodology chapter. This method improved the overall signal-to-noise ratio for the recorded trials, and issues related to overshadowed think aloud outputs were closed.

Third, the pilot study confirmed that the think aloud verbal protocol used in this experiment was similar to (a) crew coordination communications encountered during a normal flight crew interaction or (b) self-talk banter often demonstrated and used in the flight instruction environment. Verbal communication and feedback, both internal to the PF and external from a PM, is common. However, varying levels of information retention within the sample were a concern. Questions arose as to a participant's ability to retain both the stable approach criteria and required actions as well as the desired output elements from the think aloud verbal protocols. To allow for this possibility, two aide memoires were used. The first provided the stable approach criteria and actions. The second provided a list of the output elements. These were discussed during the pre-trial phase, and the participant was directed to place each of them directly below the HD monitor for future reference. The participant was then informed that time between vignettes could be used to refresh on the criteria and actions, if needed.

The pilot study fortified some aspects of the research tool design. When queried about the frequency of the secondary task cues, the majority of SMEs commented both the control and thrust inceptor cue frequencies correlated to actual airplane operation. One SME noted the workload provided by the cues was possibly higher than would be expected in actual flight and consumed considerable attention. When asked if it was detrimental, the SME responded that it was not. As to the placement of the secondary task cues on the display, the SMEs indicated it did not detract from their ability to physically observe the necessary flight parameters and boundaries.

One final point of SME feedback addressed two elements of the PFD display, the content fidelity and the secondary task cue initiation time. All SMEs agreed the PFD content provided in each of the vignettes were of the correct size, color, content, and resolution found in commonly used systems. As for secondary cue initiation times, some SMEs noted the time delay from vignette initiation to first presentation of the secondary task cues. It was felt the time gap was excessive and the participant might spend too much time in wait prior to activity. This point was well received, and the benefits of changing the video to reduce the length were considered. It was determined the delay was necessary and allowed a measure of rest time and dissociation from the previous vignette. Thus, no changes to these vignette timing aspects were made.

**Demographics Results**

The data collection was conducted over a four-month period from August through December 2023, during which 25 participants were engaged at various sites. The participant sample reflected a broad range of experience, education, and skills, and was determined to be representative of the target population. Participant qualification and general- and study-specific survey data were collected during the pre-trial phase. Each participant was asked to fill out a written instrument following informed consent concurrence. Of the 25 participants who underwent the experiment, all were deemed useable, and none were removed due to either disqualification or incomplete or errant data gathered. Participants ranged in age from 19 to 69 years, with $M = 40.0$ and $SD = 19.9$, while the $Mdn = 36.0$ years, indicating a slight positive skew favoring the younger portion of the sample group. This was reflected in a .155 skew valuation. However, graphical analysis suggested a bimodal distribution with 44% between 19-25 years and 28% between 60-68 years; the remaining 28% of participant ages spanned 28-59 years.

Examining the certification levels of the sample, four (16.0%) held a Private pilot certificate, nine (36.0%) held a Commercial pilot certificate, and 12 (48.0%) held an ATP certificate. The distribution reflected a reasonable balance between the various certificate levels upon which an instrument rating could be conferred. Thus, no experience bias was expected in the overall study results. As a requirement for study participation, all held a current driver license or FAA medical certificate, possessed an instrument rating or were certificated at the ATP level, indicated either an airplane single- or multi-engine land category and class rating, and were instrument current to 14 CFR Part 61 definition in the past five years.

Categorically, the highest levels of academic education obtained were reported as one (4.0%) at the high school level, five (20.0%) having some undergraduate study, and nine (36.0%) holding an undergraduate degree. Another three (12.0%) indicated some graduate study, while six (24.0%) had been conferred a graduate degree, and one (4.0%) claimed having completed some amount of postgraduate study.

The type of flight operation under which the participant was currently employed or engaged was also considered. Four categories were identified for participant selection: training, corporate, commercial air carrier, and military. The demographic results indicated 18 (72.0%) engaged within a training operation under Parts 61/141/142/Other, three (12.0%) executing corporate flight operations, and four (16.0%) working within commercial air carrier operations. No military flight operation participants were used.

Participants were queried as to their level of flight experience. Values included total years of flight experience, total flight time, total instrument flight time, and total number of instrument approaches flown. A summary of participant experience is presented in Table 7. The results indicated a broad diversity in terms of years of flight experience, exposure to simulated or actual instrument flight conditions, and the execution of instrument approach events. Once again, the presence of extreme values for flight-rated hours and instrument events reflected the bimodal sample characteristic and the positive skew. The number of high values at more experienced pilot mode, compared with the relatively small values associated with the low experience pilot mode, impacted the *SD* of each measure such that they exceeded the *M* values. It should be noted that the sample was nearly perfectly split between low- and high-experienced pilots, with 13 of the 25 participants possessing less than 1,500 hrs total flight time and the remaining 12 reporting higher values. This allowed for qualitative observations between the two levels of experience.

**Table 7**

*Participant Flight Experience*

| Measure | *N* | *M* (*SD*) | *Mdn* | Min | Max |
|---|---|---|---|---|---|
| YRS | 25 | 20.9 (17.8) | 18.0 | 2 | 55 |
| TFT | 25 | 4,237.4 (4,963.9) | 1,900.0 | 140 | 16,000 |
| IFT | 25 | 739.7 (1,051.7) | 120.0 | 50 | 4,500 |
| IAF | 25 | 1,272.3 (2,157.0) | 200.0 | 30 | 10,000 |

*Note*. Descriptive statistics for participant flight experience, where *N* is the number of data points, *M* is the mean value, *SD* is the standard deviation, and *Mdn*, Min, and Max indicate the median, minimum, and maximum values, respectively. Measure descriptions YRS = Cumulative flight experience in years; TFT = Cumulative flight time in hours; IFT = Cumulative instrument flight time in hours; IAF = Total instrument approaches conducted in numeric count.

Previous experience with PFD, ND, EFIS, or other electronic types of instrument displays meeting the requirements of a TAA (FAA, 2021a) was a criterion for entry in the study. Study participants were asked as to their degree of experience with such displays. Flight time using electronic instrument displays ranged from 90 to over 13,000 with an average of 2,638.4 total hrs (*SD* = 3,644.6) and a median electronic flight display flight experience of 750.0 hrs. Display experiences were those found in typical training aircraft, such as the Garmin 1000, as well as those in use in the Boeing, Airbus, Dassault, and various military platforms.

Equally, prior experience with stable approach criteria was an area of interest for which data were collected. No definition as to what defined the term *criteria* was provided in the survey, although the participant was provided general guidelines indicating some type of documented criteria. Prior experience using criteria ranged from 0 to 37 years with an average of 12.1 years (*SD* = 11.5) and a median period of 7.0. Average instances of having found oneself in an unstable approach under established criteria was 44.3 (*SD* = 72.8) events with the *Mdn* = 20.0. The average number of months since the most recent unstable approach experience was 25.3 (*SD* = 35.0) with the *Mdn* = 12. Those participants having experienced a recent unstable approach characterized the primary indication as follows: six (24.0%) excessive airspeed, three (12.0%) insufficient airspeed, four (16.0%) excessive descent rate, one (4.0%) inappropriate thrust setting, zero (0.0%) incorrect airplane configuration, one (4.0%) incomplete checklist/briefing actions, and 10 (40.0%) flight path deviation. Actions taken upon awareness of the most recent unstable approach included 16 (64.0%) continued the approach to a successful landing, one (4.0%) continued the approach to a less-than-successful landing, three (12.0%) continued the approach and later executed a go-around, and four (16.0%) executed an immediate go-around. One participant responded to the question as “Not Applicable,” even as an unstable approach experience was noted. This suggested one of the alternatives provided in the survey did not directly match the scenario.

**Descriptive Statistics**

Study participants were presented a series of vignettes in which they were asked to role relate to a pilot flying an ILS approach. Each participant was presented the same set of 12 approach vignettes for scoring. During presentation of the flight profile, video and audio methods were used to gather measures of RA to real-time unstable approach presentations.

Data scoring occurred over a consecutive four-day period to minimize the possibility of inconsistent assessment and valuation. The *M* and *SD* for the RA DV were calculated. The RA results are presented in Table 8. Values for RA were measured as participant correct responses raw score values, provided the response was received within

5 s of the criteria being exceeded. Each participant viewed the vignettes and offered think aloud comments concerning the presence of an unstable condition, the criteria being exceeded, and the action to be undertaken. Model responses for each vignette were established by SME review during the pilot study and used to assess participant RA. Incorrect responses were neither counted nor applied against the correct responses selected, nor were responses provided incorrectly when unstable conditions were not present.

The 12 vignettes were created using a core of three videos. Each was modified to provide a baseline, boundaries, alert, and boundaries with alert treatment. In doing so, the total number of model responses across all scored vignettes was consistent. The three vignettes varied in complexity from high to low, having 10, four, and three unstable events, respectively. This allowed a total of 17 unstable events for each collection of treatments and a potential for 51 points, provided all the required responses sought were observed during unstable events. Individual participant performance in vocalizing the unstable condition was then measured using the scale value. Inherent interactions resulting from an unstable event introduced compound events with some overlap in the presentation of each.

Of the total points available for RA, the *M* values indicated slightly less than half the possible SA information was identified by the participant. The resulting RA *M* scores supported increased pilot SA when any of the display enhancements were applied over the baseline Treatment. The treatments ranked in order of RA score *M* value from Treatment 1 to Treatment 4. The highest RA *M* scores were achieved when the alert enhancement was applied, either in isolation or in combination with the use of boundaries, for Treatments 3 and 4, respectively. With the exception of Treatment 2, the resulting *SD* decreased even as the *M* values increased for each treatment. This suggested pilots not only experienced enhanced SA from the treatments, but also exhibited less variance in SA scores across the experiment sample. Pilots would be more likely to identify an unstable condition with greater RA consistency if the enhancements were present.

**Table 8**

*Mean Scores and Standard Deviation – RA*

| Treatment | DV | *M* | *SD* | *N* |
|---|---|---|---|---|
| 1 | RA | 20.32 | 10.209 | 25 |
| 2 | RA | 23.08 | 10.969 | 25 |
| 3 | RA | 23.96 | 9.594 | 25 |
| 4 | RA | 24.96 | 9.195 | 25 |

*Note*. Mean scores (*M*), standard deviations (*SD*), and sample size (*N*) for dependent variables response accuracy (RA). RA is measured using a continuous ratio variable count of the total assessed score. Factorial design treatments are defined as follows: (1) baseline control with no enhancements, (2) treatment with only boundaries present, (3) treatment with only alerts present, and (4) treatment with boundaries and alerts present.

## Assumption Testing Results

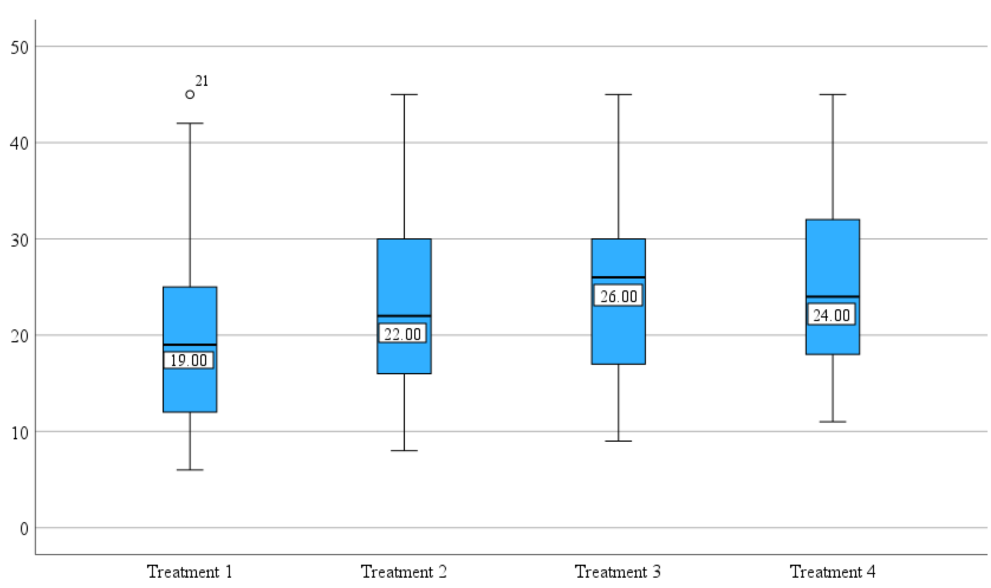
Assumptions were made in support of two-way, repeated-measures factorial ANOVA analysis and associated parametric testing used to infer the results and conclusion for the study. A number of these assumptions were met by study design. The assessable assumptions included the absence of significant outliers, normality of the DV data distribution, and homogeneity and sphericity of the DV variance. The data were assessed for compliance using the tools provided within SPSS and verified through simple spreadsheet calculations.

***Absence of Significant Outliers***

Visual inspection of the boxplots for RA values indicated a single case in which the total RA scores data value fell outside the upper fence for the baseline Treatment 1, where neither boundaries nor an alert was present and was therefore categorized by SPSS as an outlier. A boxplot of total scores data generated for each group is presented in Figure 12. There were no outliers present in the total events boxplot data for all other treatments. Median score values are provided for each treatment. The data set for the specific participant was reevaluated to verify accuracy of the value. Review of the recorded data reflected the value assigned during data scoring was consistent and measurement error did not occur. Data entry error was subsequently checked and found not to have occurred.

**Figure 12**

*Boxplot of Scores Data Distribution and Outlier*



The SPSS software was used to further assess the presence of statistical outliers. Values for the studentized residuals for each treatment were collected. Investigation of the RA score residuals for each of the participants and treatments indicated none of the values exceeded the ±3 measure for the presence of an outlier. The participant determined an outlier for RA scores on the baseline treatment indicated a studentized residual value of 2.47, which reflected the near-perfect response of 45 of 51 possible points. Examining the participant's other results, it was apparent the performance was a generally unusual case, the performance on the other treatments being 1.20, .43, and .78 for Treatment 2, Treatment 3, and Treatment 4, respectively. However, the participant performance on the complex scenario was the greatest contributor to the score, and examining the assigned Latin Square progression, this particular vignette occurred in the latter half of the test, following two previous iterations of the core video. This suggested the possibility of learning effect influencing the data but was dismissed when comparing the placement of the other treatment set vignettes where the residuals were not as strong. While inclusion might contribute to a Type II error, the determination was made to include the data for this participant.

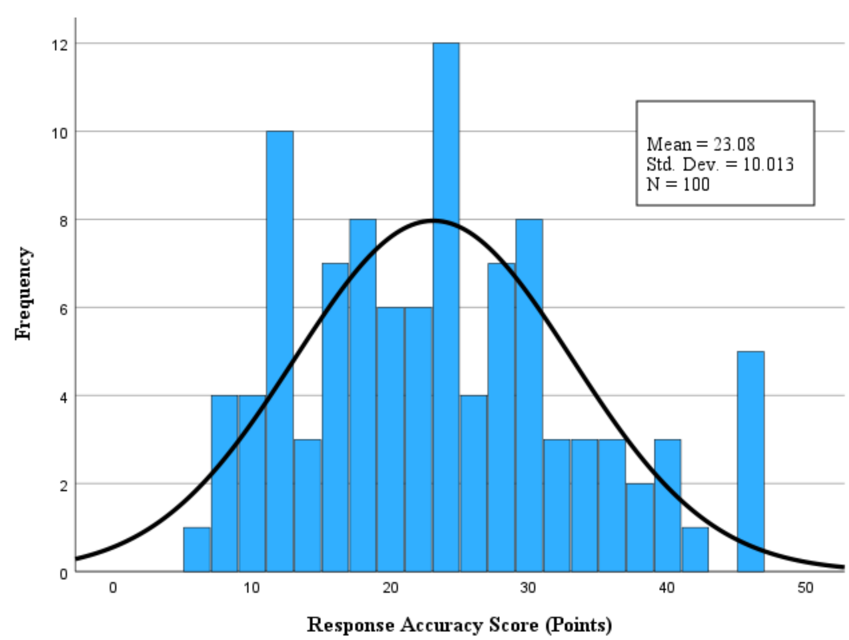
The RA data values were complete with no missing values. However, the results from the post-trial usability survey had two instances of MCAR data. In one case, the participant failed to notice the back side of the survey and did not provide data for the two questions addressing the perception and predictive value of the treatments. The other case resulted from the participant failing to properly indicate the order of preference for each of the treatments. Failure to provide the data was determined to be MCAR due to participant error. For this MCAR data, no effort to impute scores was pursued, and the data were coded in SPSS as missing values and identified in the SPSS Missing Values feature when calculating descriptive statistics for the survey elements.

### *Normality*

The values for the DV were assessed for normality using both graphical analysis and statistical testing methods. The frequency of RA values for total SA scores drawn from the 12 scored vignettes were plotted as a function of the score measured. The histogram for the totality of RA scores, N = 100, is presented in Figure 13.

**Figure 13**

*Histogram for Response Accuracy*

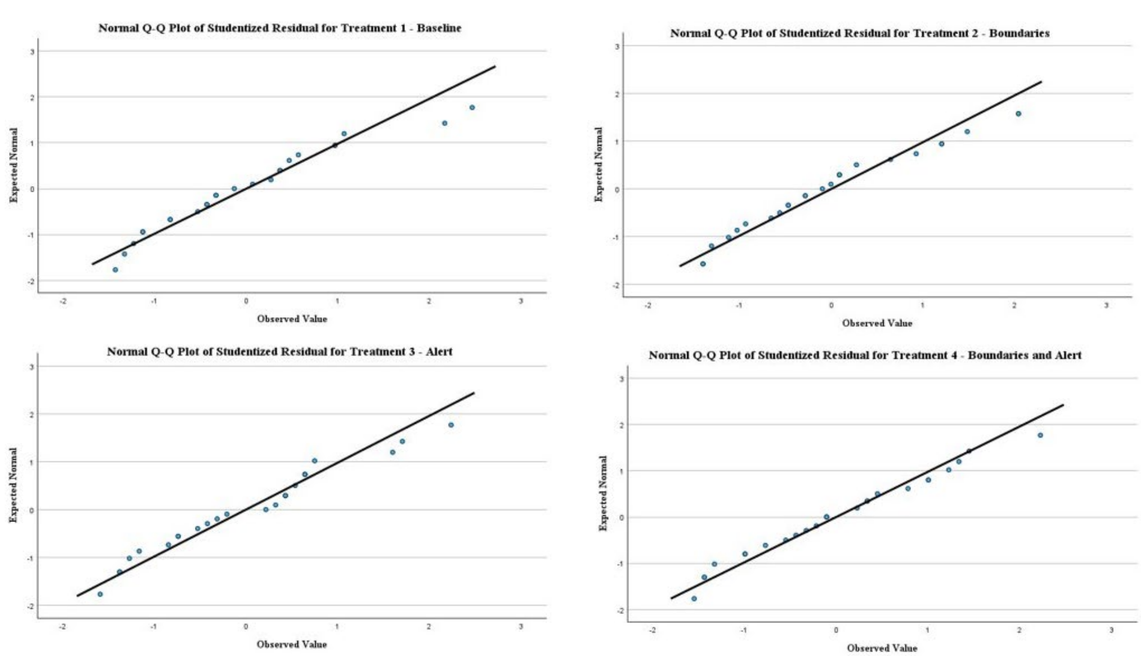



The general shape of the RA histogram of the aggregated 100 valid RA scores suggested an underlying near-normal leptokurtic shape with a slight positive skew, which was reflected in the kurtosis statistic of -.492 (with a standard error of .478), a skew statistic of .411 (with a standard error of .241). While the data indicated the majority of scores falling within a normal distribution, the small number led to an overall numeric assessment of positive kurtosis. The moderate negative skew statistic likely resulted from the rather significant RA frequency at approximately 45 points. Both square root and $\log_{10}$ transformations were pursued, and while the square root improved the skewness to -.028, it also raised the kurtosis value to -.693, an undesired change. Plots of the studentized residuals for each treatment were then examined, and each followed an approximate normal shape.

Field (2013) notes that stealth values hidden within such data summations can provide a false sense of normality. As such, it is imperative that the data for each of the treatments be investigated individually. Four Q-Q plots for each of the treatments supplied by the SPSS Analyze Descriptive Statistics feature are presented in Figure 14.

**Figure 14**

*Q-Q Plots for Normality*



The data indicated similar scatter patterns across the separate treatments. The right skew of the data was apparent in data points placement at the right of the reference line for all the treatments tested. Further, there existed a slight heavy-tailed nature, as the values deviated from the mean reference. The Q-Q plots supported a determination of data normality for each of the treatments. The boxplots shown in Figure 12 also supported this position, in that while uneven whiskers favored the higher values, to indicate an asymmetric shape to the data distribution, the absolute length of each IQR box was consistent, and the mean and median values were relatively close.

The Kolmogorov-Smirnov and Shapiro-Wilk tests were used to measure data normality. The results of these SPSS test are shown in Table 9. The Shapiro-Wilk test resulted in the significance levels, as shown. The significance values for the treatment RA scores indicated values greater than $p > .05$. Thus, the tests for each treatment were not statistically significant, and the sample data were acceptably normal in distribution.

**Table 9**

*Kolmogorov-Smirnov and Shapiro-Wilk Tests – RA*

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| Treatment | Statistic | df | Sig | Statistic | df | Sig |
| 1 | .107 | 25 | .200* | .943 | 25 | .170 |
| 2 | .112 | 25 | .200* | .955 | 25 | .328 |
| 3 | .147 | 25 | .174 | .942 | 25 | .162 |
| 4 | .102 | 25 | .200* | .968 | 25 | .599 |

*Note*. The * indicates a lower bound for true significance. The a denotes Lilliefors Significance Correction.

### *Homogeneity and Sphericity of the Variances*

A cursory examination of the boxplots showed consistency in the spread of data, suggesting homogeneity. Field (2013) and Truong (2017) offer an analysis of the scatterplots of the values of the residuals against the outcome predicted by the model. If homogeneity holds true, no observed system relationship should be indicated between the predicted values and the errors in the model. Plots of standardized residuals against predicted values indicated no systematic relationship between the error and the model. Levene's Test of Equality of Variances was completed between the four treatments and scores for each treatment. The test compared variances based on the median score between each treatment group for equality. The test showed homogeneity of the variances was present, as measured against the test value $p < .05$. For score values across the four treatments, the test was non-significant, and the assumption was met, $F(3, 96) = .181$, $p = .909$.

**Reliability and Validity Testing Results**

Study internal reliability was facilitated through a number of data collection and scoring methods. All data were scored by one individual. As previously mentioned, recorded data allowed for the analysis of the data in a concentrated period of four days, the detailed review of that data in both real-time and reduced frame rates, and the subsequent revisit of each participant's performance following the scoring period. A conservative approach to the allocation of points was taken, providing points for perception of the unstable condition, awareness of the nature of the instability, and the action to be taken when actually spoken aloud. No allowance for inference of the three required outputs was granted across all participants. Allowance was, however, granted for minor variation of the exact wording spoken aloud by the participant, underscoring that not all pilots communicate in the same fashion, even though the essential message is effectively conveyed.

Instrument validity concerns followed the decision to reduce the number of vignettes presented from 20 to 12. At issue was whether the participant would easily discern the fact that only three core flight profiles were used with each treatment applied to the core to create the 12 total vignettes, thus affecting the results. To assess the merits of that concern, eight participants – approximately 30% of the study sample – were selected at various points throughout the study for feedback. These participants demonstrated medium- to high-degrees of attentiveness and RA values during the within-trial phase, suggesting the possibility of detecting patterns. The participants were queried following completion of all post-trial data collection as to how many unique vignettes were employed. To provide clarity, they were asked: "Was there any duplication in the flight profiles of the vignettes presented?"

Of the eight, four participants stated there was no duplication. Three participants commented that while there were a number that appeared similar, none were the same. Finally, one participant stated that prior to initiating within-trial data collection the participant had posited it would be difficult to compare the enhancements without using a common profile, but further noted that there was no recall of an instance of pattern recognition. This suggested that the combination of the secondary task demands and participant immersion in the vignette visual and audio experience prevented detection.

**Data Scoring External Audit**

The reliability of the study results was dependent upon the consistency with which the data were scored between participants, as well as within the vignettes of each participant. While absolute DV RA values had meaning in this study and reflected on the individual participant, relative DV RA values were of greatest significance. Scoring was limited to one individual, the researcher, and thus interrater reliability issues were avoided.

Inconsistency in scoring each set of vignette responses was avoided through four actions. First, defined criteria for scoring the presence of the three elements of RA data sought were established. Second, scoring of all participant vignette responses was accomplished over a limited, uninterrupted period. This minimized possible time-related variation in how the criteria were applied by the researcher. Third, the vignette response set for each participant was scored entirely before moving to the next participant set. Finally, the entire experiment was captured on video, and the researcher scored a sampling of vignettes multiple times and revisited scores, when necessary, to ensure internal consistency.

To ensure scoring was consistent with the criteria, an external nonstatistical audit was accomplished by impartial third parties. A nonstatistical method was chosen as a low risk of scoring differences existed due to strong internal scoring controls using verified event timeline scripts. Colbert (2011) notes a nonstatistical audit may be more appropriate than a statistical audit when there are a small number of transactions or possible substantive errors can’t be easily identified. Two current air carrier pilots operating Boeing B-737 airplanes under Title 14 CFR Part 121 were tasked to score a sample of 24 of the 300 individual vignettes chosen through judgement sampling. Each auditor held Doctor of Philosophy degrees in various disciplines, had been trained as internal raters within their respective air carriers, and were not affiliated with the study. The external audit scoring results were then assessed against researcher scoring. These reviews indicated a consistent application of the scoring methodology and concurrence with the scores assessed for each vignette.

***Interaction Effect***

A two-way, repeated-measures ANOVA was completed for the interaction between boundaries and alerts, as measured against the baseline treatment. The results indicated no significant interaction between the boundaries and alert factors, suggesting the two treatments did not interact in some fashion, $F(1, 24) = .646$, $p = .429$, partial $\eta^2 = .026$. The partial $\eta^2$ of.026 suggested a low effect size against the accepted .01 value (Field, 2013; Kirk, 1996) for the interaction and thus was not considered significant to the study outcome. As the interaction between boundaries and the alert was insignificant, the effect of an alert on participant RA scores was not dependent on the presence of boundaries; that is, it did not matter whether or not an alert was displayed. Thus, the interaction provided no discernable gain or detriment to the individual treatments when used in combination. Any potential concern with interference, clutter, amplification, or other display aspect was unfounded. This did not negate concurrent use of both treatments, as the *M* scores suggested better performance when doing so, but rather suggested that doing so was inconsequential in terms of impact to each other.

***Main Effect of Boundaries***

The use of boundaries provided the participant a visual presentation of the values used to define a stable approach. A two-way, repeated-measures ANOVA was completed for the main effect of boundaries against the baseline treatment. In instances where boundaries were applied, the mean scores for RA were greater than that of the baseline condition. The results indicated a statistically significant main effect, $F(1, 24) = 6.563$, *p*

= .017, partial $\eta^2$ = .215. The partial $\eta^2$ indicated a large effect size as measured against the suggested .14 value (Kirk, 1996; Field, 2013) and had a significant effect on the study outcome.

***Main Effect of Alerts***

The use of alerts provided the participant a visual indication as to when one of the values used to define a stable approach were exceeded, without specifically noting which criteria was exceeded. A two-way, repeated-measures ANOVA was completed for the main effect of alerts against the baseline treatment. The results indicated a statistically significant main effect, $F(1, 24) = 4.247$, $p = .050$, partial $\eta^2$ = .150. The partial $\eta^2$ again indicated a large effect size as measured against the suggested .14 value (Field, 2013; Kirk, 1996) and also had a significant effect on the study outcome.

**Hypothesis Testing Results**

A single research question was posed for this study: Is pilot SA, when engaged in an unstable approach, affected by applying enhancements to the flight instrument displays? From this question, a total of six hypotheses, each presenting a corresponding null and alternative hypothesis, were proposed. They were:

- Hypothesis 1: Pilots provided an enhanced flight display employing stability criteria boundaries exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented no enhanced flight displays.

- Hypothesis 2: Pilots provided an enhanced flight display employing an alert message exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented no enhanced flight displays.

- Hypothesis 3: Pilots provided an enhanced flight display employing stability criteria boundaries exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented an enhanced flight display employing an alert message.
- Hypothesis 4: Pilots provided an enhanced flight display employing both stability criteria boundaries and an alert message exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented no enhanced flight displays.
- Hypothesis 5: Pilots provided an enhanced flight display employing both stability criteria boundaries and an alert message exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented an enhanced flight display employing an alert message.
- Hypothesis 6: Pilots provided an enhanced flight display employing both stability criteria boundaries and an alert message exhibit no significant difference in RA on SA queries related to detection of an unstable approach condition than pilots who are presented an enhanced flight display employing stability criteria boundaries.

Data for values of participant RA mean scores for each treatment were collected and analyzed. Tests were conducted to determine whether differences in the RA means for each sample were larger than that which would occur in the population, if the means were actually equal (Truong, 2017). The ANOVA test accomplished this by examining the variances of the data for each treatment group (Truong, 2017). The assessment of the hypotheses is presented in Table 10. The results indicated that the presence of either the boundary or alert enhancements benefited participant SA. Thus, Treatments 2 and 3 provided SA perception and cognitive gain to the participant. This outcome aligned with previously presented values for *M* and *SD* which showed the RA scores improved with enhancement application. The *M* values allowed an inferred ascending order of improvement with each of the treatments: Treatment 1, followed by Treatment 2, then Treatment 3, and finally the combination of boundaries and an alert, Treatment 4.

The test results indicated statistical significance and thus were very unlikely to have occurred by chance (Truong, 2017). Of note, four of the hypotheses, $H_03_{RA}$ through $H_06_{RA}$, were not assessed due to a lack of statistically significant interaction effects. However, merit should be given to the operational significance of the difference in the *M* and *SD* values for those treatments.

**Table 10**

*Hypothesis Testing Results – RA*

| Hypothesis | *F-ratio* | Outcome |
|---|---|---|
| $H_0 1_{RA}$: $\mu_{enhanced\ (boundaries)} = \mu_{unenhanced\ (baseline)}$<br>$H_A 1_{RA}$: $\mu_{enhanced\ (boundaries)} \neq \mu_{unenhanced\ (baseline)}$ | $F(1, 24) = 6.563, p = .017$ | Reject |
| $H_0 2_{RA}$: $\mu_{enhanced\ (alert)} = \mu_{unenhanced\ (baseline)}$<br>$H_A 2_{RA}$: $\mu_{enhanced\ (alert)} \neq \mu_{unenhanced\ (baseline)}$ | $F(1, 24) = 4.247, p = .050$ | Reject |
| $H_0 3_{RA}$: $\mu_{enhanced\ (boundaries)} = \mu_{enhanced\ (alert)}$<br>$H_A 3_{RA}$: $\mu_{enhanced\ (boundaries)} \neq \mu_{enhanced\ (alert)}$ | No interaction | n/a |
| $H_0 4_{RA}$: $\mu_{enhanced\ (boundaries\ +\ alert)} = \mu_{unenhanced\ (baseline)}$<br>$H_A 4_{RA}$: $\mu_{enhanced\ (boundaries\ +\ alert)} \neq \mu_{unenhanced\ (baseline)}$ | No interaction | n/a |
| $H_0 5_{RA}$: $\mu_{enhanced\ (boundaries\ +\ alert)} = \mu_{enhanced\ (alert)}$<br>$H_A 5_{RA}$: $\mu_{enhanced\ (boundaries\ +\ alert)} \neq \mu_{enhanced\ (alert)}$ | No interaction | n/a |
| $H_0 6_{RA}$: $\mu_{enhanced\ (boundaries\ +\ alert)} = \mu_{enhanced\ (boundaries)}$<br>$H_A 6_{RA}$: $\mu_{enhanced\ (boundaries\ +\ alert)} \neq \mu_{enhanced\ (boundaries)}$ | No interaction | n/a |

*Note*. F-ratio values for within-subject effects for each treatment. Lack of significant interaction between boundaries and the alert made testing hypotheses $H_0 3_{RA}$, $H_0 4_{RA}$, $H_0 5_{RA}$, and $H_0 6_{RA}$ unnecessary. The use of n/a denotes not applicable.

Overall, the statistical analysis indicated that there was improved pilot SA with the boundaries and alert enhancements, and there was no significant interaction between the specific enhancements. Thus, interactions following from having combined boundaries and an alert, as generated in Treatment 4, were not observed in a significant manner. Use of a combined presentation added value from an operational standpoint. While the test results lacked statistical significance, both applied in combination indicated better SA responses.

## Summary

The study was conducted as to whether pilot SA, when presented an unstable approach, was affected by applying flight instrument display enhancements. A quantitative 2x2 factorial experiment design was used to analyze the data. No data were found to be missing in the RA score values. One outlier was identified in SPSS, and it was isolated to the baseline Treatment 1. Studentized residual values supported retention of the data point. The data were found to be normal and linear, as evidenced both graphically and through SPSS exploration. Main effects indicated boundaries and alerts impacted the outcome of the test and there was no significant interaction between each of the treatments. Of the six hypotheses presented to address RA measurement of pilot SA when presented an unstable approach condition under the various treatments, the two for main effects were rejected. However, lack of interaction between boundaries and the alert prevented further testing of hypotheses $H_03_{RA}$, $H_04_{RA}$, $H_05_{RA}$, and $H_06_{RA}$.

# Chapter V: Discussion, Conclusions, and Recommendations

The purpose of this study was to conduct a quantitative 2x2 factorial repeated-measures experiment on a sample of low- and high-experience instrument rated pilots to determine whether an improvement was achieved in providing various treatments to the baseline flight instrument display. The results of the experiment demonstrated the effect each display enhancement, and the combination thereof, had on pilot RA scores. Given the experiment data collected, it is clear that an enhanced PFD design, inclusive of unstable approach boundaries and/or alerts to indicate the presence of an unstable approach condition, will enhance pilot SA. It informs not only the presence of an unstable approach, but the nature of the unstable condition as well. Thus, the inclusion of some type of display enhancement was better than no enhancement at all. While the results answered the core research question and indicated two of the treatments were beneficial in improving pilot SA, there exist a number of topics worthy of discussion.

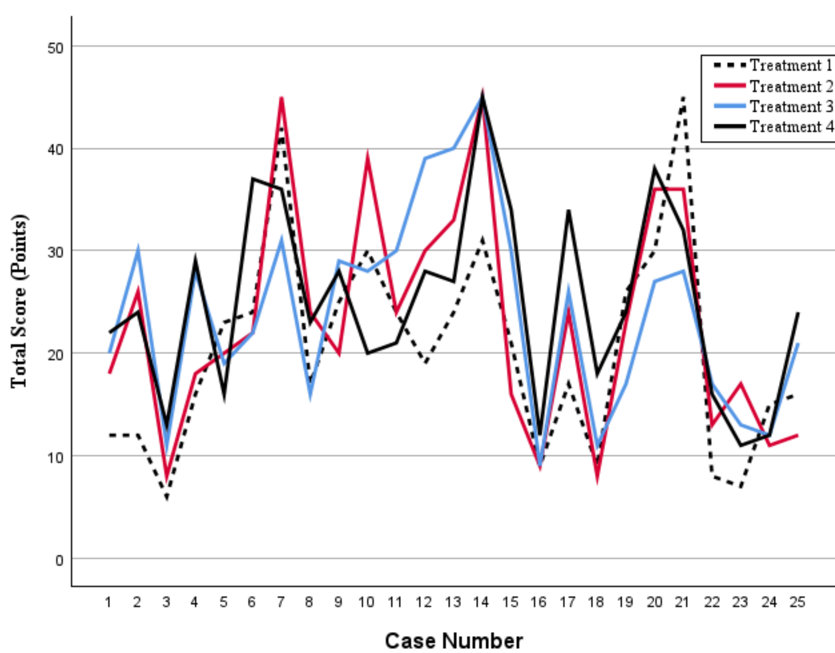
When this research topic was presented to a number of flight technical pilots, safety pilots, and flight instructors from various organizations, the response was typically lukewarm. While each pilot had an opinion as to why it was felt the presentation of unstable approach criteria tools would or wouldn't benefit pilots, none could point to empirical data to support that position. Concerns about inclusion of such display enhancements centered on the feeling (a) such presentations would add no real value or (b) including such presentations would only serve to clutter the PFD. They tendered the belief there would be no benefit to their inclusion, even as some manufacturers were considering the merit of real-time or predictive systems.

**Discussion**

Opinions are easy to form when it comes to instrument scan pilot skills and tend to be egocentric in perspective. By regulation, training programs developed to prepare pilots for instrument flight operations include the fundamentals. The measure of a successful instrument scan may be limited to an assessment of the approach outcome, rather than the accuracy of the scan. The data from this experiment indicated those positions may be unfounded in that there was a significant spread in the mean RA scores resulting from each treatment. Minimum and maximum RA score values across all participants are best visualized using a line plot of scores by participant, as shown in Figure 15.

**Figure 15**

*Total Score by Participant*



In general, scoring between the enhanced treatments correlated for each participant in terms of both magnitude and direction. With minor exception, those participants who tended to score higher on RA scores than the average for all participants did so across all four treatments; those who tended to score lower compared to the average for all participants followed suit on the opposite side of the spectrum for all treatments. Participant RA scores did not demonstrate large individual treatment slope sign differences from one participant to the next, further supporting correlation.

As seen in Figure 15, the baseline Treatment 1 individual SA scores consistently underperformed the other treatments, in terms of SA support. In only one case did Treatment 1 result in a higher individual SA score than the other treatments. The boundaries enhancement, Treatment 2, provided better unstable approach SA support than Treatment 1, as indicated by consistently higher participant RA scores. Treatments 3 and 4 also indicated an improvement over Treatment 1 for nearly all participants. However, the relative order of improvement for each of the enhanced display treatments varied rather inconsistently between the participants. For example, the combination Treatment 4 indicated an overall strong support for pilot SA and did so to a greater extent than the other treatments among those exhibiting lower levels of SA. Less so was the benefit realized with higher scoring participants, who appeared to prefer the Treatment 2 or 3 presentation over that of Treatment 4.

The results followed the theory which suggested that providing the pilot increased information as to an unstable approach would result in cognitive easing and aid SA. This outcome countered parties queried prior to the experiment who postulated the treatments would prove more detrimental than good, and thus actually diminish pilot SA. It was unknown at the beginning of the study whether the application of any one particular enhancement would be better than another, in comparison to a standard PFD or in comparison with each other, or whether the simultaneous presence of both treatments would be excessive in all cases. These unknowns were resolved in the study results.

Boundaries alone allowed the pilot to self-determine the presence of an unstable condition but required vigilant attention to the bound demarcations. Timing of the crosscheck gaze and subsequent visual saccade through the various path and performance parameters might account for some of the difference. The alert, on the other hand, provided the pilot a measure of SA in the visual announcement of an unstable condition. But barring the pilot's ability to ascertain the source, the alert alone might leave the pilot with the knowledge of an altitude-dependent action required, such as a go-around, but no real understanding as to why to execute that action. This could impact the safety of that maneuver, such as a low energy state resulting from unstable airspeed and the action of pitching up to commence a go-around. An approach to stall condition may be the undesired outcome of that action. Hence, it is critical to know why the approach is unstable prior to a corrective action. In that sense, it is still postulated the combination treatment of both boundaries and alerts might further that fundamental understanding. Unfortunately, this possibility was unable to be tested, as the research sought merely the action to be undertaken and not the determination as to whether that action would be prudent in the given airplane state.

***Event Occurrence***

By design, the vignettes presented a number of events, some complex in which there were combined events and were presented in rapid succession. Others were more benign, comprised of moderate or less unstable event complexity. The expectation during experiment conceptualization and vignette scripting was that most participants would fall short of a perfect SA score. The data supported that expectation, as indicated by the resulting total RA scores, approximating 50% of the available total for the complex scenario.

As analysis progressed, it became evident that while the RA *score* reflected on the overall SA of the participant, equally valuable to measuring improved SA might be data indicating the number of *events* each participant observed. Data was extracted to analyze the events in which the participant identified the unstable condition, regardless of any subsequent think aloud protocol information provided. In this instance, inference was allowed when it was apparent that the participant voiced at least one data element. Such a determination is of great value: Researchers, industry, and regulators seeking solutions to continued flight in unstable conditions would be asking whether the presence of the treatment triggered the pilot to perceive an unstable event, even when they may not express cognition and prediction of the current and future state. The RA event results are shown in Table 11.

**Table 11**

*Mean Events and Standard Deviation – RA*

| Treatment | DV | *M* | *SD* | *N* |
|---|---|---|---|---|
| 1 | RA | 7.40 | 3.428 | 25 |
| 2 | RA | 8.32 | 3.400 | 25 |
| 3 | RA | 9.20 | 3.109 | 25 |
| 4 | RA | 9.48 | 3.016 | 25 |

*Note*. Mean events (*M*), standard deviations (*SD*), and sample size (*N*) for dependent variables response accuracy (RA). RA is measured using a continuous ratio variable count of the total events recognized. Factorial design treatments are defined as follows: (1) baseline control with no enhancements, (2) treatment with only boundaries present, (3) treatment with only alert present, and (4) treatment with boundaries and alerts present.

Mean RA event results showed no difference from mean RA scores in terms of the ranking of Treatment 1 and all other treatments. The events *SD* values across the treatments improved in similar fashion to RA scores, though the *SD* for Treatment 2 was improved in ranking over Treatment 1. The highest *M* and lowest *SD* were once again attributed to Treatment 4, indicating higher levels of event detection and greater consistency in recognition across the sample. The use of display enhancements again supported improved participant SA when considering event *M* and *SD* values.

Means differences for the events were subjected to ANOVA testing, and the results paralleled those of the RA scores in terms of the comparisons between the baseline Treatment 1 and Treatments 2 and 3, furthering the validity of the previous RA scores results. Treatment 2 indicated a strong main effect for the application of boundaries when compared to Treatment 1, $F(1, 24) = .5.333$, $p = .030$, partial $\eta^2 = .182$. Treatment 3 indicated a very strong main effect for the application of an alert when compared to Treatment 1, $F(1, 24) = 10.621$, $p = .003$, partial $\eta^2 = .307$. Comparing

Treatment 3 to Treatment 1, the mean difference proved to be insignificant, $F(1, 24) = .891$, $p = .355$, partial $\eta^2 = .036$, indicating lack of interaction between the boundaries and alert when considering unstable approach events. However, the *M* values indicate the combination of boundaries and the alert were found to be beneficial when recognition of unstable events was considered. Even though the test results lacked statistical significance, both applied in combination should provide better responses.

Pilots and air carrier operations would perhaps be more concerned in knowing which treatment captured the greatest number of unstable *events*, as opposed to which may have afforded the most comprehensive level of SA. Indeed, two well-recognized and understood unstable events with the correct action followed may be less beneficial than four events being recognized, even if an improper understanding was present. Indeed, close review of each participant's RA score tallies showed that in certain cases, participants were able to perceive the presence of an unstable condition but were unable to determine, or voice a determination of, the criteria exceeded or the action to follow.

***Missed Cases***

Not all unstable approach events were recognized during the vignettes. Signal Detection Theory categorizes the states of the world into one of four quadrants: hits, misses, false alarms, and correct rejections (Wickens et al., 2014). Previous discussion has centered on those states considered to be hits, where the participant correctly identified an unstable event. The opposite end of the spectrum, correct rejections, carried little relevance to this study, as non-response to a non-event is trite. One other state, those deemed to have been instances of missed perception and an unstable event, are of value. Failure to perceive, understand, and act on those cases can bring tragic results.

In total, the vignettes provided for 68 unstable events presented to each participant, for a total of 1,700 events across all 25 participants. These were allocated as follows: 40 events to the complex A/D/G/I vignettes, 16 events to the moderate B/E/H/K vignettes, and 12 events to the simple C/F/I/L vignettes. Across the participant sample, there were respectively 587, 159, and 95 instances in which no score was recorded for the event. This meant 58.00% of the events were missed in the complex scenarios and 39.75% and 31.67% for the moderate and simple complexity scenarios, respectively. These values were significantly higher than would be expected. The high perception miss rate for the complex scenario drove the overall valuation. The rather poor rate was not unexpected, as the frequency of the events and the inclusion of compound events contributed to the likelihood of missed unstable condition alarms.

Moreover, there were 89 cases, or 5.2%, where the participant was able to perceive the unstable condition and either could not identify the criteria missed or the action to be executed. One or the other was missing. These events favored the identification of the criteria missed over the action to follow. Anecdotally, this small fraction of the total events occurred when the participant appeared to be saturated in terms of vocalizing the action to be followed. Still, instances where an action was called out without identification of the criteria exceeded did occur but were rare.

Most interesting were the 32 cases, 1.8% of the total, where the participant was only able to perceive the unstable condition. In 87.5% of these cases the alert, either in isolation or in combination with boundaries, was present. This highlights the benefit derived from the presence of the alert. While boundaries were indicated as more favorable in enhancing SA, in a few isolated cases, the alert served as the last defense in an otherwise unrecognized condition. During an approach, pilots can experience a number of physical and environmental challenges – individually or concurrently – that can deplete available mental resources and thus compromise the ability to affect a successful instrument scan. This can occur perhaps to the point of near-complete saturation. Even when faced with such a depleted state, the rather bold, amber ALERT cue stands out as a beacon to the pilot undergoing that fate.

***False Alarm Cases***

Data was collected on the occurrence of cases in which participants indicated the presence of an unstable condition, but did so when none actually was presented. Wickens et al. (2004) term this a false alarm within the Signal Detection Theory framework. Such instances are a failsafe position, where the outcome of an inappropriate determination is the acceptable correction to the condition for which one is actually not required or go-around procedure when none is actually necessary. Research by the Flight Safety Foundation (FSF, 2017) would imply that the execution of an unnecessary go-around may, in fact, be detrimental, and the continuation of an unstable approach is actually preferred. This is worthy of consideration as the industry experiences incidents and accidents involving proper completion of the go-around procedure (FSF, 2017). If this is true, the occurrence of a false alarm is undesirable and should be avoided in providing enhanced displays.

There was a total of 121 false alarm cases across the entire data set. These false alarms occurred at a rate 7% of the actual unstable conditions. The incidence of false alarms was associated as follows: Treatment 1 (35 cases), Treatment 2 (26 cases), Treatment 3 (31 cases), and Treatment 4 (29 cases). The data suggest that false alarms were more likely when no boundaries were present, either alone or in combination with the alert, supporting the presentation of boundaries as a clear demarcation for validation of the condition. Surprising was the higher occurrence when only the alert was present. It would be expected that such a presentation would condition the participant to rely upon its display to indicate such conditions, knowing that system design would prevent mislabeling of an unstable condition. And yet this treatment presented nearly as high a false alarm rate as the baseline treatment.

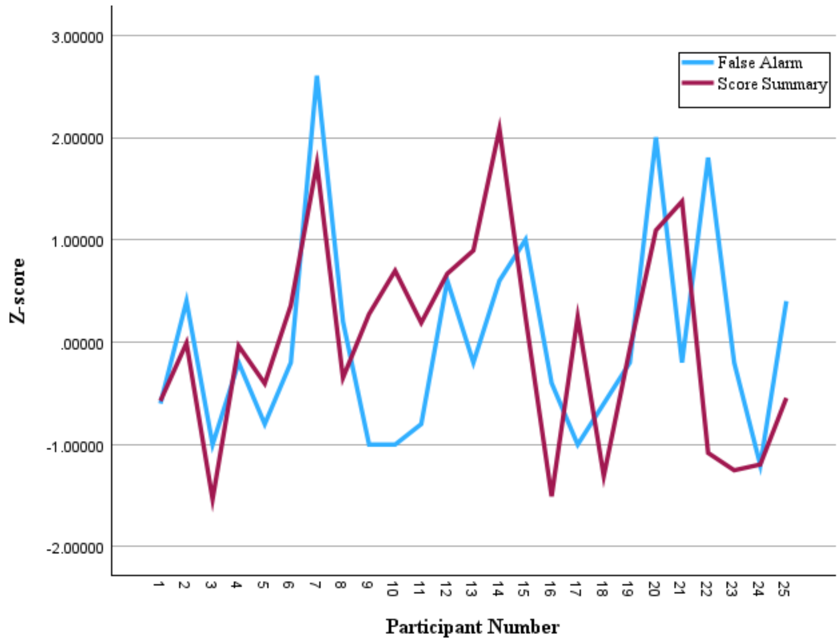
Of the 121 false alarm cases, the majority 80 events, 66%, of the total, involved a misidentification of a roll criteria exceedance. This was not unexpected, as the majority of the unstable and near-unstable events were allocated to the roll axis, a purposeful profile design intended to reflect moderate to high turbulence levels. The rolling motion of an airplane is very evident to the pilot when the amplitude and rate are elevated; the intensity of the experience primes the participant to respond and focuses attention. This was more often the case in the complex and moderate vignettes. More insidious are those events where the amplitude and rate are rather benign. Prolonged application of roll can generate a lateral path deviation. Due to the higher amplitudes and rates, participants were primed to expect a possible roll angle and lateral path exceedances with motion initiation.

False alarms for vertical speed were the next most frequent, with 15 events occurring. Equally evident to the pilot would be pitching actions, with the same effect of amplitude and rate on perception. The glideslope path, vertical speed, and airspeed are cross-coupled factors during an approach. For example, an aggressive pitch of the airplane downward not only drives the airplane to be low on the path but also elevates the rate of descent and increases airspeed, barring any thrust adjustment. Less prevalent were the remaining pitch-related stable approach criteria. Excessive vertical speed was followed by glideslope (13 events), airspeed high or low (11 events), and finally lateral path (2 events). These events tend to be subtler. For example, with the exception of a windshear encounter or uncommanded loss of thrust, airspeed tends to slowly change when not driven by a pilot-induced change of state, as might occur following a level off without commensurate thrust adjustment or other significant pitch change.

Examination of the data raised the possibility of correlation between achieving higher RA scores for unstable events and the number of false alarms. Did enhanced vigilance drive a tendency to respond at lower thresholds even as higher success was recognized? Data for RA score summations and false alarm case summations across all treatments were studentized and a line plot of the z-scores assessed for trends, as shown in Figure 16. The plot indicates false alarms tended to follow SA score totals, in terms of both magnitude and direction. This suggests the higher the participant performance on the measure of total RA scores, the greater the chance the participant would exhibit a tendency to invoke false alarms. This seems plausible, as a pilot exhibiting an elevated state of alarm and therefore primed to respond would tend to trigger on events proximate as well as beyond the criteria. The pilot may even call out an event when the conditions are far from present. Indeed, if the participant voiced an unstable condition whenever near the boundaries, their performance would be high, but would also be less discerning than an individual whose performance was lower.

**Figure 16**

*Z-score Comparison for False Alarm Events*




Normality of the false alarm and RA scores studentized data was assessed visually using Q-Q plots and analytically using the Shapiro-Wilk test. The plots indicated the observed values nicely tracked the expected normal values for studentized RA scores, with a skewness of .227 and a kurtosis of -.526; false alarms, on the contrary, presented a tail end sag at both ends of the Q-Q plot with a skewness of 1.169 and kurtosis of .933. The Shapiro-Wilk test results indicated that while the z-score data for SA scores suggested normality of the data, $W(25) = .967$, $p = .570$, studentized false alarms did not, $W(25) = .884$, $p = .008$. Regardless, a bivariate Pearson correlation test was conducted to assess the relationship between the two studentized data sets. The correlation established a moderate, positive relationship of statistical significance, $r(25) = .408$, $p = .043$. Completing a square root transformation with studentized values for both RA scores and false alarms slightly improved normality of the false alarm z-score data and impacted the Pearson test results, $r(25) = .394$, $p = .051$, but improved the model validity in that it just crossed the threshold value for statistical significance.

***Participant Experience Level***

Participants were conveniently split in terms of flight time experience. Those possessing greater than 1,500 hrs total flight time counted 13 against the remaining 12 possessing less than 1,500 hrs total flight time. The sample was grouped according to the experience metric and analyzed. Crosscheck skills were assumed to be correlated to experience level. Under this assumption, the application of boundaries was expected to be favored by pilots possessing well-established and entrenched instrument crosscheck skills, whereas the use of the alert was thought to favor those pilots less skilled. The alert was thought to fortify an otherwise developing, uncertain, or stagnant instrument scan discipline. An independent-samples, one-tail *t*-test was accomplished to compare the means for treatment scores for each group. The results did not reflect the assumptions.

Across the treatments, on average, the low-experience group outperformed the high-experience group. For Treatment 1, that under which most pilots currently operate, those of lower experience performed measurably better ($M = 23.42$, $SD = 8.415$) than those identifying a higher experience level ($M = 17.46$, $SD = 11.185$). For Treatment 2, the same outcome between the groups occurred with $M = 26.75$ and $SD = 11.655$ for the lower experienced group versus $M = 19.69$ and $SD = 9.499$ for the higher experience group. This trend followed for Treatment 3, $M = 28.00$ and $SD = 10.041$ for the experienced versus $M = 20.23$ and $SD = 7.316$ for those less experienced. Finally, the combination Treatment 4 favored the less experienced (*M* = 27.92, *SD* = 9.346) over the more experienced (*M* = 22.23, *SD* = 8.497) participant.

Means testing statistical significance occurred at a means difference of 7.769 at the 95% CI [.378, 15.161], $t(23) = 2.174$, $p = .020$ for Treatment 3. Tests for all other treatments were insignificant, although values for Treatment 1 [$t(23) = 1.494$, $p = .074$], Treatment 2 [$t(23) = 1.666$, $p = .055$)], and Treatment 4 [$t(23) = 1.594$, $p = .062$] were only slightly outside the acceptable standard. As such, they would be considered to have meaning toward interpretation and acceptance of the results.

Two items are worthy of note. First, a review of the means for each group indicated there was no difference in terms of the order of scores for the treatments, either in comparison with the overall test results or between the groups. But it also highlights that experience level didn't impact the type of enhancement applied to the PFD. This outcome was surprising. The other note for consideration dealt with currency. While the study required currency within the past five years for all participants, some participants had more recently actively flown in an operational context or had been engaged in training. All the low-experience pilots were currently involved in flight training, either as a recipient of it or a provider of it. All the high-experience pilots were engaged in either line flight operations with lower approach event to flight time ratios or flight simulation training where the instructor was seated in an observer position. The lower frequency of hands-on, repetitive instrument-based flight activity may have outweighed experience and led to better outcomes for the low-experience group.

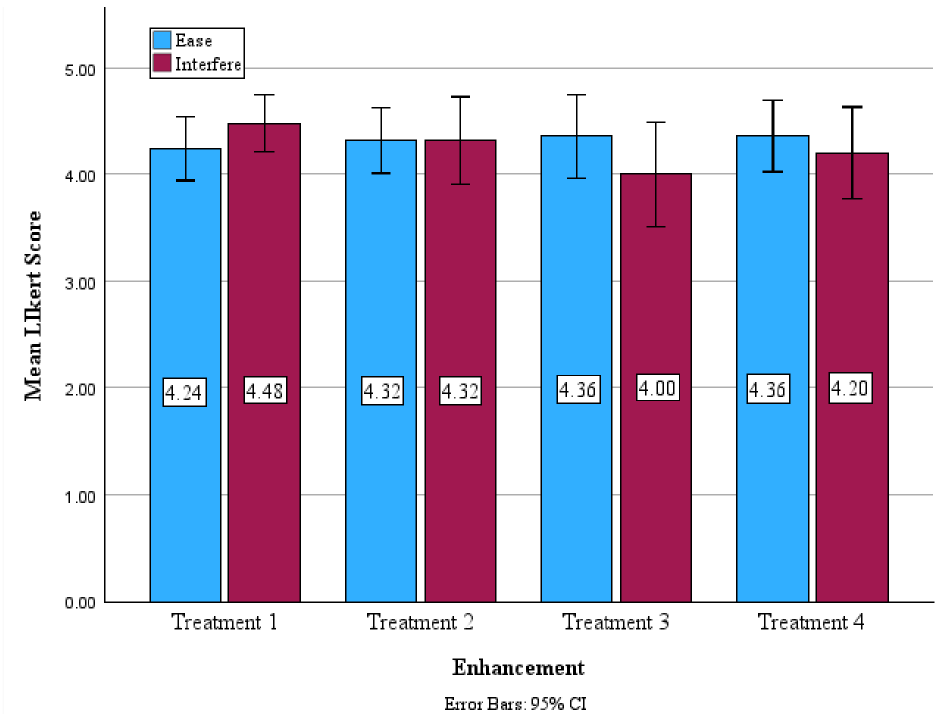
***Post-trial Questionnaire Results***

Participants provided valuable direct objective evaluation data in terms of the effect the baseline and enhanced displays had on SA. A post-trial questionnaire allowed the participant to evaluate two aspects: impact of the treatments on their SA and usability in flight operations and flight training. For the SA-focused portion, each question posed was framed in terms of the treatment presentations. Participants were asked whether the display with its enhancement (as applied) was easy to read, did not interfere with the monitoring task, improved perception and understanding of the current airplane stable approach state, and enhanced prediction of the future airplane stable approach state. The desire was to (a) see whether the participant felt differently about each of the treatments and (b) whether the internal perception was congruent with the indirect RA score data.

Figure 17 shows participant responses when scoring the degree to which the display was easy to read and did not interfere with the completion of the instrument scan monitoring task. Boxed data indicate the mean value and error bars to a 95% CI are provided. The *SD* values for ease were .723, .748, .952, and .810 for Treatments 1 through 4. The *SD* values for interference were .653, .988, 1.190, and 1.040. As can be seen, participant scores for both ease of use and non-interference with the instrument scan were high for all treatments. The ease of use score was consistent across all treatments, which was unsurprising given the design of each enhancement. Of note is the recognition that while all enhancements indicated overall value in terms of SA score outcomes from the think aloud protocols, the participant direct measure for interference suggested Treatment 1 was least impactful while Treatment 3 was of greatest impact. This was counterintuitive in that the presentation of continuous boundaries below 1,000 ft would've been thought to consume display space and perhaps clutter the baseline presentation and thus score lower than the less obtrusive Treatment 1 or Treatment 3.

**Figure 17**

*Likert Score Means for Ease and Interference*




*Note.* Values are based on a 5-point scale ranging from Strongly Disagree (1) to Strongly Agree (5).

These results may be due to a level of comfort with the current PFD layout. Current displays used for Treatment 1 provide a wealth of performance and navigation data, and the presentation of more data may be initially viewed internally as clutter, even as it brings benefit to the RA score. The incongruence of Treatment 2 and 4 scoring better than Treatment 3 in terms of interference may have less to do with the amount of display space employed than the manner in which it was presented. As a message that displayed only when an unstable condition was present, it may have been of such size and placement that attention capture was considered detrimental, even as RA scores indicated it improved SA. This possible explanation, however, did not hold when examined against Treatment 4 where even greater data is presented. Determination of the root cause of this apparent dichotomy would be addressed in further research, especially given the relatively broader CI shown for Treatment 3 over Treatments 2 and 4. Overall, however, the data indicated a sense of ease in use and a general lack of interference resulting from the increase in PFD information.

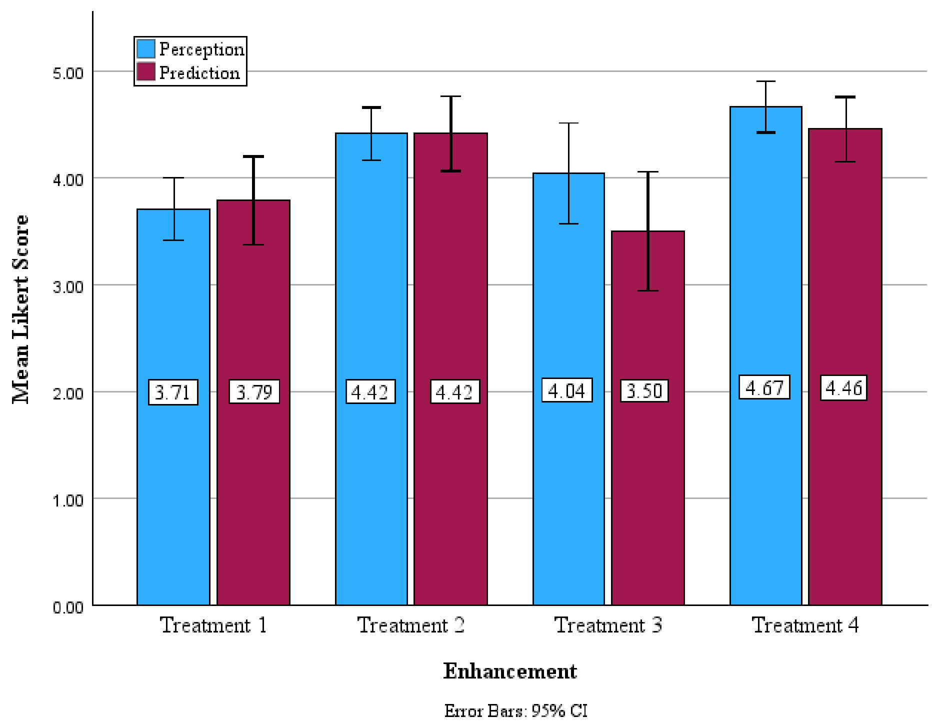
Figure 18 shows participant responses when scoring the degree to which the display enhanced perception and understanding of the current airplane stable approach state and enhanced prediction of the future airplane stable approach state. Boxed data indicate the mean value, and error bars to a 95% CI are provided. The *SD* values for ease were .690, .584, 1.122, and .565 for Treatments 1 through 4. The *SD* values for interference were .977, .830, 1.319, and .721. As can be seen, measurable differences in direct objective evaluation of the display enhancements were present. Participants found their perception was improved by all the enhanced treatments over Treatment 1. The data showed the presence of boundaries to be a common factor in driving higher Likert scores, with Treatment 4 recognized as the most beneficial, followed by Treatment 2. This result was surprising, as the indirect measures for RA score means showed Treatment 3 supported participant SA better than Treatment 2. The placement of the Likert values for Treatment 4, however, did correlate to the results observed for RA score means.

Prediction value Likert scores mirrored those of perception, with little difference between the two in terms of Treatment 1 and Treatment 2 on the basis of a comparison of perception to prediction value. It was in the use of Treatments 3 and 4 where a perception and prediction Likert scoring deviation was present. From Figure 18, it is apparent that

Treatment 4 was felt to be of greatest benefit, where the *M* values for both perception and prediction exceeded those of all other treatments. Still, the presence of the alert in Treatments 3 and 4 appeared to be of less benefit in prediction than in perception. Across the treatments, the PFD enhancement was generally viewed to benefit *perception* of an unstable approach condition than *prediction* of the future state. The exception to that was Treatment 1, where the PFD was viewed to slightly benefit prediction. Given the CI presented, however, it would be difficult to derive any definitive position as to the significance of the mean difference.

**Figure 18**

*Likert Score Means for Perception and Prediction*




*Note.* Values are based on a 5-point scale ranging from Strongly Disagree (1) to Strongly Agree (5).

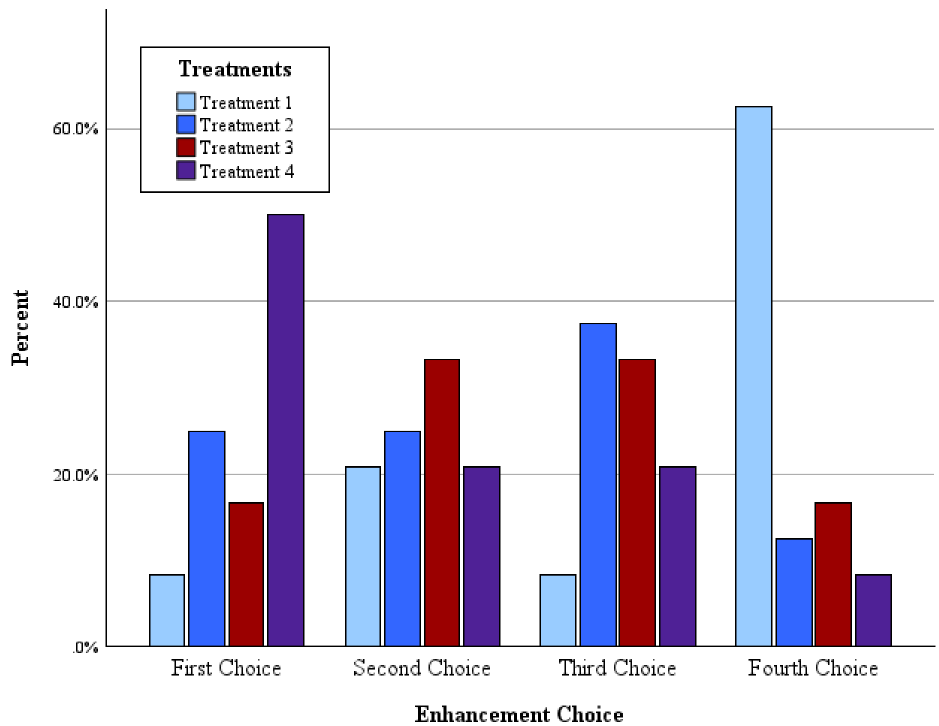
One final topic from the post-trial usability survey was the order in which the participant would place each treatment in terms of enhancement of pilot unstable approach situation awareness. As can be seen in Figure 19, the ranked order of treatment choices reflected the data provided in Figure 18. Figure 19 provides greater granularity as to the relative participant preferences. The combination Treatment 4 was the primary choice for enhancing SA, as was Treatment 1 dominant as the last choice. Surprisingly, a small percentage of participants felt the baseline Treatment 1 was better than any of the enhancements in improving their unstable approach SA. As the two participants scoring this treatment were of higher age and experience, 55 and 68 years with 10,000 and 3,500 hrs respectively, their possession of well-entrenched instrument scan patterns and practices may have favored the efficiency and accuracy of their time-proven methods and the innate desire to retain them.

It is in the relative placement of the other treatments that piques interest. Treatment 2 placed second in the first choice ranking for a number of individuals, mirroring the Likert data collected on perception, cognition, and prediction, once again implying that the boundaries, alone or in combination, are a strong draw for the participants. As a second choice, nearly all treatments held equal value – with the exception of Treatment 3 – indicating a near indifference to which treatment is better. This includes Treatment 1, which served as the control. That Treatment 3 bettered Treatment 2 as second choice reflects that while the use of boundaries in some fashion affected the primary choice, it became less relevant as a second choice.

In the third choice, Treatments 2 and 3 held greater interest, while Treatment 1 was the choice participants expressed as least likely to be beneficial. These middle choices should not carry much weight, as it is often easy for an individual to identify which product they have the greatest affinity and distaste, but individuals aren't often asked to rank those choices falling between. As such, this data supports the notion that participants were very adamant as to which treatment was most and least beneficial, but were less consistent in their middle choices.

**Figure 19**

*Treatment Order for Enhanced Situation Awareness*




### *Participant Commentary*

During and following data collection, participants provided insightful comments regarding unstable approaches. Some of the more interesting included the following topics. These comments are presented in that they would prove useful in future research and potential industry practices for unstable approaches.

**Stability Definition.** One participant discussed concern over a perceived "single-event" definition used in industry to indicate the presence of an unstable approach.

Emplacing an arbitrary value and operating to it, without needed context, was not useful to the participant and was believed it may generate unnecessary actions. When a pilot is flying an approach and the environmental conditions generate, for example, strong turbulence, one or more of the unstable criteria may be exceeded. Those exceedance events alone do not denote the overall stability of the approach. Rather, it is the overall airplane flight path that does. The corollary is the instance in which none of the criteria are exceeded, but the rate of change in the various performance and flight path parameters are exceedingly gross and indicate a deficit in the pilot's ability to manage to the condition. Such a case would be far more unstable, even if on the desired path and performance condition on average, than one in which minor transgressions occur. So, it is possible to have overall stability within a condition where minor unstable events are occurring.

The participant felt the instance of being relatively stable within an unstable state should not be cause for an immediate go-around. Such a case may appear to the pilot to be correctable within the bounds of safety but would indicate in recorded flight data the need to complete the go-around procedure. The essence of this position follows the FAA guidance on momentary deviations during testing and checking; if a momentary deviation occurs and the pilot executes corrective action to return within the tolerances, the task being completed is not necessarily failed on that basis alone. To the contrary, the correction is viewed as proper course of action and an indication of proper SA levels. Perhaps such a stance would benefit the unstable approach definition and on-going pilot compliance challenges.

**Instrument Scan Impacts.** A number of the participants, specifically those engaged in the ab initio training environment, felt the introduction of boundaries and alerts would help inculcate the unstable approach concept early on in pilot flying careers. However, more experienced participants voiced concern that the use of the alert would serve as a crutch and facilitate an undesired loss of proficiency in pilot instrument scan patterns. Such alerting is currently used in EICAS displays and benefits the pilot in avoiding the fatigue often encountered in hours of monitoring parameters of an otherwise stable system. While the systems monitored by EICAS tend to operate within a quiescent state, the majority of a typical flight profile, approach performance, and path measures often are not.

Given that missed events were observed in large number, it is possible to envision how pilots might relax their instrument scan pattern over time, knowing that an alert would be presented when the criteria for an unstable approach were exceeded. Erosion of core pilot skills over time is of great concern in academia and industry, especially when automation aides such as the autopilot or autothrottle systems failures occur. If accepted, difficult decisions on training regimens must be made to ensure an underlying ability to recognize and respond to the criteria exceedance notification and to manage the airplane correctly when such systems are not available.

**Presentation Philosophy.** One participant, in particular, equated the unstable approach issue to that of playing a video game. While the boundaries enhanced SA, the alert was more salient to that participant. The challenge is how to best convey the current condition. The participant felt momentary embedding of the specific criterion being met such as LEFT BANK, VVI, and AIRSPEED HI in lieu of the simple UNSTABLE alert used in Treatments 3 and 4 would allow SA to evolve beyond the psychomotor reactions to construct a narrative assessment of the specific unstable condition. In the video game context, the player knows how they failed the video game and are not left to deduce the root cause themselves. Perhaps the solution is not in the visual spectrum, but rather in some other attention channel such as audio feedback or the use of haptic means such as seat vibration when approaching an unstable approach parameter.

Another participant remarked on the overall philosophy as to when to display the information. Specifically, the participant wondered whether there was benefit in allowing the PM of a crewed airplane to vocalize the unstable approach condition prior to system notification. In such crewed flight deck environments, the PF and PM roles are well-defined and based upon an interactive communication model. Providing an immediate indication may not allow for each of the roles to be exercised. The participant noted that vocalized information sharing is critical to crew awareness. An airplane state presenting an unstable approach condition doesn't necessarily mean a bad outcome in terms of the landing; however, it has introduced uncertainty for the flight crew as to whether it is possible to remain on the runway if the landing is completed. Second, the magnitude of the system parameters set before indicating an unstable alert may have to be undesirably large to allow the PM and PF to react within a defined human information processing model. Finally, to be directive as opposed to suggestive, the color would need to indicate an urgency, perhaps by using a red or cyan display in the appropriate location.

Across the industry, there exists a common practice to voice when airspeed or other performance parameters fall outside a given tolerance. Application to the unstable criteria would be but an extension of this practice. However, waiting to present a defined condition to allow PF and PM crew interaction is misaligned with other types of PFD caution notifications, such as the presence of a windshear, ground proximity, or an opposing traffic flight path conflict. As such, an unstable approach condition may rise to the level of these types of flight risks and present an immediate cue. However, participants were clear in the belief that whatever presentation method was used, even if it improved pilot SA of an unstable approach, there should be “positive discretion” allowed to the pilot on the decision to continue or execute the go-around procedure.

**Instrument Scan Patterning.** All too often, the assumption is made that pilots possess an accepted universal method for executing their instrument scan. That assumption may be inaccurate as indicated by some of the participant commentary. One participant remarked having initially used a typical egocentric FD-focused scan pattern, shifting the area of gaze from the FD center point to the various performance scales in rapid sequence. However, during some vignettes, a more exocentric approach was used, taking a displaced perspective of the PFD and attempting to gather data using the peripheral cues while focusing the attention to the center of the display. In essence, the saccade gave way to an attempted attention dwell on the broader presentation. The participant believed no difference in unstable approach detection was perceived when attempting this method. But it does highlight that pilots may use different methods for scanning their instruments, depending on the conditions present. In some cases, the pilot may be able to step back from the situation and apply a wider field of regard, seeking meaningful performance and path trends worthy of attention, rather than following a well-established pattern. At other times, the pilot may have to focus every available resource on the FD guidance, foregoing all other information presented and missing crucial information on an undesirable state, such as approach instability.

These comments added color to the research. Participants brought their own perspectives and experiences to the study, providing a glimpse into the mindset a small sample of pilots may hold. While investigation of each comment was unwarranted within in this study, the comments are worthy of capture for further investigation.

## Conclusions

The study results supported a number of theoretical and practical contributions. The theoretical contributions focused on the use of the secondary task to emulate flying the airplane and the application and placement of the secondary task cues. The practical contribution reflected the benefit of the treatments used to enhance the PFD in terms of unstable approach condition SA.

### *Theoretical Contributions*

There are three theoretical contributions provided by this research. First, the use of a secondary task as a means to measure SA or degree of task saturation is pervasive in the literature. Less pervasive – and in fact not observed within the constraints of the literature review – is the inversion of task roles. In this application, the secondary task was used to *create* a demand for resource allocation at a level commensurate with actively managing an airplane final approach flight path. Each participant was monitored for secondary task cue response, and the degree of compliance was extremely high. While some at times had issue with depressing the correct switch, occasionally cross-activating when a cue appeared, nearly complete response compliance occurred. Participant feedback indicated the cues placed a moderate to high mental demand on them. As for the study design, few participants voiced a belief that the cue response accuracy was not being captured. To the contrary, the pre-trial briefing specifically ordered the tasks to be undertaken by the participant such that the cue response was the *first* item addressed. That placement pre-loaded the participant for compliance.

The second contribution is the placement and timing of the secondary cues. Special care was taken to ensure they fell within the pilot field of regard and did so without impacting the ability to complete the instrument scan. Given the participant cue response compliance and the lack of comments concerning loss of display data, it would appear the placement was effective. As to the timing of each of the cues, participants indicated they were presented at an appropriate frequency and allocation for the inceptor they were intended to replicate.

The final contribution touches on potential benefit in addressing undesirable authority gradients in the flight deck. Considerable study has been undertaken as to the presence of authority gradient relationships in a number of occupational fields to include transportation, engineering, and medicine. The existence of an authority gradient between two individuals seeking to achieve a common task objective was first recognized in the aviation field (Alkov et al., 1992; Allen, 2021; Edwards, 1975). Such gradients are not exclusively a high power-distance structure and can manifest in both strong and weak authority. Differences in defined organizational rank and established authority contribute to this gradient. Equally, in the aviation field the level of flight experience and perceived aviation expertise held by each individual can possibly have a detrimental effect on team communication and operational success (Cosby & Croskerry, 2004).

In a high authority relationship, pilots of inferior organizational placement or skill and knowledge level are, or feel they are, unable to vocalize concerns as to conditions affecting the continued safe operation of the airplane. The impact of this reticence to openly comment or question the actions of the PF can be significant, and recent airplane accidents have listed a trans-cockpit adverse gradient as a causal factor (AIBD, 2019; AAIIC, 2018). While efforts to improve open communication using Crew Resource Management and other training and operations tools have been undertaken, issues persist. Chow et al. (2014) analyzed cockpit automation and culture issues highlighted by the Asiana Airlines flight 214 accident, noting high power-distance societal culture can pressure flight crew into silence. Even in the unstable approach scenario encountered in this accident, flight crew perceiving the airspeed erosion remained silent. And while the passengers of Asiana 214 were largely spared, needless hull loss and subsequent tangential loss of life occurred. The on-going challenge is this: How to turn around highly-steeped cultural practices that present as deference to seniority and rank, even when faced with a potentially life-threatening situation?

The use of messaging to indicate a potentially hazardous condition is one solution. Current airplane designs make use of messaging, either through prominent alerts on the PFD or EICAS messaging, and flight crews from all cultures have demonstrated a tendency to enforce strict compliance to these indications. These messages remove the cultural aspect from the safety equation by providing either a boundary to be avoided or a communication to be recognized and addressed. No longer is communication of a key performance or flight path parameter impacted by the authority gradient, but rather compliance is suggested by an agnostic parcel of information. Communication of this information by the PM is viewed as a duty to be performed in a timely and accurate fashion, much as completion of a checklist, and would likely prove an impetus to speak up. The use of an unstable approach alert or depiction of criteria boundaries would further progress toward achieving a balanced and effective use of authority and teamwork in the flight deck.

***Practical Contributions***

Industry groups continue to tackle the issue of unstable approaches, and a number of companies have employed tools to aid the pilot. Occasionally, opinions are taken as fact, and viable solutions to real-world problems are lost or dismissed. This study demonstrated that enhancements meant to improve pilot SA of an unstable approach condition are beneficial, if properly implemented. Not only did the data reflect that, but the study participants so stated in their individual comments and post-trial usability survey data.

A number of open questions were answered during the analysis of the data. Assumptions as to the ability of pilots to consistently perceive an unstable condition were found to be less than accurate. Study participants, while having achieved the defined skills to obtain an instrument rating, exhibited diverse ability to perceive an exceedance of clearly defined boundaries. Individual performance on complex, moderate, and simple unstable approach vignettes indicated but a small number were able to identify nearly every event presented. In fact, the participant sample collectively missed the vast number of events. They did so, even when provided clear indications in a manner commonly used in production flight instrument displays.

An additional point of interest was the degree to which participant experience influenced the ability to perceive and understand an unstable event. It had been postulated that pilots possessing high levels of experience would find the tasks presented them to be challenging, but much less so than pilots of lower experience. That was not the case, and the opposite appeared to be true. The discussion highlighted that low flight time participants were more actively engaged in repetitive, high frequency approach event activity, whereas those of higher experience were more passively involved, either due to long flight durations at altitude followed by a single instrument approach or through passive observation while instructing.

Concerns as to the presentation of unstable approach information introducing undesired clutter to an already information-rich PFD appear to be ill-founded based on participant response to questions concerning ease with which the display could be read and the potential minimized for interference with the monitoring task. Participants indicated high favorability, with Likert values meeting or exceeding the 4.0 value out of a maximum 5.0.

Flight instrument display designers and manufacturers may seek to use this data to support the broad question as to whether to pursue an unstable approach notification feature in their products, and the narrower question as to whether boundaries, alerts, a combination of both, or perhaps some other concept would be most beneficial in which to commit corporate resources.

**Limitations of the Findings**

Experiments have limitations, and those limitations can affect the relevance of the study to the operational environment. Experiments, by nature, are typified by application of a controlled environment supportive of causality determinations. Protections put in place to avoid confounding variables and bias are never perfect. This study encountered a number of limitations and delimitations, as previously identified in the Introduction chapter. Additional limitations were observed during the study and are addressed.

Every effort was made to ensure each participant was uniformly briefed, trained in the display presentations to be encountered, and provided an orientation vignette to experience the data collection tool. For those needing additional orientation, a practice vignette was provided. Despite these efforts, nearly all participants experienced some degree of adaptation as the study progressed. Anecdotal observation indicated it could take a number of vignettes before the participant settled into the routine. The use of the Latin Squares method disseminated the impact of the learning curve over diverse vignettes, reducing the overall study impact. Still, the study would have benefitted from additional practice sessions before data collection. This benefit would have come at a cost though, as total participation time was approximately 70 min, and fatigue would have become a greater concern.

The study presented vignettes created to highlight specific unstable approach criteria exceedance events. Seven vignettes were created to capture the diversity of criteria desired and allow for a near-even distribution of event types. During the transition to a within-subject, repeated-measures design, the scoreable vignettes were culled to reduce the total number from 24 to 12. In doing so, an imbalance occurred such that roll angle exceedances outweighed other criteria. The missed events data obtained may have been disproportionately affected by this imbalance, as a number of the missed events were preceded by events approximate to, but not in excess of, the roll angle criteria. The study would benefit from a more balanced presentation.

**Recommendations**

This study lays the groundwork for further investigation of pilot SA during an unstable approach. Historical and current unstable approach research focuses on identifying conditions common to sampled documented unstable approaches, the merits of the currently used unstable approach criteria, and potential predictive tools for alerting the crew of the airplane energy state. One key aspect is what, if any, impact the display of data pertaining to unstable conditions enhances pilot SA. Understanding how the three rather fundamental enhancements used in this study benefitted the pilot helps guide needed training and operational solutions to the on-going problem, one that has not yet been addressed adequately through extensive and continual awareness messaging.

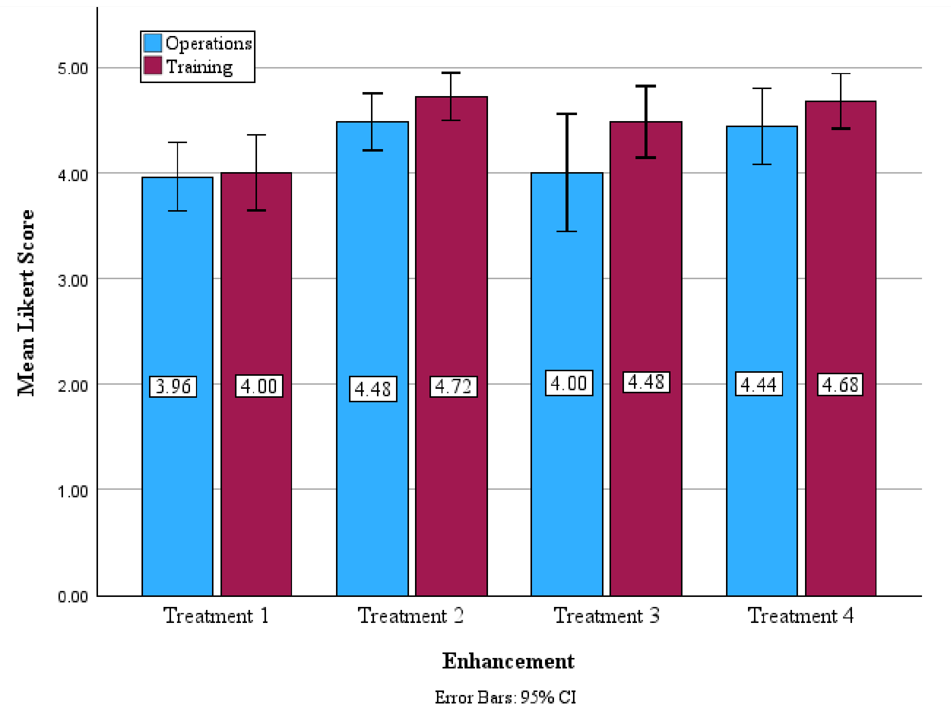
***Recommendations for Industry***

A number of considerations for industry were identified in the study. While the various approaches to solve the problem with continuing unstable approaches may prove fruitful, a number of near-future improvements may prevent near-term incidents and accidents.

Data collected using the post-trial survey provided insight as to the usefulness of the various display treatments in operations and training. Participants provided perspectives representing a broad range of flight experience, as indicated previously. Treatments were scored separately to allow differentiation between each of the enhancements presented to the participant. The results of the survey are shown in Figure 19. Boxed data indicate the mean value for each treatment, and error bars to a 95% CI are provided. Data are presented for each of the questions posed: (a) “The display presented would be beneficial in *flight* operations” and (b) “The display presented would be beneficial in *training* operations.”

**Figure 19**

*Likert Score Means for Operations and Training*




*Note.* Values are based on a 5-point scale ranging from Strongly Disagree (1) to Strongly Agree (5).

Participants were informed the treatments were not an end-state design but rather a notional representation of two basic display presentation concepts. Consequently, the responses do not represent a particular treatment configuration but rather the overall representation of the treatment. The data indicated participants desired to include some form of unstable approach enhancement to the current PFD format in both operations and training. The use in training appeared to outweigh that of use in operations. Participants were desirous of the boundaries, both alone and in combination, in operations but even more so in training. This occurred despite a handful of participant comments finding the continuous presentations of the boundaries to somewhat clutter the display.

The participant preference did not warrant complete disregard for the application of boundaries as a tool but rather to either display the boundaries and have the exceeded criteria highlight in red or present the alert at the time of the event and “pop up” the exceeded boundary. Both approaches were viewed to reduce clutter while providing the desired information to the pilot. The overall value of the display to industry was best summarized by one participant who noted, “The presentation of boundaries and the alert would be useful when tired; by the end of a long flight my crosscheck is slower and it would be very helpful in completing the flight safely.” This potential contribution to safety should not be overlooked.

Through survey of various industry publications, review of various airplane manufacturer procedures, regulatory guidance, and anecdotal commentary from study participants, there appears to be little established guidance on the vocalization of unstable approach conditions. There exists, however, guidance on other types of messaging that would benefit pilot SA, such as approaching key altitudes during final approach segment, the presence of an airplane upset condition, airplane system alerting, takeoff thrust setting, and critical speed values. The extension of an unstable callout would align with practices employed in other phases of flight.

Observation of participant think aloud protocols highlighted two facts. First, the participants initially demonstrated some difficulty in voicing the information sought. While the information was forthcoming, it was apparent some discomfort emanating from unfamiliarity with the activity was present. However, this discomfort was rapidly overcome, and the majority of participants adapted to the task within the first few vignettes. In fact, many of the participants eventually were able to consistently voice unstable conditions in a timely manner, and do so in a consistent, standardized format. Second, the manner by which the information was vocalized varied by individual. While participants were directed a suggested order – announce the state as unstable, indicate under which criteria it was unstable, and direct the action to be followed – they tended to adapt to a structure that best suited them. While some would offer “Unstable for airspeed, correcting and continuing the approach,” others might use “Airspeed unstable, correct and continue.” The structure of the information didn’t seem to impact the performance of the participant. In post-trial discussion, a number of participants noted the inclusion of such a callout would benefit the industry.

***Recommendations for Future Research***

There exists a delicate balance between ensuring the validity of the research results and the ability to generalize those results to a broader population. The sample used for this research comprised a broad spectrum of pilot experience, as determined by total flight hours. Members of each of the two experience levels were subjected to vignettes applying the four separate treatment groups. While this supported generalizability of the results to both low- and high-time pilots, it did not allow for greater granularity in understanding causation among a number of considerations. In fact, it could be argued that extending participation in this study to pilots holding different pilot certifications, total flight times, recency of experience, and exposures to stable approach criteria may, in fact, confound the results. This follows from the numerous identified and unidentified variables that cannot be controlled with such a broad spectrum of participant knowledge, skills, and attributes. Study results may vary with a more homogeneous sample and should be investigated. This leads to a general recommendation to focus this research to groups of greater demographic homogeneity.

While this study focused on four specific airplane performance values as indicators of an unstable approach – airspeed, vertical speed, roll angle, and path – current academic and industry thought includes a number of additional parameters that were either assumed to be validated or were not a consideration. The presence of these additional parameters may increase the complexity of the subject. Common themes in runway excursion events include the presence of significant weather factors such as gusty wind conditions and onerous ceilings and visibilities, non-normal and emergency conditions, and elevated stress placed upon the flight crew. Such an expansion in the study would require the use of more sophisticated simulation measures to create the desired environment. This study was also limited in its ability to present a realistic flight environment to the participant pilot. Perception of unstable conditions can be foreshadowed by readily-obtained inflight information and subtle proprioceptive, kinesthetic, and equilibrioceptic cues during the approach that could not be replicated using a tabletop simulation.

A final recommendation would be to investigate for variation in the efficacy and acceptance of varying methods of presenting unstable approach cueing. The presentation used for this research was intentionally simplistic due to limited financial and time resources. Additional time and funding would support novel presentations that may be of greater benefit in advising pilots as to when an unstable approach condition is present.

## Summary

This study gathered data to determine how notional flight instrument display enhancements might improve pilot SA as to the presence of an unstable approach condition. The data showed that pilots experienced statistically significant RA means differences and thus higher levels of SA compared to the baseline treatment when boundaries and alerts were applied to the PFD.

The penultimate objective of this research was to provide empirical data to the academic and manufacturing community as to potential unstable approach SA benefits associated with incorporating bounds and/or alerts into existing flight instrument displays. Serendipitous findings helped better characterize the results. First, identification of unstable events may be of greater overall benefit in assessing pilot SA as to their presence. Second, flight experience did not have a significant effect on the RA scores, suggesting the benefit of enhancements occurred across the entire demographic. Third, boundaries appear to be the most significant factor in both RA scores and the objective evaluation of SA improvements. Finally, all participants in the study scored the use of enhanced unstable approach displays very high for both operations and training, seeing the benefit of providing the pilot some indication that the quality of airplane performance and path guidance are reduced.

While the data showed affirmation of the research question in two of the six treatment comparisons, specifically against the baseline PFD configuration, further research is necessary to determine the impact at greater granularity for a number of factors not addressed in this study. Follow-on research should examine unstable approach

SA in simulations capable of higher complexity and focus on homogeneous samples to allow greater certainty of the results.

**References**


Accident Investigation Board Denmark [AIBD] (2019). *Report on accident to D-CAWM (Cessna 560 XLS+) at Aarhus (EKAH) on 5-8-2019* (Report No. 2019-327). https://skybrary.aero/sites/default/files/bookshelf/6058.pdf

Adamski, A. J., & Doyle, T. J. (2005). *Introduction to aviation regulatory process* (5th ed.). Hayden-McNeil.

Adobe Premier Pro® (Version 23.0). [Computer software]. Adobe Systems.

Aeronautical Accidents and Incidents Investigation Commision [AAIIC] (2018). *Aircraft investigation report* (Report No. ZKM KHAIA 001/R). https://skybrary.aero/sites/default/files/bookshelf/32814.pdf

Airbus. (2016). *A380 aircraft characteristics airport and maintenance planning.* https://www.airbus.com/aircraft/support-services/airport-operations-and-technical-data/aircraft-characteristics.html

Alkov, R. A., Borowsky, M. S., Williamson, D. W., & Yacavone, D. W. (1992). The effect of trans-cockpit authority gradient on Navy/Marine helicopter mishaps. *Aviation, Space, and Environmental Medicine, 63*(8), 659–661.

Allen, Robert D. (2021) *Pilot perception of cockpit organizational framework's impact on flight safety and subordinate pilot behavior*. (Doctoral dissertation, Embry-Riddle Aeronautical University). https://commons.erau.edu/edt/705

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*, 63–71. doi:10.1016/j.jarmac.2014.04.010

Anderson, V. L., & McLean, R. A. (2018). *Design of experiments: A realistic approach* (1st ed.). Routledge.

Aukstakalnis, S. (2017). *Practical augmented reality: A guide to the technologies, applications, and human factors for AR and VR.* Pearson.

Baker, E. R. (2017). *Comparative analysis of conventional electronic and OZ concept displays for aircraft energy management* (Doctoral dissertation, Embry-Riddle Aeronautical University). http://voyager.db.erau.edu:7008/vwebv/holdingsInfo?searchId=1&recCount=50&recPointer=488&bibId=474832

Banks, V. A., Stanton, N. A., & Harvey, C. (2014). What the drivers do and do not tell you: Using verbal protocol analysis to investigate driver behaviour in emergency situations. *Ergonomics*, *57*(3), 332–342. https://doi.org/10.1080/00140139.2014.884245

Bardou, N. & Owens, D. (2014). *Hard landing, a case study for crews and maintenance personnel* [Airbus Safety First #17]. Airbus.

Boeing Commercial Airplanes [Boeing]. (2012). Reducing runway landing overruns. *AERO*, *12*(3), 15-19. Retrieved from http://www.boeing.com/commercial/aeromagazine/articles/2012_q3/pdfs/AERO_2012q3.pdf.

Boeing Commercial Airplanes [Boeing]. (2022). *Statistical summary of commercial jet airplane accidents, worldwide operations, 1959-2022.* https://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf

Brams, S., Ignace, T. C. H., Ziv, G., Dauwe, S., Evens, K., De Wolf, T., Levin, O., Wagemans, J., & Helsen, W. F. (2018). Does effective gaze behavior lead to enhanced performance in a complex error-detection cockpit task? *PLoS One, 13*(11) doi:http://dx.doi.org.ezproxy.libproxy.db.erau.edu/10.1371/journal.pone.0207439

Burian, J. M. (2011). 2011: The year in review. *Proceedings of the 64th Annual International Air Safety Seminar (IASS)*. Flight Safety Foundation.

Burian, B., Barshi, I., & Dismukes, K. (2005). *The challenge of aviation emergency and abnormal situations*. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20060023295.pdf

Campbell, A., Zaal, P., Shah, S., & Schroeder, J. A. (2019). Pilot evaluation of proposed go-around criteria for transport aircraft. *AIAA Aviation 2019 Forum.* doi:10.2514/6.2019-3610

Cherulnik, P. D. (2001). Specifying the research variables ii: artifacts and construct validity. *Methods for behavioral research: A systematic approach* (pp. 99-134). SAGE Publications. doi:10.4135/9781452233666.n4

Chow, S., Yortsos, S., & Meshkati, N. (2014). Asiana Airlines flight 214: Investigating cockpit automation and culture issues in aviation safety. *Aviation Psychology and Applied Human Factors*, *4*(2), 113–121. https://doi.org/10.1027/2192-0923/a000066

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159. https://web-b-ebscohost-com.ezproxy.libproxy.db.erau.edu/ehost/pdfviewer/pdfviewer?vid=1&sid=3fba5933-b334-4e93-86e4-d29404c7f7b1%40pdc-v-sessmgr05

Colbert, J. L. (2011). Statistical or non-statistical sampling: Which approach is best? *Journal of Applied Business Research*, *7*(2), 117-. https://doi.org/10.19030/jabr.v7i2.6252

Cooper, G. E., & Harper, R. P. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities*. (Report No. NASA TN-D-5153). https://ntrs.nasa.gov/search.jsp?R=19690013177

Cosby, K. S., & Croskerry, P. (2004). Profiles in patient safety: Authority gradients in medical error. *Academic Emergency Medicine*, *11*(12), 1341–1345. https://doi.org/10.1197/j.aem.2004.07.005

Cowan, N. (2005). *Working memory capacity*. Psychology Press.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* SAGE Publications.

Curtis, M. T., Jentsch, F., & Wise, J. A. (2010). Aviation displays. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (pp. 439-478). Academic Press.

Daidzic, N. E., & Shrestha, J. (2008). Airplane landing performance on contaminated runways in adverse conditions. *Journal of Aircraft, 45*(6), 2131-2144. 10.2514/1.38056

De Boer, R. J., Coumou, T., Hunink, A., & Van Bennekom, T. (2014). The automatic identification of unstable approaches from flight data. *6th International Conference on Research in Air Transportation*. https://www.semanticscholar.org/paper/The-automatic-identification-of-unstable-approaches-Boer-Coumou/bfe873bf03dd96c2afbf5e2ac298dd8a34ee1d3f

DeCelles, J. L. (1991). *The delayed GPWS syndrome.* Aviation Research and Education Foundation.

Dillard, M. B., Warm, J. S., Funke, G. J., Funke, M. E., Finomore, V. S., Matthews, G., Parasuraman, R. (2014). The sustained attention to response task (SART) does not promote mindlessness during vigilance performance. *Human Factors: The Journal of Human Factors and Ergonomics Society, 56*(8), 1364-1379. doi:10.1177/0018720814537521

Dismukes, R. K. (2010). Understanding and analyzing human error in real-world operations. In E. Salas & D. Maurino (Eds.), Human factors in aviation (pp. 335-374). Academic Press.

Dole, C. E., Lewis, J. E., Badick, J. R., & Johnson, B. A. (2017). *Flight theory and aerodynamics: A practical guide for operational safety* (3rd ed.). John Wiley & Sons.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives, 38*(1), 105-124. http://search.proquest.com.ezproxy.libproxy.db.erau.edu/docview/1020696142?accountid=27203

Durso, F. T., & Alexander, A. L. (2010). Managing workload, performance, and situation awareness in aviation systems. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (pp. 217-248). Academic Press.

Durso, F. T., Bleckley, M. K., & Dattel, A. R. (2006). Does situation awareness add to the validity of cognitive tests? *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48*(4), 721-733. doi:10.1518/001872006779166316

Ebbatson, M. (2009). *The loss of manual flying skills in pilots of highly automated airliners* (Doctoral dissertation, Cranfield University). https://dspace.lib.cranfield.ac.uk/handle/1826/3484

Edwards, E. (1975). Stress and the airline pilot. British Airline Pilots' *Association Technical Symposium: Aviation Medicine and the Airline Pilot*. British Airline Pilots' Association.

Endsley, M. R. (1990). A methodology for the objective measurement of pilot situation awareness. In *Situation Awareness in Aerospace Operations* (AGARD-CP-478, pp. 1-1-1-9). Advisory Group for Aerospace Research and Development.

Endsley, M. R. (1995a). A taxonomy of situation awareness errors. In R. Fuller, N. Johnston, & N. McDonald (Eds.), *Human factors in aviation operations* (pp. 287–292). Avebury Aviation.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 147-174). Lawrence Erlbaum.

Endsley, M. R. (2013). *Designing to support situation awareness: SA theory, SA demons, &SA errors* [PowerPoint presentation]. https://erau.instructure.com/courses/73903/pages/1-dot-4-guided-support?module_item_id=3918430

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making, 9*(1), 4-32. doi:10.1177/1555343415572631

Endsley, M. R., Bolstad, C. A., Jones, D. G., & Riley, J. M. (2003). Situation awareness oriented design: From user's cognitive requirements to creating effective supporting technologies. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 47*(3), 268-272. doi10.1177/154193120304700304

Endsley, M. R. & Jones, D. G. (2012). *Designing for situation awareness: An approach to user-centered design (2nd ed.).* CRC Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211-245.

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.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2014). G*Power (Version 3.1.9.2) [Computer software]. Heinrich-Heine-Universität Düsseldorf.

Federal Aviation Administration [FAA]. (2002). Criteria for Approval of Category I and Category II Weather Minima for Approach. Advisory Circular 120-29A. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-29A.pdf

Federal Aviation Administration [FAA]. (2014b). Aircraft Wake Turbulence. Advisory Circular 90-23G. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_90-23G.pdf

Federal Aviation Administration [FAA]. (2014a). Electronic Flight Displays. Advisory Circular 25-11B. Retrieved from https://www.faa.gov/documentLibrary/media/Advisory _Circular/AC_25-11B.pdf

Federal Aviation Administration [FAA]. (2014). Mitigating the Risks of a Runway Overrun upon Landing. Advisory Circular 91-79A. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_91-79A.pdf

Federal Aviation Administration [FAA]. (2016). Aircraft Certification Service – Mission, Vision, Organization Structure and Functions. Order 8100.5B. https://www.faa.gov/regulations_policies/orders_notices/index.cfm/go/document.information/documentID/1030266

Federal Aviation Administration [FAA]. (2017b). Aeronautical Information Manual. https://www.faa.gov/air_traffic/publications/atpubs/aim_html/index.html

Federal Aviation Administration [FAA]. (2017c). Air Traffic Control. Order JO 7110.65X. https://www.faa.gov/regulations_policies/orders_notices/index.cfm/go/document.current/documentNumber/7110.65

Federal Aviation Administration [FAA]. (2017d). FAA English Language Standard for an FAA Certificate Issued Under 14 CFR Part 61, 63, 65, and 107. Advisory Circular 60-28B. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_60-28B.pdf

Federal Aviation Administration [FAA]. (2017e). Responsibility and Authority of Pilot in Command, 14 C.F.R. §91.3. https://www.law.cornell.edu/cfr/text/14/91.3

Federal Aviation Administration [FAA]. (2020). General Definitions, 14 C.F.R. §1.1. https://www.ecfr.gov/cgi-bin/text-idx?rgn=div5&node=14:1.0.1.1.1

Federal Aviation Administration [FAA]. (2021a). General Definitions, 14 C.F.R. §61.129. https://www.ecfr.gov/cgi-bin/text-idx?pitd=20180727&node=se14.1.61_1129&rgn=div8

Federal Aviation Administration [FAA]. (2021b). U.S. Terminal Procedures Publication, Northwest (NW) Vol 1 of 1. https://aeronav.faa.gov/upload_313-d/terminal/2021-06-17/NW1.pdf

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

Flach, J. M. (1995). Situation awareness: Proceed with caution. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 149-157. doi:10.1518/001872095779049480

Flight Safety Foundation [FSF]. (2010). Killers in aviation: FSF Task Force presents facts about approach-and-landing and controlled flight into terrain accidents. *Flight Safety Digest*, *17*, 1-256.

Flight Safety Foundation [FSF]. (2017). *Go-around decision making and execution project.* Flight Safety Foundation.

Greer, A. A., Hedges, D.E., & Hunter, D. J. *Display modifications for enhanced flight crew stabilized approach situation awareness.* Unpublished manuscript, Department of Aviation Studies, Embry-Riddle Aeronautical University.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Harbour, S. D. (2015). *A neuroergonomic quasi-experiment: Predictors of situation awareness and display usability with USAF pilots while performing complex tasks* (Doctoral dissertation, Northcentral University). ProQuest Dissertations Publishing (Order No. 3685732).

Harrivel, A. R., Liles, C., Stephens, C. L., Ellis, K. K., Prinzel, L. J., & Pope, A. T. (2016). *Psychophysiological sensing and state classification for attention management in commercial aviation*. NASA/Langley Research Center. Retrieved from http://ezproxy.libproxy.db.erau.edu/login?url=https://search-proquest-com.ezproxy.libproxy.db.erau.edu/docview/2128126689?accountid=27203

Haslbeck, A., Gontar, P., & Shubert, E. (2012). The way pilots handle their control stick – Effects shown in a flight simulator study. *30th European Annual Conference on Human Decision-Making and Manual Control*

Hays, R. T. (1980). *Simulator fidelity: A concept paper* (Report No. ARI-TR-490). http://www.dtic.mil/docs/citations/ADA101475

Hess, R. A. (2003). Pilot control. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 265-309). CRC Press.

Ho, G., Hunter, D. J., Thirtyacre, D. L., Thomas, R. (2016). *14 CFR part 60 fight training simulation device synthesis: Pilot perception of system anomalies.* Unpublished manuscript. Department of Aviation, Embry-Riddle Aeronautical University.

Human, R., Henry, M., Jacobs, W. J., & Thomas, K. G. F. (2018). Elevated cortisol leaves working memory unaffected in both men and women. *Frontiers in Human Neuroscience.* doi:http://dx.doi.org.ezproxy.libproxy.db.erau.edu/10.3389/fnhum.2018.00299

*iFly Jets 737NG®* (Version 3.2.3.4) [Computer software]. Flight One Software.

International Air Transport Association [IATA]. (2016). *Unstable approaches: Risk mitigation policies, procedures and best practices* (2nd ed.). International Air Transport Association.

International Business Machines [IBM]. (2023). IBM SPSS Statistics for Windows. (Version 29.0.1.0) [Computer software]. IBM Corporation.

Janelle, C. M. (2002). Anxiety, arousal and visual attention: A mechanistic account of performance variability. *Journal of Sports Sciences*, 20, 237–251. doi:10.1080/026404102317284790

Jones, D. G., &. Endsley, M. R. (1996). Sources of situation awareness errors in aviation. *Aviation, Space and Environmental Medicine, 67*(6), 507-512.

Jones, D. G., & Endsley, M. R. (2000). Overcoming representational errors in complex environments. *Human Factors*, *42*(3), 367– 378.

Kennedy, Q., Taylor, J. L., Reade, G., & Yesavage, J. A. (2010). Age and expertise effects in aviation decision making and flight control in a flight simulator. *Aviation, space, and environmental medicine*, *81*(5), 489–497. doi:10.3357/asem.2684.2010

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Education and Psychology Measurement, 56*(5), 746-759.

Kumar, B., De Remer, D., & Marshall, D. M. (2005). *An illustrated dictionary of aviation*. McGraw-Hill.

Lau, N. & Boring, R. (2018). Situation awareness in sociotechnical systems. In H. M. Cuevas, J. Veláquez, & A. R. Dattel (Eds.), *Human factors in practice: Concepts and applications* (1st ed.) (pp. 55-70). CRC Press.

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). *Designing for people: An introduction to human factors engineering*. CreateSpace.

Loft, S., Bowden, V., Braithwaite, J., Morrell, D. B., Huf, S., & Durso, F. T. (2015). Situation awareness measures for simulated submarine track management. *Human Factors: The Journal of Human Factors and Ergonomics Society, 57*(2), 298-310. doi:10.1177/0018720814545515

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, *51*(3), 485–504. https://doi.org/10.1002/ejsp.2752

Martin, W. L., Murray, P. S., Bates, P. R., Lee, P. S. Y. (2015). Fear-potentiated startle: A review from an aviation perspective. *The International Journal of Aviation Psychology, 25*(2), 97-107. doi:10.1080/10508414.2015.1128293

Masson, D. (2023). *Balanced Latin square generator* (Initial Version) [Computer software]. https://damienmasson.com/tools/latin_square/

Mitchell, J., Vermeulen, L. P., & Naidoo, P. (2009). Flying glass: A qualitative analysis of pilot perceptions of automated flight-decks after 20 years. *International Journal of Applied Aviation Studies, 9*(1), 13–28.

Moriarty, D., & Jarvis, S. (2014). A systems perspective on the unstable approach in commercial aviation. *Reliability Engineering and System Safety, 131*, 197-202. doi:10.1016/j.ress.2014.06.019

Morris, C. H., & Leung, Y. K. (2006). Pilot mental workload: How well do pilots really perform? *Ergonomics*, *(49)15*, 1581-1596. doi:10.1080/00140130600857987

Mosier, K. L. (2010). The human in flight: From kinesthetic sese to cognitive sensibility. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (pp. 147-173). Academic Press.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International Journal of Aviation Psychology, 8*(1), 47-63. doi:10.1207/s15327108ijap0801_3

National Aeronautics and Space Administration [NASA]. (2020). Aviation safety reporting system database online [Database]. https://asrs.arc.nasa.gov/search/database.html

National Transportation Safety Board [NTSB]. (1973). *Aircraft accident report Eastern Air Lines, Inc. L-1011, N310EA, Miami, Florida, December 29, 1972* (Report No. NTSB-AART-73-14). https://www.ntsb.gov/investigations/AccidentReports/Pages/AAR7314.aspx

National Transportation Safety Board [NTSB]. (2021). *Aircraft investigation final report* (Report No. DCA19MA143). file:///C:/Users/n629597/AppData/Local/Temp/Report_DCA19MA143_99367_12_22_2023%206%2036%2005%20PM.pdf

National Transportation Safety Board [NTSB]. (2023). *Aircraft investigation final report* (Report No. ERA21LA208). file:///C:/Users/n629597/AppData/Local/Temp/Report_ERA21LA208_103036_12_22_2023%206%2019%2003%20PM.pdf

Nielson, J. (1994). Enhancing the explanatory power of usability heuristics. *Chi '94 Proceedings* (pp. 152-158). Association for Computing Machinery.

Noble, C. (2002). The relationship between fidelity and learning in aviation training and assessment. *Journal of Air Transportation, 7*(3), 33. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.472.8909

O'Hare, D. (2003). Aeronautical decision making: Metaphors, models, and methods. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 201-238). CRC Press.

Papua New Guinea Accident Investigation Commission [AIC]. (2018). *Final aircraft accident investigation report* (Report No. AIC 18-1004). https://www.aic.gov.pg/sites/default/files/2020-08/FINAL%20REPORT%2C%20issued%2015.07.2019.pdf

Pew, R. W. (2000). The state of situation awareness measurement: Heading toward the next century. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness: Analysis and measurement* (1st ed.). Lawrence Erlbaum Associates.

Pierce, R. S. (2012). The effect of SPAM administration during a dynamic simulation. *Human Factors*, *54*(5), 838–848. https://doi-org.ezproxy.libproxy.db.erau.edu/10.1177/0018720812439206

Prepar3D® (Version 4.0) [Computer software]. Lockheed Martin.

Pritchett, A. R., & Hansman, R. J. (1997). Pilot non-conformance to alerting system commands during closely spaced parallel approaches. *16$^{th}$ Digital Avionics Systems Conference.* doi:10.1109/DASC.1997.637290

Rao, A. H., & Puranik, T. G. (2018). Retrospective analysis of approach stability in general aviation operations. *2018 Aviation Technology, Integration, and Operations Conference.* doi:10.2514/6.2018-3049

Reason, J. (1990). *Human error*. Cambridge University Press.

Ritter, F. E., Kim, J. W., Morgan, J. H. & Carlson, R. A. (2013). Preparation for running experiments. In *Running behavioral studies with human participants: A practical guide* (pp. 19-56). SAGE Publications. doi:10.4135/9781452270067.n2

Ross, G. E. (2018). *Human factors contributing to unstabilized approaches and landings in commercial aviation incidents: An analysis of ASRS reports* (Master's thesis, Embry-Riddle Aeronautical University). https://commons.erau.edu/edt/380/

Roth, W., & Mavin, T. J. (2015). Peer assessment of aviation performance: Inconsistent for good reasons. *Cognitive Science*, *39*(2), 405–433. https://doi.org/10.1111/cogs.12152

Sembiring, J., Liu, C., Koppitz, P., & Holzapfel, F. (2018). Energy management for unstable approach detection. Paper presented at the 1-6. doi:10.1109/ICARES.2018.8547140

Shappell, S. A., & Wiegmann, D. A. (2000). *The human factors analysis and classification System - HFACS.* (No. DOT/FAA/AM-00/7). National Technical Information Service.

Shelby, C. A., Cross, D., & Szathmary, K. (2013). Integrating flight operational quality assurance into the corporate pilot safety culture. *Journal of Interdisciplinary Collaboration, 1*(1), 4.

Shelton, C. L., Kinston, R., Molyneux, A. J., & Ambrose, L. J. (2013). Real-time situation awareness assessment in critical illness management: Adapting the situation present assessment method to clinical simulation. *BMJ Quality & Safety, 22*(2), 163. doi:10.1136/bmjqs-2012-000932

Sheridan, T. B. (2010). The systems perspective on human factors in aviation. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (pp. 23-63). Academic Press.

Sherry, L., Wang, Z., Kourdali, H. K., & Shortle, J. (2013). Big data analysis of irregular operations: Aborted approaches and their underlying factors. *2013 Integrated Communications, Navigation and Surveillance Conference (ICNS).* doi:10.1109/ICNSurv.2013.6548548

Shish, K. H., Kaneshige, J., Acosta, D. M., Schuet, S., Lombaerts, T., Martin, L., & Madavan, A. N. (2015). Trajectory prediction and alerting for aircraft mode and energy state awareness. *AIAA Conference Proceeding*s. doi:10.2514/6.2015-1113

Singh, N. P., Goh, S. K., & Alam, S. (2020). Real-time unstable approach detection using sparse variational gaussian process. *2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT)*, 1-10. doi:10.1109/AIDA-AT48540.2020.9049174

Stanton, N. A., Salmon, P. M., Walker, G. H., Salas, E., & Hancock, P. A. (2017). State-of-science: Situation awareness in individuals, teams and systems. *Ergonomics, 60*(4), 449-466. doi:10.1080/00140139.2017.1278796

Stelzer, E. M., & Wickens, C. D. (2006). Pilots strategically compensate for display enlargements in surveillance and flight control tasks. *Human Factors, 48*(1), 166-81. http://ezproxy.libproxy.db.erau.edu/login?url=https://www-proquest-com.ezproxy.libproxy.db.erau.edu/scholarly-journals/pilots-strategically-compensate-display/docview/216460474/se-2?accountid=27203

Stinton, D. (1987). *The anatomy of the aeroplane*. BSP Professional Publications.

Sumwalt, R., Cross, D., & Lessard, D. (2015). Examining how breakdowns in pilot monitoring of the aircraft flight path. *International Journal of Aviation, Aeronautics, and Aerospace, 2*(3). https://doi.org/10.15394/ijaaa.2015.1063

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press.

The Boeing Company [TBC]. (2007). 777 flight crew training manual: Qatar airways (Rev. 1). http://www.anythingaboutaviation.com/wp-content/uploads/2013/01/Boeing-B777-FCTM-Flight-Crew-Training-Manual.pdf

Thaler, R. H., & Sunstein, C.R. (2021). *Nudge: The final edition*. Penguin Books.

Thirtyacre, D. L. (2021). *The effects of remotely piloted aircraft command and control latency during within-visual-range air-to-air combat*. (Doctoral dissertation, Embry-Riddle Aeronautical University). ProQuest Dissertations Publishing (Order No. 28493218).

Transport Safety Board of Canada [TSB]. (2018). Aviation investigation A18A0085. http://www.bst-tsb.gc.ca/eng/enquetes-investigations/aviation/2018/a18a0085/a18a0085.asp

Truong, D. (2017). *Introduction to multivariate statistics* [Video file]. https://www.youtube.com/watch?v=F12jDF69X3E.

Tsang, P. S. (2003). Assessing cognitive aging in piloting. In P.S. Tsang & M. A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 507-546). CRC Press.

Tsang, P. S., & Vidulich, M. A. (2003). Introduction to aviation psychology. In P.S. Tsang & M. A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 1-19). CRC Press.

United Kingdom Civil Aviation Authority [UK CAA]. (2012). *Flight data monitoring based precursors project: Part 1 – runway excursions.* Civil Aviation Authority.

Vidulich, M. A. (2000). The relationship between mental workload and situation awareness. In *Proceedings of the IEA 2000/HFES 2000 Congress* (Vol. 3, pp. 115-146). Human Factors and Ergonomics Society.

Vidulich, M. A. (2003). Mental workload in situation awareness: Essential concepts for aviation psychology practice. In P.S. Tsang & M. A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 115-146). CRC Press.

Vidulich, M. A., Wickens, C. D., Tsang, P. S., & Flach, J. M. (2010). Information processing in aviation. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (pp. 175-215). Academic Press.

Vine, S. J., Uiga, L., Lavric, A., Moore, L. J., Tsaneva-Atanasova, K., & Wilson, M. R. (2015). Individual reactions to stress predict performance during a critical aviation incident. *Anxiety, Stress, & Coping*, *28*, 467–477. doi:10.1080/10615806.2014.986722

Vogt, P., Gardner, D., & Haeffele, L. (2012). *When to use what research design*. The Guilford Press.

Vogt, P., Vogt, E., Gardner, D., & Haeffele, L. (2014). *Selecting the right analyses for your data*. The Guilford Press.

Wang, Y., Guo, X., Liu, Q., Xiao, H., & Bai, Y. (2018). Three-dimensional measurement applied in design eye point of aircraft cockpits. *Aerospace Medicine and Human Performance, 89*(4), 371.

Wang, Z. (2016). *A methodology for nowcasting unstable approaches* (Order No. 10135037). http://search.proquest.com.ezproxy.libproxy.db.erau.edu /docview/1811448449?accountid=27203

Wang, Z., Sherry, L., & Shortle, J. (2015). Airspace risk management using surveillance track data: Stabilized approaches. *2015 Integrated Communications, Navigation and Surveillance Conference (ICNS).* doi:10.1109/ICNSURV.2015.7121351

Watts, B.D. (2004). *Situation awareness in air-to-air combat* (McNair Paper No. 68). National Defense University.

West, P. D. (2011). Solution design. In G. S. Parnell, P. J. Driscoll, & Dale L. Henderson (Eds.), *Decision making in systems engineering and management* (2nd ed., pp. 353-393). John Wiley & Sons.

Whitlow, S. D., & Dillard, M. B. (2019). Intelligent modules and advanced displays to support pilot airplane system state awareness. *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC).* doi:10.1109/DASC43569.2019.9081762

Wickens, C. D. (2000). The tradeoff of design for routine and unexpected performance: Implications of situation awareness. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 211-226). Lawrence Erlbaum.

Wickens, C. D. (2003a). Aviation displays. In P.S. Tsang & M. A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 147-200). CRC Press.

Wickens, C. D. (2003b). Pilot actions and tasks: Selections, execution, and control. In P.S. Tsang & M. A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 239-263). CRC Press.

Wickens, C. D. (2005). Attentional tunneling and task management. *2005 International Symposium on Aviation Psychology*, 812-817. https://corescholar.libraries.wright.edu/isap_2005/121

Wickens, C. D. (2016). Situation awareness and workload in aviation. *Current Directions in Psychological Science : A Journal of the American Psychological Society*, *11*(4), 128-133. doi:10.1111/1467-8721.00184

Wickens, C. D., & Carswell, C. M. (1997). Information processing. In G. Salvendy (Ed.), *Handook of human factors and ergonomics* (pp. 89-129). John Wiley & Sons.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology & human performance* (3rd ed.). Pearson Prentice Hall.

Wickens, C. D., Hollands, J. G., Banbury, S., and Parasuraman, R. (2015). *Engineering psychology & human performance* (4th ed.). Psychology Press.

Wickens, C. D., Lee, J. D., Gordon, S. E., & Liu, Y. (2004). *An introduction to human factors engineering* (2nd ed.). Pearson Prentice Hall.

Wickens, C. D., & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 30*(5), 599-616. doi:10.1177/001872088803000505

Wilson, D., & Binnema, G. (2014). *Managing risk: Best practices for pilots*. Aviation Supplies & Academics.

Yeh, M., & Chandra, D. (2004). Issues in symbol design for electronic displays of navigation information. *Proceedings of the 23rd DASC Conference*. Salt Lake City, UT.

Young, L. R. (2003). Spatial orientation. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 69-113). CRC Press.

Yu, C., Wang, E. M., Li, W., & Braithwaite, G. (2014). Pilots' visual scan patterns and situation awareness in flight operations. *Aviation, Space, and Environmental Medicine, 85*(7), 708-714. doi:10.3357/ASEM.3847.2014

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. doi:10.1080/10508414.2017.1313096

## Appendix A: Baseline Primary Flight Display Presentation

The Primary Flight Display (PFD) allows only for display of information necessary to follow the approach guidance, such as vertical and lateral path, airspeed, altitude, heading, and roll angle. Currently, such information is presented only in raw data form and is limited to target speeds and highlighted altitudes. An example of the PFD installed on variants of the Boeing B-737 is presented in Figure A1.

**Figure A1**

*B-737 Primary Flight Display*

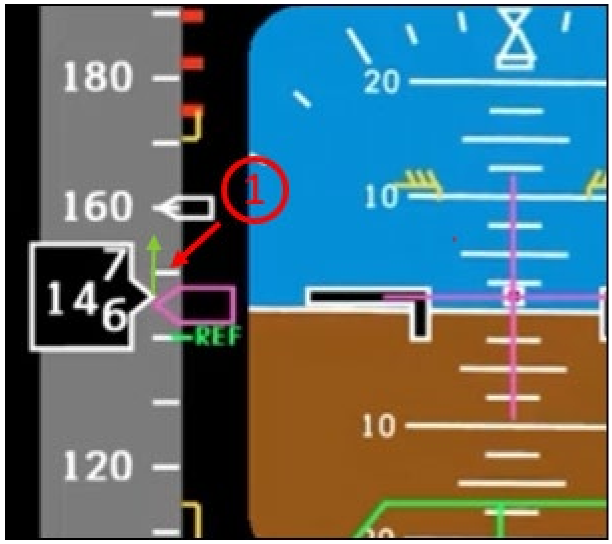

*Note*. Adapted from "Prepar3D®" by Lockheed Martin, 2017 (http://www.prepar3d.com/prepar3d-store/). Copyright by Lockheed Martin.

The particular PFD presentation shown is representative of a B-737 on a stabilized Instrument Landing System (ILS) final approach. Callout 1 identifies the programmed approach reference speed, 140 kts, calculated by, and entered through, the FMS as a function of airplane weight and the selected landing flap setting. Callout 2 shows the magenta command airspeed cue for the approach, manually set by the pilot to 145 kts, and is nominally set to 5 kts above the reference speed in a no-wind condition. Callout 3 provides the actual airspeed and indicates slightly greater than 146 kts. At Callout 4, the magenta vertical and lateral FD guidance bars provide indication of flight path corrections to be followed to track the ILS localizer and glideslope path; it provides the pilot the appropriate direction and magnitude of correction to return to the flight path, saving the pilot from calculating the needed response. Course and path correction accuracy is achieved by placing the adjacent small white box, representative of the airplane longitudinal reference, so as to be superimposed over the FD guidance bars. Callout 5 identifies the bank angle pointer, which measures deflection in roll from the vertical against a scale of increments set at 10°, 20°, 30°, 45°, and 60°. Callout 6 indicates the current RA in both round analog and digital format. The display shows the airplane at 740 ft RA. Callouts 7 and 11 point to the ILS glideslope and localizer magenta diamonds, respectively, that provide raw data information on displacement from the path. The current altitude box, Callout 8, depicts the airplane at 1,070 ft pressure altitude – a measure against a sea level datum. Callout 9 identifies the vertical speed scale and pointer, indicating the rate of climb or descent the airplane is undergoing. The current value is 800 fpm of descent. Finally, Callout 10 shows the minimum RA for the approach, set at 250 ft.

An additional display feature, the velocity trend vector, is shown by Callout 1 in Figure A2. The display uses a green arrow that grows out from the tip of the current airspeed box. This green arrow depicts the projected airspeed for 10 s in the future, based on the current acceleration or deceleration. The tip of the arrow aligns with the expected value. There is no digital presentation of the value.

**Figure A2**

*Velocity Trend Vector*



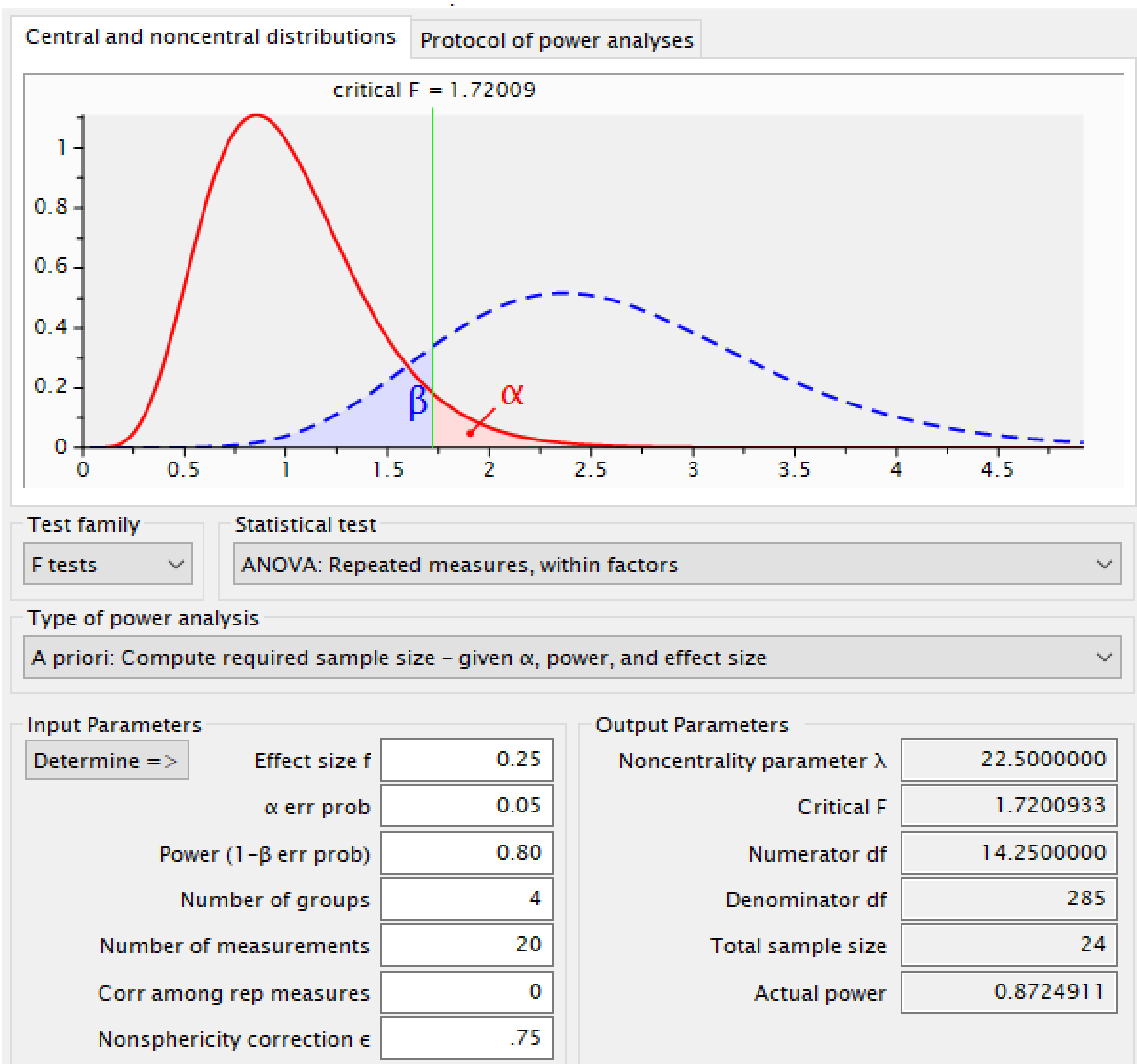

*Note*. Adapted from "Prepar3D®" by Lockheed Martin, 2017 (http://www.prepar3d.com/prepar3d-store/).

Copyright by Lockheed Martin.

## Appendix B: Impact of Effect Size and Power on Total Sample Size

**Figure B1**

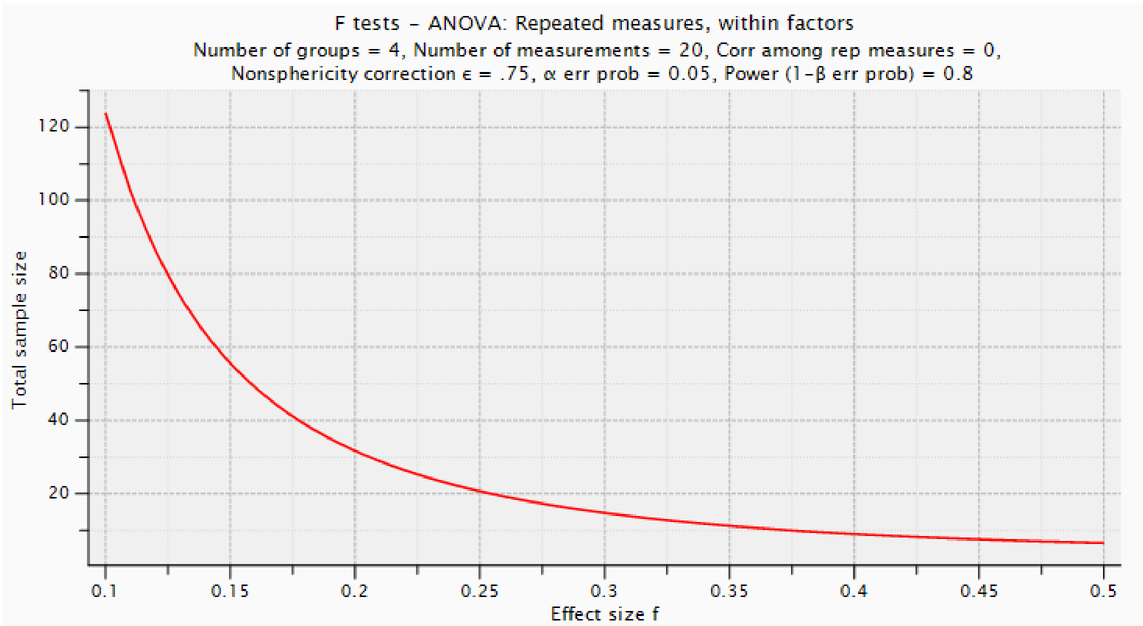
*G*Power Assessment of Sample Size*




*Note*. Excerpted from G*Power (Version 3.1.9.2), by F. Faul, E. Erdfelder, A. G. Lang, and A. Buchner, 2014. Copyright 2014 by Heinrich-Heine-Universität Düsseldorf.

**Figure B2**

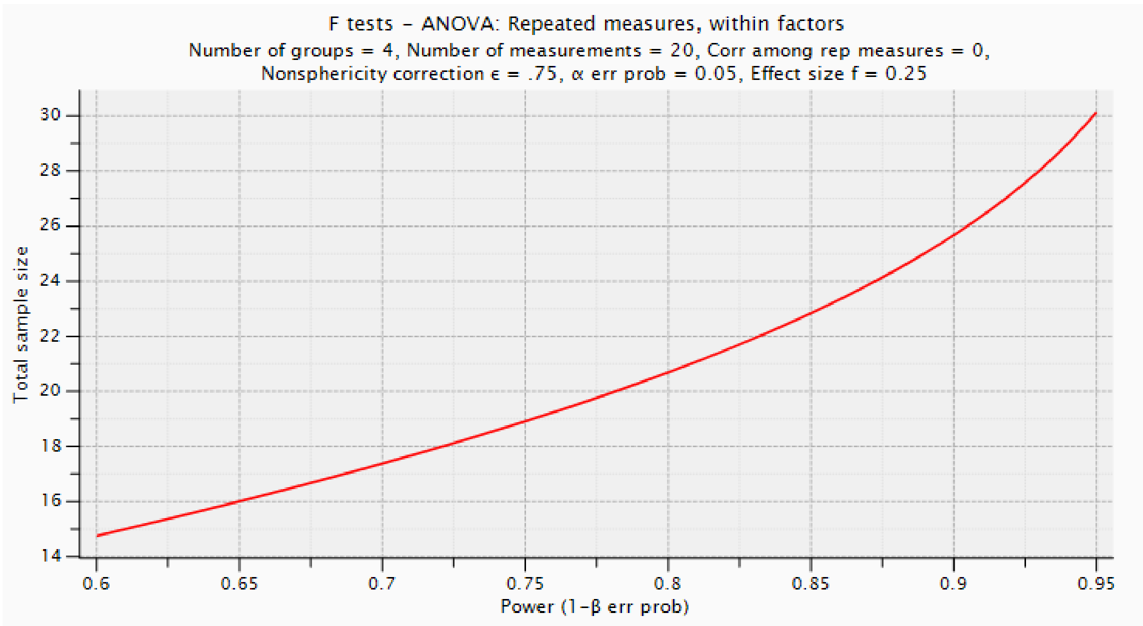
*The Effect of Effect Size (f) on Total Sample Size*




*Note*. Excerpted from G*Power (Version 3.1.9.2), by F. Faul, E. Erdfelder, A. G. Lang, and A. Buchner, 2014. Copyright 2014 by Heinrich-Heine-Universität Düsseldorf.

**Figure B3**

*The Effect of Power (1-β) on Total Sample Size*




*Note*. Excerpted from G*Power (Version 3.1.9.2), by F. Faul, E. Erdfelder, A. G. Lang, and A. Buchner, 2014. Copyright 2014 by Heinrich-Heine-Universität Düsseldorf.

# Appendix C: Goal Directed Task Analysis

## Targeted Role Description

Pilots are assigned the duty of safe operation of an airplane, and this duty extends to fellow flight crew. This responsibility includes landing on the assigned runway and being able to decelerate to a complete stop within the longitudinal and lateral confines of the surface. The Code of Federal Regulations (C.F.R.) Title 14 Part 91.3(a) addresses the responsibility and authority of the pilot in command, stating "The pilot in command of an aircraft is directly responsible for, and is the final authority as to, the operation of that aircraft" (FAA, 2017e).

Airplanes may be operated by one or more pilots, depending on the airplane type design. During the manufacturer design process and subsequent airplane type certification, the FAA Aircraft Evaluation Group mandates the minimum flight crew complement. With few exceptions, transport category airplanes incorporate a two-pilot flight crew that act as a team. Areas of responsibility are established by the individual operator and align under either the pilot flying (PF) or the pilot monitoring (PM) role. The PF is primarily tasked with physical control of the airplane to achieve the desired flight path; the PM is normally tasked with monitoring the status of the airplane systems, staying abreast of the current and projected flight path, running checklists and communicating with ATC, and supporting the PF. To operate the airplane properly, pilots must collectively maintain a high degree of situation awareness and exhibit effective communications.

This design study targets the role of the collective flight crew team as it relates to the task of landing the airplane safely on the runway surface. Failure to consider the crew aspect risks the possibility of missing information due to bias resulting from the distribution of tasks. A Goal-Directed Task Analysis (GDTA) was completed with the intent of gathering detailed information on the PFD presentation and its relationship to the landing task.

**Method Applied**

The expanse of runway excursion causal factors was constrained to unstable approaches in order to limit the scope of the design study. Consequently, the flight phase prior to the instrument approach final approach fix, which marks the beginning of the final descent to runway, was deemed irrelevant to the study. Equally, actions following the initiation of the airplane flare maneuver nominally accomplished 50 ft above the runway were not considered. Accomplishing a stable approach to arrive at the point of flare initiation (1) within the appropriate landing zone, and (2) devoid of excess energy or roll angles was the targeted task. Thus, data collection focused on flight crew cognitive goals during the flight period bracketed by those action points.

The GDTA was completed through three SME interviews conducted with flight crew members. The flight crew members were fully trained and qualified to operate their assigned airplane and held operational currency, meaning they had received recurrent training and/or operational experience within the last six months. Airplanes represented included the Dassault Falcon 900 and the Boeing 737. Questions were crafted following the guidance provided by Endsley and Jones (2012) and focused on the goals, subgoals, decisions, and situation awareness requirements of flight crew during the approach task. Design study group members with flight experience were tasked with drafting the initial set of questions to be presented.

Interviews were conducted on an individual, one-on-one basis and captured for later reference and documentation. Each session began with an administrative briefing on the expected duration of the interview and privacy protections for both data and identification, an introduction to the purpose of the interview and the intended use of the data collected, and a short explanation of the GDTA process. Interview participants were briefed on the intent to gather both written notes and audio recordings during the session. Although each interview began with a list of suggested questions to follow, the flow of conversation was allowed to seek its own course when the SME felt specific material should be covered. Focus was maintained on specific cognitive goals and information requirements to avoid fixation on detailed technology solutions or PFD display specifics. The duration of interview for participants 1, 2, and 3, were 40, 25, and 30 m, respectively.

At the completion of the interview sessions, data were transcribed from the audio recordings, collated with written notes, and summarized; information garnered from the interview process was combined with knowledge obtained through academic research and an in-depth review of various PFD displays and flight crew instrument approach procedures. The data were then analyzed to identify the structure of the GDTA, inclusive of the goals, subgoals, decisions, and situation awareness requirements at the perception (Level 1), comprehension (Level 2), and prediction (Level 3) levels.

## Appendix D: Proposed Primary Flight Display Presentations

The display enhancements are presented in Figure D1, D2, and D3. Proposed display enhancement Treatment 1 seeks to address four performance parameter boundaries to be called out on a continuous basis, against which the pilot assesses the current state to determine whenever these parameters are exceeded.

**Figure D1**

*Application of Boundaries*


*Note*. Flight instrument display modifications for enhanced pilot stabilized approach situation awareness using airspeed, roll angle, and vertical speed boundaries presentation. Adapted from "Prepar3D®" by Lockheed Martin, 2017 (http://www.prepar3d.com/prepar3d-store/). Copyright by Lockheed Martin.

Callout 1 in Figure D1 addresses the airspeed criteria boundaries for a stabilized approach. Using an amber horizontal line, the marked areas reflect both the high- and low-speed values beyond which the approach becomes unstable. Callout 2 establishes the roll angle deviation allowance to remain stable, using the criteria of ±6° deviation. For this study, an amber arc would display upon reaching 1,000 ft RA. In Callout 3, the maximum descent rate of 1,000 fpm is depicted using an amber diagonal line. Callout 4 assists in recognition of the point of 1/2 dot vertical glidepath deviation. Callout 5 aids in defining the 1/2 dot localizer course deviation.

The incorporated PFD velocity trend vector would support the design in demonstrating when the projected airspeed would fall above or below the limit values. Roll and descent rates necessary for Level 3 SA prediction of the flight path could be assessed by time-sequenced repetitive sampling of the current roll angle and vertical speed. Note that the boundary presentations all fall within a narrow vertical segment of the display, incorporating the concept of SAOD Principle 21, which proposes grouping information to support Level 2 and 3 situation awareness requirements, and does so in support of display density reduction espoused by Principle 22. As the FSF criteria do not address level flight conditions or climbs, these cases are not considered.

Proposed display enhancement Treatment 2, depicted in Figure D2, seeks to address pilot perception of stabilized approach parameter exceedance, but does so through the application of a single caution-level alert cue. Specific criteria values are not presented, as is the case in Treatment 1. The amber *UNSTABLE* caution display at Callout 1 would present continuously whenever any of the criteria boundaries – airspeed, roll, vertical speed, course, and vertical and lateral path – are exceeded. Crews would be expected to execute the approach without depiction of the trigger criteria for display of the caution on the PFD. The display would be shown until the parameter is corrected to within tolerance.

**Figure D2**

*Application of Alerts*


*Note*. Flight instrument display modifications for enhanced pilot stabilized approach situation awareness using an alert display. Adapted from "Prepar3D®" by Lockheed Martin, 2017 (http://www.prepar3d.com/prepar3d-store/). Copyright by Lockheed Martin.

The application of the amber *UNSTABLE* caution-level alarm on the PFD would provide a single indication of the need to correct the airplane condition or consider a go-around procedure and do so without the directive nature of a red warning-level display. This caution-level placement would support the pilot making the decision to terminate the approach.

The final display enhancement is presented in Figure D3. This treatment employs both the unstable approach boundaries provided in the first treatment and the alert display provided in the second treatment. The interpretation of Callouts 1, 2, 3, 4, and 5 remain the same, as does the inclusion of the amber *UNSTABLE* alert at Callout 6.

**Figure D3**

*Application Using Combined Boundaries and Alerts*


*Note*. Flight instrument display modifications for enhanced pilot stabilized approach situation awareness using a combination of the stable approach boundaries and alert displays. Adapted from "Prepar3D®" by Lockheed Martin, 2017 (http://www.prepar3d.com/prepar3d-store/). Copyright by Lockheed Martin.

## Appendix E: Measurement Instruments

### Pre-trial Demographic Questionnaire

**Participant Number: ___________________**

**Participation Requirements**

1. What is your current age?
   __________ Years

2. Do you possess a current driver license or FAA medical certificate?
   a. Yes
   b. No

3. Which pilot certificate do you currently hold?
   a. Private
   b. Commercial
   c. ATP

4. Do you hold an instrument rating or an ATP?
   a. Yes
   b. No

5. Do you hold an airplane single- or multi-engine land category and class rating?
   a. Yes
   b. No

6. Do you possess instrument currency within the past five years?
   a. Yes
   b. No

**General Information**

7. What is your highest level of education?
   a. High School
   b. Some undergraduate study
   c. Undergraduate degree
   d. Some graduate study
   e. Graduate degree
   f. Post graduate study

8. In which type of flight operation are you currently employed/engaged?
   a. Training (Part 61/141/142/Other)
   b. Corporate
   c. Commercial (Part 121/125/135/Other)

d. Military

9. Approximately how many years have you been flying?
__________ Years

10. What is your approximate total flight time in hours?
__________ Hours

11. What is your approximate total number of flight hours logged in simulated or actual instrument conditions?
__________ Hours

12. What is your approximate number of instrument approaches conducted?
__________ Approaches

**Study-specific Information**

13. What is your prior experience, in terms of flight hours, with electronic flight displays – PFD, ND, EFIS, or other “glass cockpit” presentation – in your flight history?
__________ Hours

14. What is your prior experience, in terms of years, in flying under defined stable approach criteria?
__________ Years

15. How many unstable approaches have you experienced in your flying career?
__________ Approaches

16. How many months have passed since your most recent unstable approach?
__________ Months

17. What was the primary indication of your unstable approach?
    a. Excessive airspeed
    b. Insufficient airspeed
    c. Excessive descent rate
    d. Inappropriate thrust setting
    e. Incorrect airplane configuration
    f. Incomplete checklist/briefing actions
    g. Flight path deviation
    h. Not applicable

18. What was your action upon awareness of the most recent unstable approach?
    a. Continued the approach to a successful landing
    b. Continued the approach to a less-than-successful landing
    c. Continued the approach and later executed a go-around

d. Executed an immediate go-around
e. Not applicable

**Situation Awareness Data Collection Questionnaire**

When presented an unstable approach as define by the criteria provided, state aloud the following information as *timely* and *accurately* as possible. You will have five (5) seconds to provide the information.

| | |
|---|---|
| □ Condition: | “Unstable” |
| □ Exceedance: | “Left of Course,” “Airspeed Low,” or others |
| □ Action: | “Turn Right,” “Increase Airspeed,” or others (1,000 - 500 ft AGL) |
| | “Go-around” (below 500 ft AGL) |

**Post-trial Usability Questionnaire**

**Participant Number: __________________**

For each question below, circle the number that corresponds to your perceptions of the displays used for this study.

1. The display presented was easy to read.

**Basic**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries + Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

2. The display presented did not interfere with completion of the monitoring task.

**Basic**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries + Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

3. The display presented enhanced my perception and understanding of the current airplane stable approach state.

**Basic**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries + Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

4. The display presented enhanced my prediction of the future airplane stable approach state.

**Basic**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries + Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

5. The display presented would be beneficial in *flight* operations.

**Basic**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries + Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

6. The display presented would be beneficial in *training* operations.

**Basic**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

**Boundaries + Alert**

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

7. Rank order the displays in terms of pilot situation awareness enhancement. Indicate from highest to lowest – 1$^{st}$, 2$^{nd}$, 3$^{rd}$, and 4$^{th}$:

   Basic _______
   Boundaries _______
   Alert _______
   Boundaries + Alert _______

8. For the final question, do you have any additional comments or suggestions concerning the proposed displays as they relate to improving your awareness of an unstable approach?

# Appendix F: Institutional Review Board Documents

**Embry-Riddle Aeronautical University**
**Application for IRB Approval**
**Limited or Expedited Determination**

**Principal Investigator:** David J. Hunter

**Other Investigators:** Andrew R. Dattel, Ph.D.

**Role:** Student **Campus:** Daytona Beach **College:** Aviation/Aeronautic

**Project Title:** SITUATION AWARENESS ASSESSMENT OF ENHANCED STABLE APPROACH FLIGHT INSTRUMENT DISPLAYS

**Review Board Use Only**

---

**Initial Reviewer:** Teri Gabriel **Date:** 05/08/2023 **Approval #:** 23-120

**Exempt:** No

IRB Member Reviewer Signature: Christine Walck Digitally signed by Christine Walck Date: 2023.05.22 14:24:31 -04'00'

Dr. Beth Blickensderfer IRB Chair Signature: Elizabeth L. Blickensderfer Digitally signed by Elizabeth L. Blickensderfer Date: 2023.05.23 15:23:46 -04'00'

**Brief Description:**
The purpose of this proposed study is to determine whether pilot situational awareness (SA) when engaged in an unstable approach can be improved by applying enhancements to the flight instrument displays. The proposed study applies the principles of user-centered design to identify proper application of display presentation. Further, it measures pilot SA under a representative flight instrument display used on current air carrier airplane systems. The study then assesses the effectiveness of a number of stabilized approach display treatments on the ability of the pilot to achieve an improved state of SA. Participants will be asked to observe a series of instrument approach video vignettes of the instrument displays. The vignettes will be designed to either highlight a particular element of an unstable approach to be observed by the participant. Participants will be assessed as to their level of SA during each trial.

This research falls under the expedited category as per 45 CFR 46.110 (b) because one of the following apply:

(1) ☒ some or all of the research appears on the list provided by the Office of Human Research Protections and/or are found by the reviewer(s) to involve no more than minimal risk;

(2) ☐ minor changes in previously approved research during the period for which approval is authorized;

(3) ☐ research for which **Limited IRB review** is a condition of Exemption;

[Under an expedited review procedure, the review may be carried out by the IRB chairperson or by one or more experienced reviewers designated by the chairperson from among members of the IRB. In reviewing the research, the reviewers may exercise all of the authorities of the IRB except that the reviewers may not disapprove the research. A research activity may be disapproved only after review in accordance with the nonexpedited procedure set forth in §46.108(b).]

a. Prospective collection of biological specimens for research purposes by noninvasive means.

b. Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies. (NOTE: Some research in this category may be exempt from the HHS regulations for the protection of human subjects §46.101(b)(2) and (b)(3). This listing refers only to research that is not exempt.) [This means research that presents *more than minimal risk to human subjects.*]

(3) Research for which **limited** IRB review is a condition of exemption as follows:

a. Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) if the information obtained is recorded by the investigator in such a manner that the identity of the human subjects can readily be ascertained, directly or through identifiers linked to the subjects. §46.104(d)(2)(iii)

b. Research involving benign behavioral interventions in conjunction with the collection of information from an adult subject through verbal or written responses (including data entry) or audiovisual recording if the subject prospectively agrees to the intervention and information collection and the information obtained is recorded by the investigator in such a manner that the identity of the human subjects can readily be ascertained, directly or through identifiers linked to the subjects. §46.104(d)(3)(i)(C)

c. Storage or maintenance for secondary research for which broad consent is required: Storage or maintenance of identifiable private information or identifiable biospecimens for potential secondary research use. §46.104(d)(7)

d. Secondary research for which broad consent is required: Research involving the use of identifiable private information or identifiable biospecimens for secondary research use, if the following criteria are met:

(i) Broad consent for the storage, maintenance, and secondary research use of the identifiable private information or identifiable biospecimens was obtained.

(ii) Documentation of informed consent or waiver of documentation of consent was obtained. §46.104(d)(8)

EMBRY-RIDDLE
Aeronautical University

**2 May 2023**

# Human Subject Protocol Application

| | | | |
|---|---|---|---|
| Campus: | Daytona Beach | College: | COA |
| Applicant: | David Hunter | Degree Level: | Doctorate |
| ERAU ID: | 2381822 | ERAU Affiliation: | Student |
| Project Title: | Situation Awareness Assessment of Enhanced Stable Approach Flight Instrument Displays | | |
| Principal Investigator: | David J. Hunter | | |
| Other Investigators: | Andrew R. Dattel, Ph.D. | | |
| Submission Date: | 05/02/2023 | | |
| Beginning Date: | 05/08/2023 | | |
| Type of Project: | Experiment | | |

## Questions

1. **Background and Purpose**: Briefly describe the background and purpose of the research. Include how the study contributes to existing knowledge; spell out acronyms the first time they are used; and use consistent terminology.

The use of instrumentation in developing pilot situation awareness (SA) has been demonstrated in both theory and application in the aviation field. Much of the historic research focus has been directed at novel methods of presenting airplane performance parameters and airplane system monitoring. Less prevalent is academic research as to the impact of enhanced flight instrument displays in improving flight crew, and more specifically individual pilot, SA as to the presence of an unstable approach.

Pilots currently must depend on the timely recall of stable approach criteria stored in long-term memory, application of those criteria against current flight conditions, perception and cognition of deviation from that criteria, and determination of potential courses of action. Such recall is fallible. Enhanced flight instrument displays have the potential to assist pilots in assessing the current airplane performance state against proven stable approach criteria and serve as an effective means to support the aeronautical decision-making process as to whether to continue or abort an approach. Unfortunately, little research has been conducted to validate this potential.

The purpose of this study is to determine whether pilot SA when engaged in an unstable approach can be improved by applying enhancements to the flight instrument displays. The study applies the principles of user-centered design to identify proper application of display presentation. Further, it measures pilot SA under a representative flight instrument display used on current air carrier airplane systems. The study then assesses the effectiveness of a number of stabilized approach display treatments on the ability of the pilot to achieve an improved state of SA.

2. **Design, Procedures and Methods**: Describe the details of the procedure(s) to be used; how the data will be collected and/or what will be done to collect the needed data.

Quantitative research methodology using an experimental design will be used. A within-group, repeated measures experiment will gather the data necessary to investigate the research question and hypotheses. Participants will be asked to observe a series of instrument approach video vignettes presenting the instrument displays. The vignettes will be designed to highlight a particular element of an unstable approach to be observed by the participant. Participants will be assessed as to their level of SA during each trial.

The hypotheses in the study are established based on a 2x2 within-group design with two independent variables (IVs) and one dependent variable (DV). The first IV manipulates the flight displays by employing stability criteria boundaries, and the second IV manipulates the flight displays by employing an alert message. The state of each of the IVs will address the presence or lack of a specific display treatment, while the DV will address the accuracy in responses to unstable conditions. Think aloud protocols will be employed.

Measures and observations can be viewed as those accomplished in the pre-trial, trial, and post-trial phase. In the pre-trial phase, the participant will be asked to complete a pre-trial demographic questionnaire to ensure minimum participation requirements and gather data on participation commonalities and diversity. During the trial phase, participants will be asked to respond in as timely and accurately as possible to an unstable condition, noting (1) the presence of an unstable condition, (2) the boundary being exceeded, and (3) the action to be taken. The principal data to be collected, response accuracy, will be gathered using commercial off-the-shelf recording software. In the post-trial phase, participants will complete a post-trial usability questionnaire designed to capture participant views of the treatments experienced. This questionnaire will incorporate a 5-point Likert design with open field comment and focus on subjective impressions of treatment effectiveness and overall display usability, as well as solicit comments and offer suggestions for improvement.

EMBRY-RIDDLE
Aeronautical University

**2 May 2023**

Following an introductory vignette to calibrate each pilot to the particulars of the display and the experiment process, pilots will be presented 20 different scored trial scenarios. The number of trials chosen was to ensure a representative sampling of each of the bounds and random presentation order will be used to reduce the potential for learning effects and experimental fatigue. Each vignette is approximately 150 seconds duration. Within the experiment, participants will observe each vignette on a high-definition monitor. Common modern flight instrumentation will be displayed. The environment will be controlled to avoid the introduction of external factors, – lights, noise, sound, movement – that might distract from the principal task. Upright, comfortable seating will be used.

Participants will be tasked with observing the vignette, paying particular attention to the primary flight and navigation displays afforded. Participants will not actually be flying as the variation of pilot skill would serve as a confounding variable; that is, some pilots would hold and be able to employ sufficient piloting skill to avoid a particular unstable approach in comparison to others who may not. Instead, a secondary task will be used to increase the participant workload and emulate the mechanical actions involved in moving flight and thrust control inceptors. Pilots will be presented visual markers within their field of regard at a frequency commensurate with flight and thrust control inputs experienced in actual flight under nominal conditions. Markers will be to the left and right of the display and correspond to the location of the flight control and thrust inceptors in the airplane. When each marker displays, the participant will depress an inline handheld switch in the corresponding hand, illuminating a corresponding LED that only the researcher will see. Data will not be collected on this activity as it is inconsequential to the DV data being collected. However, compliance will be monitored by the researcher and the participant reminded to pay attention to the secondary task when accuracy drops below 80%.

a. Will the activity be RECORDED?

Yes

Check all that apply: Audio, Video

Include how and on what device the data will be collected.

Audio of participant think aloud output will be recorded commensurate with the video being presented. Participants will wear a Logitech H390 lightweight headset with attached electret microphone to capture audio output. The headset will be attached to the research laptop. Vignette video will be recorded simultaneously to synchronize participant output data with each vignette. Commercial off-the-shelf components will be used for data input. The data will be collected on the research laptop system used for the study for later analysis. The system will be portable to allow for easy movement between data location locations.

b. LOCATION: Indicate where the activity will take place –

**Embry-Riddle**:

Campus

Prescott

Specify where the project will take place by including the building name and office/lab number:

Building AC1, Room 238.

**External-Outside Embry-Riddle** – specify where the project will take place:

Boeing Miami Flight Training Center – Building 7207-5709/Room 101 – Miami, FL
Boeing Global Engagement & Training/Boeing Test & Evaluation – Building 2-122, 3rd Floor – Seattle, WA
The Anschutz Corporation Flight Operation – Hanger 12/Conference Room - Centennial Airport, Englewood, CO

3. **Time**: Include how much time will be asked of each participant. Include the amount of time it takes for each activity and the total time. The total amount of time must match what is written on the Informed Consent Form (ICF), but the ICF only need include the total amount of time. (Do NOT include the amount of time needed to read the ICF.)

The total time for each participant to complete the session will be approximately 80 minutes, inclusive of issuance of standardized participant instructions, completion of a participant pre-trial demographic questionnaire, presentation of a display familiarization and demonstration vignette, accomplishment of the trial vignettes and data collection, completion of a post-trial questionnaire, and accomplishment of the debrief session. The planned flow of events and allocated time for each event is presented in the table, below:

Event: Duration

Participant Instructions: 5 minutes
Pre-trial Demographic Questionnaire: 5 minutes
Display Familiarization/Demonstration Vignette: 5 minutes
Participant Question Period: 5 minutes
Collections Vignettes: 50 minutes

Post-trial Usability Questionnaire: 5 minutes
Debrief: 5 minutes

Total Time: 80 minutes

4. **Measures and Data to be Collected**: What measures and data will be collected in the study? How will the measures and/or data be collected?

During the pre-trial phase, participant demographic data will be collected to include various categories of flight time, pilot certifications, and experience with unstable approaches. The data collection form is attached to the application. Post-trial phases data will be an assessment of suitability of the treatment to flight operations and training, using Likert scales and an open field comment option. The post-trial data collection form is attached to the application. The trial phase will collect data on vocalized participant response to observed unstable conditions to be scored against truth data derived from subject matter expert inputs. The measure collected is the accuracy of the participant response to unstable approach conditions presented during the vignette.

5. **Participant Population and Recruitment Procedures**:
   a. Who will be recruited to be participants? Check ALL that apply:

      Embry-Riddle Students, Outside Embry-Riddle

   b. Approximately how many participants do you hope to recruit?

      A minimum of 24 participants are expected to meet statistical rigor.

   c. Explain how and where recruitment will be conducted? (Emails, mailings, sign-up sheets, social media, flyers, etc.)

      Participants for the study will be recruited and randomly selected from universities, training centers, and corporate and air carrier volunteer pools. Recruitment will focus on individuals rather than through companies or institutions. Participants will be assessed as to flight experience based on the measure of total flight time. Levels are set at less than 500 hours, 500-1,500 hours, and greater than 1,500 hours total flight time. A balanced sample of experience will be employed. The proposed university location for low experience participants is Embry-Riddle Aeronautical University, Prescott, AZ and qualified pilots operating from local airports. Mid- and high-experience flight training instructors and operational pilots will be sought at Boeing operations at both Miami, FL and Seattle, WA. Mid- and high-experience corporate pilots situated at The Anschutz Centennial Airport, Denver, Colorado, will also be recruited.

      When required, notification of the study will be sent electronically to targeted industry and academic groups explaining its general purpose and the demographics of participants sought. Directions on how to contact the researcher for participation will be provided, as well as the study title, purpose, participation criteria, risks and discomforts, and method for contacting the researcher for data collection dates and locations. Pilots wishing to participate will be initially vetted to ensure they meet the eligibility requirements. Chosen participants will be categorized into levels of flight experience to ensure a balance within the group.

6. **Risks or Discomforts**: Describe any potential risks to the dignity, rights, health or welfare of the human subjects and how these risks will be mitigated. Risks may be physical, psychological, social, legal, economic, to reputation, or others. All other possible options should be examined to minimize any risks to the participants.

The risks in participating in this study are minimal and are not expected to be higher than routine daily life experienced in an office environment. As with any study, slight levels of mental and cognitive stress and fatigue may be present. If the participant experiences any discomfort, the participant may bring the trial to a stop. Centers for Disease Control and Embry-Riddle Aeronautical University sanitation, masking, and safe distancing protocols will be observed. If data collection sites require more stringent practices, they will take precedence.

7. **Benefits**: Assess the potential benefits to be gained by the participants as well as to others in general as a result of this project. If there are no benefits to the participants, state that 'While there are no benefits to the participants...' The benefits here must match what is written on the consent form; here they are written to the IRB reviewer on the consent form they are written directly to the participant.

While there are no benefits to the participants, findings from this research could provide empirical research data and conclusions that might prove beneficial to the academic community, flight instrument display manufacturers, commercial air carrier operations groups, and pilots undergoing flight training in programs that make use of scenarios to introduce the potential for unstable approaches.

8. **Informed Consent**: Describe the procedure you will use to obtain informed consent of the subjects. How and where will you obtain consent? The first page of an electronic survey must be the consent document. See ***Obtaining Participant Consent*** for more information on Informed Consent requirements.

The Informed Consent presentation and concurrence will occur during the pre-trial briefing. This will occur following standardized participant general instructions to ensure the participant is aware of the activities and actions asked of them to follow. The details of the Informed Consent document will then be explained to the participant, followed by an opportunity to read the Informed Consent Form in detail. The participant will then be asked to sign the document. Completion of the Informed Consent document will occur prior to presentation of the demographic questionnaire to ensure all participants have signed prior to any study activities. In the post-trial phase,

EMBRY-RIDDLE
Aeronautical University

**2 May 2023**

verbal debriefing will reiterate the Informed Consent particulars and address any questions they may have prior to leaving the test location.

9. **Confidentiality of Records/Data and Privacy**: Will participant information be:

Confidential

a. Justify the classification and describe the safeguards you will employ to protect participant privacy in securing, sharing, and maintaining data during the study.

Neither the demographic questionnaire, experiment data collection, nor post-trial questionnaire will gather personal identifiable information. Confidential participant information such as name, certifications held, experiences, trial responses, and perceived usability will only be accessible to the researchers and used for the purpose of (a) determining eligibility to be a member of the experiment sample group, and (b) supporting study results and conclusions. Otherwise, names or any other identifying demographics will not be able to be matched. Publication of the experiment results will refer to the data in a generic manner, not inclusive of any identifying information.

To ensure confidentiality, each participant will be assigned a random participant identification number to be attached to the data files. Each participant will subsequently be referred to by their participant identification number. Retention of specific name identification is only necessary for possible follow-up inquiries. The key code for specific names and their correlation to an identification number will be kept under strict control and locked in a separate location.

Data obtained during the study will be collected automatically and manually. Audio and video will be captured for later analysis by the researchers to extract any data from the real-time scenario of each pilot that were not gathered from the assessments. Video will be collected of the screen display and not include any capture of the participant image or identifiable imagery. All data collected in the study will be kept under strict control during execution of the study. The laptop used to collect and store the data will be physically retained in the possession of the principal investigator during trial sessions and stored in a controlled access locked room between collection periods. The laptop and data files will be password protected.

b. Indicate what will happen to data collected from participants that choose to "opt out" during the research process.

All data collected from participants who have chosen to "opt out" during the research will be deleted and purged from the study records.

c. Where and how long will participant data be kept? Include the plan for storage or destruction of data upon study completion. Stating that the data will be destroyed when the Capstone project is completed is NOT acceptable. A specific time period must be indicated. Example: Data will be destroyed three years after completion of the research.

Once the study is complete and the results published, data will be secured using the current Embry-Riddle Aeronautical University policy. Information collected as part of this research may be used or distributed for future research studies and will be retained for a period of three years. Only the PI and the listed other investigator will have access. Paper record of participant data will be destroyed using common shredding methods upon approval of the dissertation.

10. **Economic Considerations/Incentives**: Are participants going to be paid for their participation or are you providing any other type of incentive; including extra credit?

Yes

What will be the compensation or incentive –

**Gift Card** – Specify what kind of gift card, the amount of the gift card and where it can be used/redeemed

Participants will be presented a gift card incentive of $20 value, redeemable at any Starbucks location.

Describe your policy for dealing with participants who start but fail to complete the research.

All participants who initiate the research process will receive the gift card incentive, regardless of whether they complete the research trials or not. Participants who show up for the research but do not consent will not be provided the incentive.

## Appendix G: Informed Consent Form

SITUATION AWARENESS ASSESSMENT OF ENHANCED STABLE APPROACH FLIGHT INSTRUMENT DISPLAYS

**Purpose of this research:** You are being asked to take part in a study for the purpose of determining how pilot situation awareness of stable conditions during approaches can be improved by applying enhancements to the flight instrument displays. Following presentation of general instructions, collection of pre-trial demographic information, and completion of a calibration vignette on the particulars of the display and the experiment process, you will be presented 12 different scored vignettes. Vignettes are of approximately 150 s duration. You will be asked to pay particular attention to primary flight and navigation displays afforded. During the approach you will be presented visual cues to the left and right of the primary flight display. When each cue displays, you will depress an inline handheld switch in the hand corresponding to the side the cue is displayed. Further, you will be asked to verbally respond to observed unstable conditions as timely and accurately as possible, providing three pieces of information. Your audio responses and the corresponding vignette video will be recorded for data analysis. Upon completion of the data collection events, you will be asked to complete a post-trial usability questionnaire and will be provided a debriefing. The total time of your participation is estimated to be 70 minutes.

**Risks or discomforts:** The risks in participating in this study are minimal and are not expected to be higher than routine daily life experienced in an office environment. As with any study, slight levels of mental and cognitive stress and fatigue may be present. If you experience any discomfort, you may bring the trial to a stop. Centers for Disease Control and Embry-Riddle Aeronautical University sanitation, masking, and safe distancing protocols will be observed. If data collection sites require more stringent practices, they will take precedence.

**Benefits:** You will not receive any direct benefit from participating in the study. However, findings from this research could provide empirical research data and conclusions that might prove beneficial to the academic community, flight instrument display manufacturers, commercial air carrier operations groups, and pilots undergoing flight training in programs that make use of scenarios to introduce the potential for unstable approaches.

**Confidentiality of records:** Your individual information will be protected in all data resulting from this study. While the members of the research team will have access to your personal information, publication of the data will not include any identifying information. You will be assigned an identification number; the key code will be stored separately from the data. Information collected as part of this research will not be used or distributed for future research studies.

**Compensation:** You will be presented a gift card incentive of $20 value, redeemable at any Starbucks location. You will receive you will receive the gift card incentive, regardless of whether you complete the research trials or not.

**Contact:** If you have any questions about this research or would like additional information about this study, please contact David J. Hunter, hunted10@my.erau.edu, or the faculty member overseeing this project, Dr. Andrew R. Dattel, dattela@erau.edu. For any concerns or questions as a participant in this research, contact the Institutional Review Board (IRB) at 386-226-7179 or via email teri.gabriel@erau.edu.

**Voluntary participation:** Your participation in this research is completely voluntary. You may discontinue your participation at any time without penalty or loss of benefits to which you are otherwise entitled. Should you decide to discontinue the research at any time, none of the collected information will be used.

**CONSENT:** By signing below, I certify that I meet the participation criteria listed below. I further verify that I understand the information on this form, that the researcher has answered any and all questions I have about this study, and I voluntarily agree to participate in the research.

(a) Hold an instrument qualification, as demonstrated through possession of either a Private or Commercial certificate with an Instrument rating or an FAA-issued Airline Transport Pilot certificate.
(b) Hold an airplane single- or multi-engine land category and class rating.
(c) Possess instrument currency within the past five years.
(d) Have experience with current electronic flight instrument displays.
(e) Demonstrate adequate visual and auditory acuity through possession of a current FAA medical certificate, or a current driver license with demonstrated responsiveness to visual and aural cues.
(f) Be over 18 years of age.

Signature of Participant: ___________________ Date: ________________

Printed Name of Participant: ________________

## Appendix H: Participant Data Collection Script

**Introduction:** Thank you for consenting to take part in a study for the purpose of determining how pilot situation awareness of stable conditions during approaches can be improved by applying enhancements to the flight instrument displays. This briefing provides the general instructions for participating in the study. There will be three phases: Pre-trial demographic data collection, completion of the study trials, and finally the post-trial usability questionnaire. The expected time to complete the study actions is estimated at 70 minutes. If you are unable to support this amount of time, please let the researcher know now.

**Pre-trial Demographic Collection:** Please take a moment to complete the pre-trial demographic information form… **[Present form]**

**Vignette Orientation:** Now we begin the data collection phase. During the collection, you will be shown a series of video vignettes on this screen. Each will be started and stopped by the researcher and recorded for subsequent data analysis. To allow capture of your comments, a microphone will be worn. The first video will be a calibration vignette to introduce the particulars of the display and the experiment process. After that introduction vignette, you will be presented 12 different scored vignettes that present one of four display enhancements. Each vignette is approximately 150 s duration.

The approach to be displayed is the following… **[Present approach plate]**

- BUCKK fix location at 2,200 ft pressure altitude
- Minimums of 250 ft radio altitude
- Glideslope of 2.75°
- 6.1 nm from BUCKK to RWY 34R
- All briefings, checklists, and configurations complete and verified

The baseline display layout is as shown now by the researcher… **[Present display]**

Your task in this study is two-fold.

First, during the approach you will be presented red and green visual cues in the left and right segment of the primary flight display, respectively… **[Present display]**

These cues are located so as to allow detection within your peripheral vision and not impact your instrument scan. For the first 15-20 s of each vignette, these cues will be inhibited to allow you to settle into the profile. When each cue displays, you are to depress this inline handheld switch in the hand corresponding to the side the cue is displayed. Only a momentary press is needed for successful recording. The researcher will be monitoring your response actions. Should you begin to lapse in your response, the researcher will provide a “Cue Response” reminder.

Second, you will be asked to pay particular attention to primary flight and navigation displays presented and monitor performance and attitude parameters as you would for any other approach. Take a moment to study and memorize the criteria, point of application, and response action on the provided card… **[Present card]**

You will be monitoring the approach for an unstable condition using the following criteria. These criteria apply upon hearing the "one thousand" radio altitude callout. When these values are exceeded, the airplane is considered to be in an unstable approach state.

| Element | Criteria |
|---|---|
| Airspeed | Airspeed not more than $V_{REF}$ +10 kt and not less than $V_{REF}$ |
| Vertical Speed | Vertical speed no greater than 1,000 fpm |
| Roll Angle | Roll angle not in excess of 6° |
| Path | Less than one dot deviation from localizer and glideslope |

| Altitude | Action |
|---|---|
| 1,000-500 ft | Correct airplane path or energy state to achieve a stable condition |
| Below 500 ft | Immediately execute the go-around procedure |

For the study, you'll see the following enhancements to assist in determination of stability… **[Present enhancements]**

For each vignette, you will be asked to think aloud: where you are looking on the display, what you see, and your possible actions. Most critically, should you see an unstable condition, you are asked to voice (1) the presence of an unstable condition, (2) the parameter exceeded, and (3) the action to execute – correct or go around. You are asked to do so as *timely and accurately* as possible whenever observed, no matter how short in duration. Understand that there may be more than one criterion being exceeded, so be sure and speak out for each case observed. In time-critical flight phases, accurate flight parameter perception and quick decision-making and response are necessary. Responses outside five (5) seconds will not be counted, so avoid excessive deliberation and do not delay submission of the response. I'll provide feedback as we go and encourage you to keep voicing your observations and unstable condition determinations.

Here is an example:

> When presented an unstable approach as define by the criteria provided, state aloud the following information as *timely* and *accurately* as possible.
>
> - Condition: "Unstable"
> - Exceedance: "Left of Course," "Airspeed Low," or others
> - Action: "Turn Right," "Increase Airspeed," or others (1,000 - 500 ft AGL)

“Go-around” (below 500 ft AGL)

Let’s begin… **[Execute data collection]**

**Post-trial Survey:** You completed the data collection events, so now you are asked to complete a post-trial usability questionnaire based on your experience and your professional opinion. When desired, written comments may be added in the blocks provided… **[Present form]**

**Conclusion:** Thank you for your participation in this study. Here is your award for participating. Remember that you may contact the researcher at any time using the means provided in the Informed Consent Notice… **[Provide card and incentive]**