DOT/FAA/AM-26/02

Aviation Safety

Office of Aerospace Medicine

Washington, D.C. 20591

Use of a Monocular and Binocular Head-Worn Display in Lieu of a Head-Up Display During Approach, Landing, and Rollout: Human Factors Evaluation of Pilot Performance and Workload

David C. Newton, Ph.D.[1]
Eric T. Greenlee, Ph.D.[2]
Theodore C. Mofle[1]
Farzaneh Shahini, Ph.D.[3]
Rebecca DiDomenica[3]
Inchul Choi, Ph.D.[3]
Ryan Weaver[4]

[1]Federal Aviation Administration
Civil Aerospace Medical Institute
Oklahoma City, OK 73169

[2]Texas Tech University
Department of Psychological Sciences
Lubbock, TX 79409

[3]Cherokee Federal
Oklahoma City, OK 73169

[4]Federal Aviation Administration
Flight Research and Analysis Group
Oklahoma City, OK 73169

January 2026

NOTICE

This publication and all Office of Aerospace Medicine technical reports are available in full text from the Civil Aerospace Medical Institute's publications website and at the National Transportation Library's Repository & Open Science Access Portal.

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-26/02	2. Title & Subtitle Use of a Monocular and Binocular Head-Worn Display in Lieu of a Head-Up Display During Approach, Landing, and Rollout: Human Factors Evaluation of Pilot Performance and Workload
3. Report Date January 2026	4. Performing Organization Code AAM-510
5. Author(s) David C. Newton (0000-0002-3112-3316), Eric T. Greenlee (0000-0001-9433-6434), Theodore C. Mofle (0000-0003-2217-9083), Farzaneh Shahini (0000-0002-1514-5312), Rebecca DiDomenica (0009-0009-4726-578X), Inchul Choi (0000-0002-2048-7026), and Ryan Weaver	6. Performing Org Report Number N/A
7. Performing Organization Name & Address Federal Aviation Administration, Civil Aerospace Medical Institute AAM-510, Oklahoma City, OK 73169	8. Contract or Grant Number N/A
9. Sponsoring Agency Name & Address Office of Aerospace Medicine, Federal Aviation Administration 800 Independence Ave., S.W., Washington, DC 20591	10. Type of Report & Period Covered N/A
11. Supplementary Notes N/A	

12. Abstract

The approach, landing, and rollout is a complex, critical operation for pilots of fixed-wing aircraft, particularly when flight visibility is limited by weather. To enhance the safety of this operation, aircraft can be equipped with a Head-Up Display (HUD), which presents flight symbology on a transparent screen at a focal distance of optical infinity so that the pilot can view primary flight information while maintaining visual contact with the runway. The Head-Worn Display (HWD) is an emerging technology that is designed to provide the benefits of a HUD. However, it may incorporate optical differences that impact pilots' performance and workload. HWDs can be binocular (i.e., displaying symbology to both eyes) or monocular (i.e., displaying symbology to a single eye). When flying with a monocular HWD, binocular rivalry may impact the pilot's ability to use the symbology and impose greater demands on the pilot's attention. This raises questions about whether using a monocular HWD impacts pilots' flying performance, elevates workload, and increases the risk of attentional tunneling. Pilot performance and workload may also be impacted by the physical and optical differences of the HWD relative to those of the HUD. To address these concerns, a study was carried out in which 24 pilot crews, each consisting of two Airline Transport Pilot (ATP) Captains, flew approach and landing scenarios with varying visibility levels, some of which included non-normal events, in a Boeing 737 Level D-equivalent flight simulator while using flight symbology presented on a HUD, binocular HWD, and monocular HWD. Simulator motion was disabled in the study to prevent interference with the HWD head tracking system. Quantitative measures of pilot flying performance were implemented to evaluate the effects of each display type on flightpath and energy management, landing and rollout performance, and response to non-normal events. Pilots rated their workload during each scenario using the National Aeronautics and Space Administration Task Load Index (NASA-TLX). The findings of this study suggest that a monocular HWD may not have a substantial impact on a pilot's ability to manage the flightpath and energy state during approach, landing, and rollout operations. However, pilots experienced a higher workload when flying with the monocular HWD than with the binocular HWD and HUD. There were impacts on landing performance and runway incursion detection attributable to the optical characteristics of the HWD relative to those of the HUD, as well as the monocular versus binocular configuration of the HWD. Ultimately, this research contributes to the understanding of how visual attention is impacted by monocular viewing and provides operational takeaways for the use of an HWD in lieu of a HUD during low-visibility flight operations.

13. Key Words		14. Distribution Statement
Advanced vision systems, binocular rivalry, extended reality, flight technical error, head-mounted display, head-up display, head-worn display, low-visibility operations, pilot performance, pilot response time, pilot workload, runway incursions		N/A
15. Security Classification (of this report) Unclassified	16. Security Classification (of this page) Unclassified	17. No. of Pages 105

Acknowledgments

This research was completed with funding from the Federal Aviation Administration (FAA) Human Factors Division (ANG-C1) in support of the Aviation Safety (AVS) organization. The authors wish to express appreciation to numerous individuals for their contributions throughout the development and execution of this research. From the FAA Flight Standards Low Visibility Operations Unit (AFS-410C): Doug Dixon, Catherine Graham, Janet Greenwood, Matt Harmon, James Marks, and Mike Melssen for their guidance and support as technical sponsors. From the FAA Human Factors Division (ANG-C1): Dr. Victor Quach, Eddie Austrian, and Kevin Siragusa for program management support and helpful feedback on earlier drafts of this report. From the FAA Flight Standards Flight Research and Analysis Group (AFS-430): Brian Baker, Dave Guy, Jeremy Hesselroth, Vincent Liao, Pete Rhode, Doug Rodzon, and Scott Stevenson for their dedication and support during study development and data collection. From the FAA Civil Aerospace Medical Institute Flight Deck Human Factors Research Laboratory (AAM-510): Ashley Awwad, Terry King, and Dr. Daniela Kratchounova for mentorship and support during the initial development of this research. From Cherokee Nation 3S: Tara Stuckey for transcribing all of the simulator audio recordings and Victor Shortt for developing a custom approach procedure chart. From Collins Aerospace: Nate Kowash, Alex Reid, Dean Schwab, Carlo Tiana, and Felix Turcios for supporting the Collaborative Research and Development Agreement that facilitated this research and for helpful feedback on earlier drafts of this report. From Texas Tech University: Drs. HeeSun Choi, Martina Klein, and Keith Jones for serving on the lead author's dissertation committee and for providing valuable guidance and mentorship throughout this research. The authors also express gratitude to the pilot subject matter experts who supported head-worn display acceptance testing and scenario beta testing: Jason Crain, Mark Humphreys, Keith Kankelfitz, Aaron Perkins, Keith Stover, and Andy Reitz. Finally, the authors express the utmost appreciation to the Part 121 flight crews for participating in this study and to Brian Townsend for recruiting and scheduling those flight crews.

Disclaimer

The views expressed herein do not reflect the views of the United States (U.S.) Department of Transportation (DOT), the FAA, or those acknowledged.

Table of Contents

Introduction **1**
Low-Visibility Approach and Landing 1
Use of a Head-Up Display During Low-Visibility Approach and Landing 3
Emerging Head-Worn Display Concepts 6
Binocular Versus Monocular HWDs 7
The Present Study and Research Questions 9
Method **11**
Participants 11
Research Design and Independent Variables 12
IV_1: Display Type 13
IV_2: Runway Visual Range 14
Dependent Variables 16
Instrument Segment Performance 16
Threshold Crossing Performance 17
Landing and Rollout Performance 18
Pilot Flying Workload 19
Pilot Response to Non-Normal Events 19
Supplemental Measures 24
Testing Environment and Flight Scenarios 24
Flight Simulator 24
Head-Up and Head-Worn Displays 25
Simulator Scenarios 28
Procedure 29
Initial Briefing 30
Familiarization Session 31
Sighting Eye Dominance Test 32
Experimental Session 32
Post-Experiment Procedures 34
Results **34**
Frequency of Missed Approaches versus Landings 34
Instrument Segment Performance 35
Root Mean Square Flightpath Deviation 36
Root Mean Square Airspeed Deviation 37
Threshold Crossing Performance 39
Localizer Deviation at Threshold Crossing 39

Glideslope Deviation at Threshold Crossing 41
Absolute Airspeed Deviation at Threshold Crossing 43
Landing and Rollout Performance 45
Lateral Deviation from Runway Centerline at Touchdown 45
Distance from Runway Threshold at Touchdown 47
Sink Rate at Touchdown 49
Deviation from Localizer During Rollout 51
Response to Non-Normal Events 54
Runway Incursion Scenarios 54
Flight Director Failure Scenarios 57
Pilot Flying Workload 61
NASA-TLX Total Weighted Score During Normal Scenarios 61
NASA-TLX Subscale Scores During Normal Scenarios 63
NASA-TLX Total Weighted Score During Runway Incursion Scenarios 65
NASA-TLX Subscale Scores During Runway Incursion Scenarios 66
NASA-TLX Total Weighted Score During Flight Director Failure Scenarios 68
NASA-TLX Subscale Scores During Flight Director Failure Scenarios 70
Usability Questionnaire 71
Subjective Evaluation of Dominant Versus Non-Dominant Eye with Monocular Head-Worn Display 75
Discussion 75
Findings 76
RQ_1: Do pilots demonstrate differences in flightpath tracking, airspeed management, landing, and rollout performance depending on whether they fly with a monocular HWD, binocular HWD, or HUD? 76
RQ_2: Does flight visibility impact pilots' flightpath tracking, airspeed management, landing, and rollout performance when flying with a HUD, binocular HWD, and monocular HWD? 76
RQ_3: Do pilots experience different workload levels during an instrument approach and landing depending on whether they fly with a monocular HWD, binocular HWD, or HUD? 77
RQ_4: Is there a difference in pilots' ability to detect symbology failures or hazards on the runway depending on whether they fly with a monocular HWD, binocular HWD, or HUD? 78
RQ_5: Do the physical and optical differences between the HUD and HWD used in this study impact pilot performance, workload, and non-normal event detection during an instrument approach and landing? 80

RQ_6: Is pilot eye dominance a significant consideration when flying with a monocular HWD? 81
Limitations 81
Conclusions and Future Directions 82
References 84

List of Tables

Table 1 Research Design with Independent Variables and Dependent Variables 13

Table 2 Approximate Schedule of Study Session 30

Table 3 Familiarization Session Scenarios 32

Table 4 Frequency of Landings versus Missed Approaches in Normal Scenarios 35

Table 5 Adjusted Means, Standard Errors, and Confidence Intervals for Root Mean Square Flightpath Deviation with Each Display Type...37

Table 6 Adjusted Means, Standard Errors, and Confidence Intervals for Root Mean Square Airspeed Deviation with Each Display Type 39

Table 7 Adjusted Means, Standard Errors, and Confidence Intervals for Localizer Deviation at Threshold Crossing in Each Experimental Condition...41

Table 8 Adjusted Means, Standard Errors, and Confidence Intervals for Glideslope Deviation at Threshold Crossing in Each Experimental Condition...42

Table 9 Adjusted Means, Standard Errors, and Confidence Intervals for Airspeed Deviation at Threshold Crossing in Each Experimental Condition...44

Table 10 Adjusted Means, Standard Errors, and Confidence Intervals for Lateral Deviation from Runway Centerline at Touchdown in Each Experimental Condition 47

Table 11 Adjusted Means, Standard Errors, and Confidence Intervals for Longitudinal Distance from Runway Threshold at Touchdown in Each Experimental Condition..................................48

Table 12 Adjusted Means, Standard Errors, and Confidence Intervals for Sink Rate at Touchdown in Each Experimental Condition 50

Table 13 Frequency of Landings versus Missed Approaches During Runway Incursion Scenarios for Each Display Type 54

Table 14 Adjusted Means, Standard Errors, and Confidence Intervals for Distance from Runway TDZ when the Pilot Flying Pressed the TO/GA Button in Response to Runway Incursions for Each Display Type 56

Table 15 Frequency of Landings versus Missed Approaches During FD Failure Scenarios for Each Display Type 58

Table 16 Adjusted Means, Standard Errors, and Confidence Intervals for Elapsed Time from the Onset of a FD Failure to when the Pilot Flying Pressed the TO/GA Button for Each Display Type 59

Table 17 Adjusted Means, Standard Errors, and Confidence Intervals for Response to First FD Failure for Each Display Type 61

Table 18 Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Total Weighted Score during Normal Scenarios in Each Experimental Condition 62

Table 19 Results from Post-Hoc ANOVAs 63

Table 20 Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Subscale Scores during Normal Scenarios in Each Display Type Condition 64

Table 21 Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Subscale Scores During Normal Scenarios in Each Runway Visual Range Condition 65

Table 22 Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Total Weighted Score During Runway Incursion Scenarios for Each Display Type 66

Table 23 Results from Post-Hoc ANOVAs 67

Table 24 Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Subscale Score during Runway Incursion Scenarios in Each Experimental Condition 68

Table 25 Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Total Weighted Score During FD Failure Scenarios for Each Display Type 69

Table 26 Results from Post-Hoc ANOVAs 70

Table 27 Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Subscale Scores during Runway Incursion Scenarios in Each Experimental Condition 71

Table 28 Descriptive Statistics for Usability Questionnaire Items 72

Table 29 Highlighted Written Feedback on Participants’ Experiences with the Head-Up Display and Head-Worn Display 74

List of Figures

Figure 1 *The Relationship of Task Performance, Activation Level, and Mental Workload* 3

Figure 2 *Image of Head-Up Display Symbology during an Instrument Landing System Approach* 5

Figure 3 *Illustration of the Perceptual Impacts of Binocular Rivalry* 8

Figure 4 *Simulator Images Showing Runway Visual Range Levels Used in the Study* 15

Figure 5 *Images of Left and Right Runway Incursions Viewed Through the Head-Up Display* 21

Figure 6 *Depiction of Normal FD and Failed FD* 23

Figure 7 *Flight Deck of CAE Boeing 737-800 Simulator* 25

Figure 8 *Collins Aerospace HGS-6700 Head-Up Display in Simulator* 26

Figure 9 *Microsoft HoloLens 2 in Simulator* 27

Figure 10 *Box-and-Whisker Plot of Root Mean Square Flightpath Deviation with Each Display Type* 37

Figure 11 *Box-and-Whisker Plot of Root Mean Square Airspeed Deviation with Each Display Type* 38

Figure 12 Box-and-Whisker Plot of Localizer Deviation at Threshold Crossing in Each Experimental Condition 40

Figure 13 *Box-and-Whisker Plot of Glideslope Deviation at Threshold Crossing in Each Experimental Condition* 42

Figure 14 *Box-and-Whisker Plot of Airspeed Deviation at Threshold Crossing in Each Experimental Condition* 44

Figure 15 *Box-and-Whisker Plot of Lateral Deviation from the Runway Centerline at Touchdown in Each Experimental Condition* 46

Figure 16 *Box-and-Whisker Plot of Longitudinal Distance from the Runway Threshold at Touchdown in Each Experimental Condition* 48

Figure 17 *Box-and-Whisker Plot of Sink Rate at Touchdown in Each Experimental Condition* 50

Figure 18 Box-and-Whisker Plot of Root Mean Square Deviation from Localizer During Rollout 52

Figure 19 *Frequency Bar Plot of Maximum Deviation from Localizer During Rollout* 53

Figure 20 *Box-and-Whisker Plot of Aircraft Distance from Runway TDZ when the Pilot Flying Pressed the TO/GA Button in Response to Runway Incursions for Each Display Type* 55

Figure 21 *Strip Chart of Aircraft Distance from Runway TDZ when the PF Pressed the TO/GA Button in Response to the First Runway Incursion for Each Display Type* 57

Figure 22 *Box-and-Whisker Plot of Elapsed Time from the Onset of a FD Failure to when the PF Pressed the TO/GA Button for Each Display Type*..........59

Figure 23 *Box-and-Whisker Plot of Elapsed Time from the Onset of the First FD Failure to when the Pilot Flying Pressed the TO/GA Button for Each Display Type*..........60

Figure 24 Box-and-Whisker Plot of NASA-TLX Total Weighted Score during Normal Scenarios in Each Experimental Condition..........62

Figure 25 *Box-and-Whisker Plot of NASA-TLX Total Weighted Score During Runway Incursion Scenarios for Each Display Type*..........66

Figure 26 *Box-and-Whisker Plot of NASA-TLX Total Weighted Score During FD Failure Scenarios for Each Display Type*..........69

Figure 27 *Stacked Bar Chart of Usability Questionnaire Responses*..........73

List of Equations

Equation 1 Flightpath Deviation 16

Equation 2 Root Mean Square Flightpath Deviation 17

Equation 3 Airspeed Deviation 17

Equation 4 Root Mean Square Localizer Deviation 19

List of Abbreviations

ANOVA	Analysis of Variance
AGL	Above Ground Level
ATP	Airline Transport Pilot
CAT I	Category I
CAT II	Category II
CAT III	Category III
CLL	Centerline Lighting
DA	Decision Altitude
DH	Decision Height
DV	Dependent Variable
FAA	Federal Aviation Administration
FD	Flight Director
FOR	Field of Regard
FOV	Field of View
FTE	Flight Technical Error
GA	Go-Around
GPIP	Glide Path Intercept Point
HIRL	High-Intensity Runway Lighting
HUD	Head-Up Display
HWD	Head-Worn Display
IAS	Indicated Airspeed
ILS	Instrument Landing System
IMC	Instrument Meteorological Conditions
IQR	Interquartile Range
ILS	Instrument Landing System
IRB	Institutional Review Board
IV	Independent Variable

MALSR	Medium Intensity Approach Lighting System with Runway Alignment Indicator Lights
MANOVA	Multivariate Analysis of Variance
MSL	Mean Sea Level
NASA	National Aeronautics and Space Administration
NASA-TLX	NASA-Task Load Index
NTSB	National Transportation Safety Board
PDX	Portland International Airport
PF	Pilot Flying
PM	Pilot Monitoring
RA	Radio Altimeter
RMS	Root Mean Square
RVR	Runway Visual Range
RWY	Runway
TDZ	Touchdown Zone
TO/GA	Takeoff/Go-Around
VMC	Visual Meteorological Conditions
XR	Extended Reality

Abstract

The approach, landing, and rollout is a complex, critical operation for pilots of fixed-wing aircraft, particularly when flight visibility is limited by weather. To enhance the safety of this operation, aircraft can be equipped with a Head-Up Display (HUD), which presents flight symbology on a transparent screen at a focal distance of optical infinity so that the pilot can view primary flight information while maintaining visual contact with the runway. The Head-Worn Display (HWD) is an emerging technology that is designed to provide the benefits of a HUD. However, it may incorporate optical differences that impact pilots' performance and workload. HWDs can be binocular (i.e., displaying symbology to both eyes) or monocular (i.e., displaying symbology to a single eye). When flying with a monocular HWD, binocular rivalry may impact the pilot's ability to use the symbology and impose greater demands on the pilot's attention. This raises questions about whether using a monocular HWD impacts pilots' flying performance, elevates workload, and increases the risk of attentional tunneling. Pilot performance and workload may also be impacted by the physical and optical differences of the HWD relative to those of the HUD. To address these concerns, a study was carried out in which 24 pilot crews, each consisting of two Airline Transport Pilot (ATP) Captains, flew approach and landing scenarios with varying visibility levels, some of which included non-normal events, in a Boeing 737 Level D-equivalent flight simulator while using flight symbology presented on a HUD, binocular HWD, and monocular HWD. Simulator motion was disabled in the study to prevent interference with the HWD head tracking system. Quantitative measures of pilot flying performance were implemented to evaluate the effects of each display type on flightpath and energy management, landing and rollout performance, and response to non-normal events. Pilots rated their workload during each scenario using the National Aeronautics and Space Administration Task Load Index (NASA-TLX). The findings of this study suggest that a monocular HWD may not have a substantial impact on a pilot's ability to manage the flightpath and energy state during approach, landing, and rollout operations. However, pilots experienced a higher workload when flying with the monocular HWD than with the binocular HWD and HUD. There were impacts on landing performance and runway incursion detection attributable to the optical characteristics of the HWD relative to those of the HUD, as well as the monocular versus binocular configuration of the HWD. Ultimately, this research contributes to the understanding of how visual attention is impacted by monocular viewing and provides operational takeaways for the use of an HWD in lieu of a HUD during low-visibility flight operations.

Introduction

The approach, landing, and rollout is a critical operation for pilots of fixed-wing aircraft. This operational sequence, which involves transitioning from terminal operations to the intended landing runway, is associated with increased pilot mental workload relative to other phases of flight (Rosekind et al., 2000; Wilson & Hankins, 1994). This is particularly the case when weather conditions restrict pilots' ability to see natural visual information about the runway and airfield environment while they fly (Bennett & Schwirzke, 1992). Aviation safety data indicate that, despite the final approach and landing representing approximately 4% of the total flight time of a 1.5-hour flight, 47% (14 of 30) of fatal accidents involving civilian commercial jet aircraft[1] between 2014 and 2024 occurred during one of these two phases of flight (Boeing Aircraft Co., 2025). In addition to impacts on flight safety, weather conditions that restrict flight visibility during critical flight phases, such as approach and landing, are a source of measurable operational inefficiency. According to FAA data, flight cancellations, delays, diversions, and reduced throughput due to weather were responsible for up to a $3.1 billion economic loss to the flying public from 2016 to 2018 (FAA, 2022a). These data demonstrate an opportunity to explore the use of flight deck technology as a potential safety and efficiency enhancement.

Low-Visibility Approach and Landing

When pilots manually fly[2] an instrument approach and landing, they use a combination of flight instruments presented on displays inside the aircraft and visual information about the runway environment beyond the aircraft to keep the aircraft stabilized during the approach, fly the aircraft to the runway touchdown zone (TDZ), and initiate the landing flare sequence to touch down on the runway surface and decelerate during rollout. Weather conditions can restrict visibility of runway visual information, so the pilot must rely solely on the flight instruments, i.e., the instrument segment. During the instrument segment of the approach, the primary task for the Pilot Flying (PF) is to actively control the aircraft to keep it aligned with the intended lateral and vertical flightpath for the instrument approach procedure being flown. The PF must also maintain the target indicated airspeed (IAS) of the approach. To accomplish this, they must maintain awareness of airspeed information presented on the flight instruments and adjust the throttles as necessary (Campbell et al., 2019). When the aircraft descends nearer to the runway environment, visual information such as approach lighting, runway markings, and other traffic becomes available. At this point, the pilot might be able to use visual information about the runway environment to make judgments about the aircraft's position and trajectory relative to the runway TDZ. This visual information includes height information from a synthesis of the horizon line and landing aim point (Lintern & Liu, 1991), depth information communicated by

[1] Multi-engine aircraft with a maximum gross weight greater than 60,000 lbs.

[2] In this context, "manually fly" refers to controlling the flightpath and managing energy of the aircraft without the use of an autoland, autopilot, or autothrottle. Other types of flight deck automation, such as a Flight Director, may be enabled.

texture gradient from terrain and linear perspective from the runway (Lintern, 2000), and optic flow information from ambient features in the runway environment (Gibson, 1986).

Manually flying an Instrument Landing System (ILS) approach and landing is a cognitively demanding activity, as evidenced by pilots reporting higher mental workload levels while flying an approach and landing than during other phases of flight (Roscoe & Ellis, 1990). Mental workload conceptualizes the relationship between the demands of a task and the spare attentional resources available to allocate to the task: When the demands of a task are high, there is less spare attentional capacity available to absorb further increases in task demands, resulting in high workload (Wickens, 2008). Young and Stanton (2005) define mental workload as "the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support, and past experience" (p. 39-1). The relationship between task demands and task performance is not linear: If the demands of a task are too low, there can be a resulting disengagement from the task and poor task performance. Conversely, if task demands are too high, the pilot's attentional resources can be overloaded, and flying performance can suffer. Between these two zones lies an optimum zone where the task demands and mental workload are at a level where the pilot is actively engaged with the operation and is performing it well (Hancock & Ganey, 2003; see Figure 1). On the flight deck, where multiple processes with dynamically changing priorities compete for the pilot's attentional resources, attentional overload can cause pilots to neglect critical aspects of their duties at critical times. Eastern Airlines Flight 401 is a prime example of this, where the flight crewmembers were cognitively consumed with diagnosing a faulty landing gear light and failed to maintain awareness of their airspeed and altitude, leading to a ground collision and the death of 101 passengers and crew members (National Transportation Safety Board [NTSB], 1973).

Figure 1

The Relationship of Task Performance, Activation Level, and Mental Workload

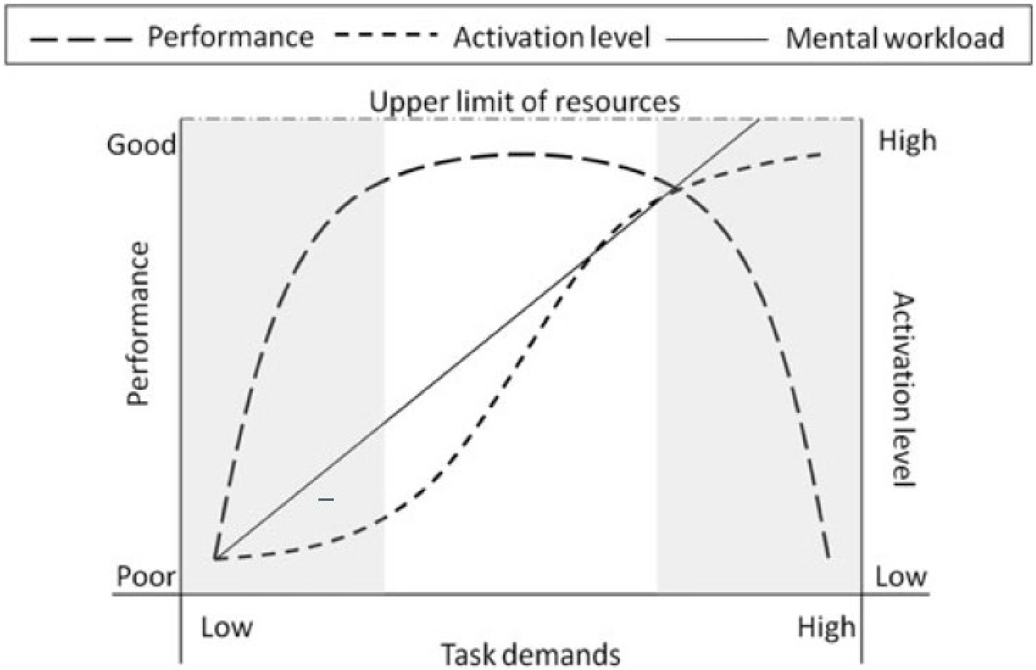


Figure is from "State of Science: Mental Workload in Ergonomics" by M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, 2015, Ergonomics, 50, p. 3. Copyright 2015 by Taylor & Francis.

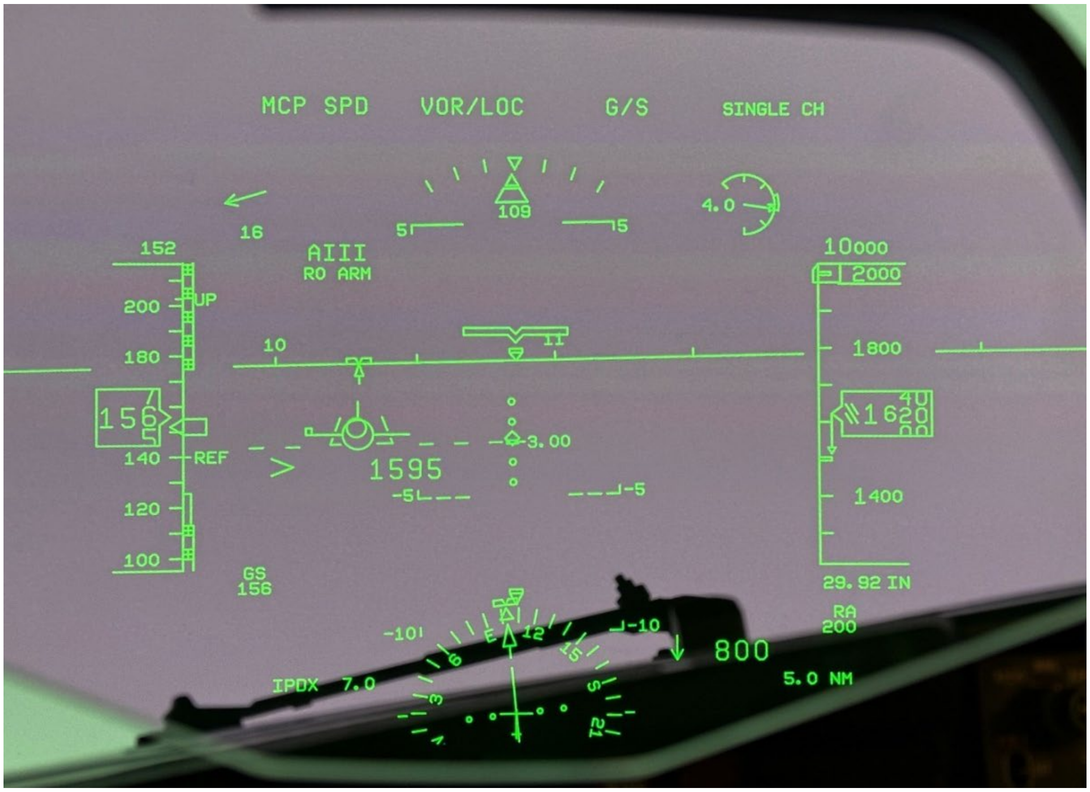
One of the most fundamental aspects of an approach and landing that can influence the pilot's performance and workload while flying is the degree to which weather conditions restrict visibility of the runway. Restricted flight visibility caused by weather phenomena such as overcast clouds and fog reduces the salience of runway visual information and narrows the time window between transitioning to the visual segment of the approach and landing. Consequentially, the PF may exhibit poorer landing performance and experience increased mental workload. Weather conditions that reduce flight visibility increase the risks involved with landing, as well as the likelihood that flights will be cancelled, delayed, or diverted. When low visibility reduces the salience of runway visual information, aerial perspective illusions can also occur, in which the pilot misperceives the height and distance of the aircraft from the runway and makes incorrect compensations to the flightpath. This can cause the pilot to land the aircraft too far ahead of or beyond the runway TDZ, too far left or right of the runway centerline, or initiate the landing flare too early or too late, resulting in a hard landing (FAA, 2011b; Gibb, Schvaneveldt, & Gray, 2008).

Use of a Head-Up Display During Low-Visibility Approach and Landing

To enhance safety when pilots fly in low-visibility conditions, aircraft can be equipped with a Head-Up Display (HUD), which superimposes information from flight instruments onto the pilot's view outside the aircraft using a transparent display (FAA, 2014; see Figure 2). The HUD is a collimated image display, meaning the optics of the display project the symbology at optical infinity (Society of Automotive Engineers [SAE] International, 2008). Collimating the symbology minimizes the degree to which the pilot's eyes must adjust focal length (i.e., accommodation) and vergence angle when switching attention between the HUD symbology and runway visual information (Weintraub & Endsing, 1992). Image collimation also minimizes the vergence-accommodation conflict, which occurs when there is incongruent perceptual information about the distance at which a virtual image appears and the focal distance required to render the image in sharp focus (Coni et al., 2016; Hoffman et al., 2008; Iaveccia et al., 1988). Superimposing flight symbology onto runway visual information converts what was once a sequential, selective attention process—look at only the flight symbology on the head-down displays or look at only runway information out the windscreen—into a parallel process, where pilots can integrate information from both domains by way of object-based mechanisms of visual attention (Kimchi, Yeshurun, & Cohen-Savransky, 2007; Wickens & Long, 1995). The design philosophy of HUD symbology also leverages space-based mechanisms of visual attention, with the assumption that multiple pieces of visual information lying within an attended region can be processed concurrently, increasing cognitive processing efficiency compared to when this information is compartmentalized into two separate locations in space, requiring larger magnitude eye and head movements (Jarmasz et al., 2005; Wickens & Long, 1995).

Figure 2

Image of Head-Up Display Symbology during an Instrument Landing System Approach



This is an example of flight symbology on a HUD during an ILS approach.

Broadly speaking, HUD technology provides critical flightpath and energy state information to the pilot while enabling an outside-the-cockpit view, and an analysis by the Flight Safety Foundation indicates that this capability can positively influence aviation accident outcomes. In an analysis of 983 commercial air carrier, business, and corporate aircraft accidents from 1995 through 2007, the Foundation found that 38% of those accidents may have been prevented if the aircraft had been equipped with a HUD. Moreover, among accidents that occurred during takeoff or landing—the most common category of accident—the Foundation's analysis found that 69% of those accidents might have been prevented if the pilot had been flying with flight symbology presented on a HUD (Flight Safety Foundation, 2009).

The use of flight symbology on a HUD can support improved pilot performance and reduced workload when flying an approach and landing. During the instrument segment of the approach, pilots can more accurately fly on the flightpath of the approach and maintain the target airspeed compared to when conducting this operation using flight instruments on traditional head-down displays (Fischer & Haines, 1980; Weintraub et al., 1984; Wickens & Long, 1995). When transitioning to the visual segment during an instrument approach procedure, pilots can acquire visual information about the runway environment more quickly and fly the approach with increased flightpath accuracy (Boucek et al., 1983; Larish & Wickens, 1991; Wickens & Long, 1995).

During landing, pilots can land closer to the runway centerline (Goteman et al., 2007). When flying in instrument meteorological conditions (IMC), pilots can track the flightpath with nearly the same accuracy as when flying in visual meteorological conditions (VMC; Boucek et al., 1983). Beringer, Domino, and Kamienski (2018) found that pilots report lower levels of mental workload when flying low-visibility instrument approach and landing operations with a HUD compared to when flying those operations without a HUD.

Because of these demonstrated safety enhancements during low-visibility approach and landing operations, the FAA currently authorizes pilots to manually fly an ILS approach and landing with Runway Visual Range (RVR)[3] as low as 600 feet if they fly with flight symbology presented on a HUD, compared to a minimum of 1800 feet for manually flying without a HUD (FAA, 1999, 2018). By enabling operations in these conditions, the widespread use of flight symbology on a HUD increases the throughput of the National Airspace System by reducing weather-related flight delays, cancellations, and diversions (FAA Flight Technologies and Procedures Division, 2022).

Emerging Head-Worn Display Concepts

While there are safety enhancements associated with flying an approach and landing using flight symbology that is superimposed onto the external visual scene, there may be additional benefits and drawbacks associated with the type of display that is used to present the symbology. The HUD possesses design characteristics that may impact pilot performance and workload relative to other types of displays that superimpose symbology onto the pilot's view outside the aircraft. Because the HUD is mounted in a fixed location on the flight deck, the pilot must keep their head stationary and position their eyes within a specific region in space to view the symbology fully and clearly (Velger, 1998). Additionally, HUD hardware typically has a large form factor, so installation is typically limited to larger aircraft with enough space to accept the system, limiting the operational benefits to a small end user population (FAA Flight Technologies and Procedures Division, 2022). To address this, there have been recent efforts to transition toward presenting HUD-like symbology with a Head-Worn Display (HWD) with the goal being to reduce the cost, weight, and size of the display compared to the HUD to potentially expand the operational benefits of the HUD to a larger end user population, including operators of space-restricted aircraft (FAA Flight Technologies and Procedures Division, 2022).

Like the HUD, the HWD enables the pilot to view flight symbology while maintaining visual contact with the runway (Yeh & Wickens, 1997). The primary distinction of the HWD is that it is worn on the pilot's head and moves in accordance with it, whereas the HUD is an installed aircraft system with a fixed position. As such, an HWD may provide the pilot with greater freedom of motion than a HUD, potentially offering an expanded field of view (FOV)[4] and field of

[3] RVR is a measure of flight visibility that is defined as the horizontal distance a pilot can see down the runway, based on sighting either the High-Intensity Runway Lights (HIRLs) or the visual contrast of other targets (FAA, 2011a).

[4] FOV refers to the angular extent of the visual information that is visible through an extended reality (XR) headset at a given moment (Interaction Design Foundation, n.d.).

regard (FOR)[5]. Because a near-to-eye display, such as the HWD, can communicate binocular disparity depth cues, some HWDs may also present certain symbology at a distance nearer to the pilot than optical infinity to minimize visual interference between symbology and the flight deck instrument panel (Thomas, 2009). This may increase the degree to which the pilot's eyes adjust focal length and vergence angle when switching attention between the symbology and the runway environment and negatively impact performance during tasks that require switching attention between those two information sources (Arefin et al., 2022; Condino et al., 2019). In addition, the near-to-eye display of an HWD may restrict the pilot's view of other information sources on the flight deck, or the added weight from wearing the display on their head may change the physical demands involved with the approach and landing operation. Previous FAA research on HWD use by general aviation pilots indicates that an HWD may serve as an effective substitute for a HUD from a pilot performance and usability standpoint and can be used to present information to rotorcraft pilots that enhances their awareness of terrain, obstacles, and other safety-related information (Beringer, 2020; Beringer & Drechsler, 2013). Based on this, a comparison between the HUD and HWD is warranted regarding impacts on pilot performance and workload during a low-visibility approach and landing in commercial jet aircraft.

Binocular Versus Monocular HWDs

The contribution of an HWD to pilot performance, workload, and flight safety may further depend on whether the system is configured to present the symbology to both eyes using independent image sources (i.e., binocular) or to one eye using a single image source (i.e., monocular). With binocular HWD systems, independent image sources generate two monocular images—one for each eye. Those monocular images are tailored to each eye, with precise locations, FOVs, and virtual distances, so that the pilot sees a fused, binocular image. To minimize vergence-accommodation conflict, modern HWD systems typically align the images so that the vergence angle of each eye matches the focal distance needed to render the image in sharp focus (SAE International, 2023). This configuration enables the pilot to view the symbology as a single, binocular image. However, this type of configuration has more complicated optics and calibration requirements than a monocular HWD system. With monocular HWD systems, the image is projected to a transparent combiner positioned over either the left or right eye, while the other eye does not receive the image. The motivations for implementing a monocular HWD are that the weight of the headset can be reduced, the alignment and calibration processes can be simpler, and the hardware may be sold at a lower price point compared to a binocular HWD (FAA Flight Technologies and Procedures Division, 2022).

Binocular Rivalry with Monocular HWDs

When viewing flight symbology on a monocular HWD, dissimilar images are visible to each eye. In this situation, the visual system is unable to fuse the separate retinal images from each eye into a stable, binocular image, so the separate retinal images compete with one another for

[5] FOR refers to the range of a virtual environment presented by an XR headset that can be viewed with physical head and body rotation (Ragan et al., 2015).

perceptual dominance. This phenomenon, known as *binocular rivalry*, causes observers to experience a temporal alternation between the stimuli presented to each eye (Blake, 2001; Patterson et al., 2006; see Figure 3).

Figure 3

Illustration of the Perceptual Impacts of Binocular Rivalry

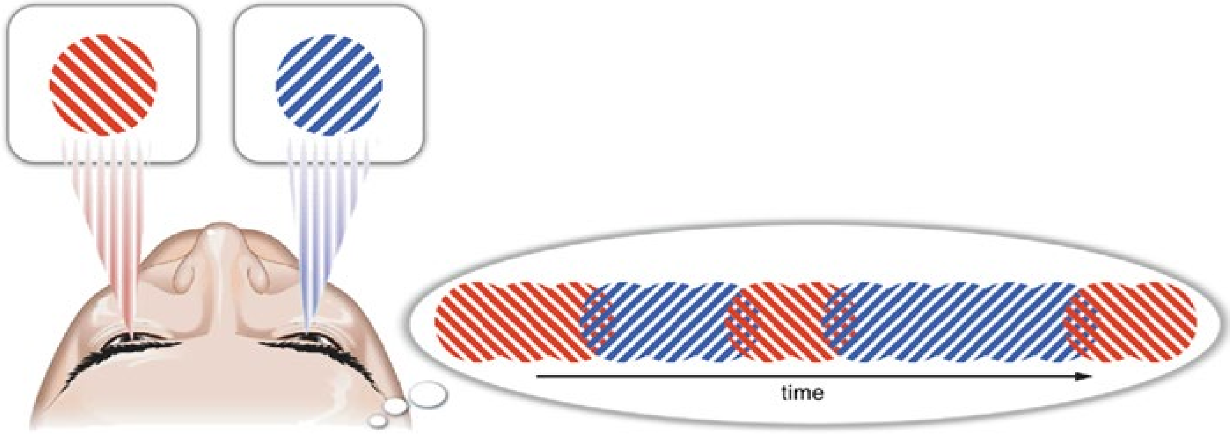


Figure is from adapted from "Understanding attentional modulation of binocular rivalry: A framework based on biased competition" by K.C. Dieter and D. Tadin, 2011, Frontiers in Human Neuroscience, 5, 155-167. Copyright 2011 by Dieter and Tadin.

Basic laboratory research demonstrates that task performance suffers when information used to complete the task is presented to a single eye. When visual information is suppressed during binocular rivalry, there is a general loss of perceptual sensitivity to that information (Blake, 2001). This loss of sensitivity prolongs reaction time to basic stimuli (Fox & Check, 1968; O'Shea, 1987). It also reduces the observer's sensitivity to changes in the shape and orientation of visual stimuli (Blake & Fox, 1974; Fox & Check, 1968; Walker, 1975). Ultimately, these effects degrade the ability to voluntarily direct attention toward task-relevant visual information that has been suppressed due to rivalry (Patterson et al., 2006; Schall et al., 1993; Winterbottom et al., 2006a). Winterbottom et al. (2006b) found that these results may translate to performance impairments in pilots who use a monocular HWD during flight. When performing a simulated in-flight target detection and tracking task with an HWD, participants were less able to detect targets presented with a monocular HWD than with a binocular HWD. These findings suggest that pilot performance may be poorer when using a monocular HWD during a low-visibility approach and landing compared to when using a binocular HWD. As a result, during approach and landing, pilots may be less able to continuously track flightpath guidance and airspeed management symbology.

Eye Dominance with Monocular HWDs

Past research suggests that there is a tendency for visual perception to be biased, such that the overall viewing experience is driven more by one eye than the other—a phenomenon known as eye dominance (Patterson et al., 2007). Individual differences in eye dominance may lead to differential impacts on a pilot's ability to use flight symbology presented on a monocular HWD. Eye dominance is commonly established using a test for sighting eye dominance, such as the

Hole-In-The-Card Test, which determines the preferred eye for monocular sighting tasks (e.g., aligning the sights on a firearm or looking through a microscope; Johannson et al., 2015). Bayle et al. (2020) found that sighting eye dominance testing outcomes are correlated with testing outcomes for near vision acuity and motion coherence, suggesting a unifactorial aspect to eye dominance. Aggregate data from tests of sighting eye dominance indicate that approximately two-thirds of the population is right-eye dominant (Aswathappa, Kutty, & Annamalai, 2011; Eser et al., 2008). The impact of sighting eye dominance on visual perception, performance, and comfort during monocular viewing was a prominent area of concern when the US Army implemented the IHADSS monocular HWD in the AH-64 Apache helicopter, which presented imagery to the pilot's right eye (Rash, 1998). To address this concern, Hiatt et al. (2004) investigated the visual experiences of pilots who were experienced with the IHADSS, including an evaluation of whether there were differences in pilot-reported visual discomfort, unintentional alternations (i.e., binocular rivalry), and visual illusions as a function of sighting eye dominance. While pilots did frequently report these experiences, there was no marked difference in experience as a function of the pilot's sighting eye dominance.

In the empirical literature, there is conflicting evidence about whether sighting eye dominance impacts visual task performance and attention during monocular viewing conditions. On one hand, previous research corroborates the findings of Hiatt et al. (2004). Furthermore, Bayle et al. (2020) evaluated target detection, identification, and tracking performance, as well as subjective comfort, during visual tasks involving monocular viewing on the dominant and non-dominant sighting eye. They found no differences in any of these outcomes as a function of eye dominance. On the other hand, Shneor and Hochstein (2005) evaluated performance on a visual feature search task with long stimulus duration, comparing outcomes between dominant and non-dominant-eye presentation, and found significantly better performance when the target was presented to the dominant sighting eye. Given the competing empirical evidence and the fact that it was derived from basic laboratory tasks, it is important to investigate the role of eye dominance during cognitively demanding flight operations where pilots fly with a monocular HWD, with a focus on any workload impacts, as well as subjective pilot feedback on other aspects of eye dominance (e.g., near and far acuity) that may drive their experiences while flying with a monocular HWD.

The Present Study and Research Questions

The low-visibility approach and landing is an operationally complex procedure that requires the pilot to use numerous types of information from aircraft instruments and the runway environment, and optimally divide their attention among those information sources to safely fly. The measurable benefits of a HUD to flightpath tracking accuracy bring safety enhancements during an approach and landing; however, there are also known pitfalls involved with HUD use, such as the risk of attentional tunneling. These benefits and drawbacks, which are demonstrated from the use of a HUD to present symbology, may not transfer if other display types are used. A binocular HWD, while presenting information as a fused, binocular image just as the HUD does, may impact the pilot's ability to distribute attention between the symbology and the runway environment if it presents the symbology at a closer distance to the pilot than the collimated symbology presented on a HUD (Laramee & Ware, 2002).

When using flight symbology on a monocular HWD, pilots will experience binocular rivalry, which may impact the ability of the PF to maintain lateral and vertical guidance, maintain an airspeed target, and land the aircraft. The PF may also experience elevated workload when flying with a monocular HWD due to the increased attentional demands of monocular visual information (Patterson et al., 2007). Binocular rivalry when flying with a monocular HWD may also compromise the ability of the PF to detect abnormalities in the flight symbology or outside the aircraft, such as symbology failures or runway incursions, respectively. Additionally, eye dominance of the PF may play a role when flying with a monocular HWD. Based on these concerns, the following research questions were identified:

1. Do pilots demonstrate differences in flightpath tracking, airspeed management, landing, and rollout performance depending on whether they fly with a monocular HWD, binocular HWD, or HUD?
2. Does flight visibility impact pilots' flightpath tracking, airspeed management, landing, and rollout performance when flying with a HUD, binocular HWD, and monocular HWD?
3. Do pilots experience different workload levels during an instrument approach and landing depending on whether they fly with a monocular HWD, binocular HWD, or HUD?
4. Is there a difference in pilots' ability to detect symbology failures or hazards on the runway depending on whether they fly with a monocular HWD, binocular HWD, or HUD?
5. Do the physical and optical differences between the HUD and HWD used in this study impact pilot performance, workload, and non-normal event detection during an instrument approach and landing?
6. Is pilot eye dominance a significant consideration when flying with a monocular HWD?

To evaluate these research questions, a study was carried out in which Airline Transport Pilot flight crews manually flew low-visibility ILS approach and landing scenarios in a Boeing 737 Level D-equivalent flight simulator. The PF flew these scenarios with a HUD, binocular HWD, and monocular HWD from the left seat, while the Pilot Monitoring (PM) sat in the right seat and monitored without a HUD or HWD. The HUD in this study was a production-quality, collimated display that presented symbology at a focal distance of optical infinity, whereas the HWD was a Microsoft HoloLens 2 that could be configured as a monocular or binocular display and presented the symbology at a focal distance of six feet ahead of the PF. This latter characteristic is important to highlight because display collimation is a major design difference between the two devices, and it was not possible to control for this factor between HUD and HWD conditions. Because display collimation is a foundational characteristic of certified HUDs, the use of a non-collimated HWD may limit the ability to generalize certain findings from this research to all HWD systems developed for use on the flight deck.

The approach and landing scenarios featured varying levels of flight visibility to evaluate whether there are any differences in pilot performance and workload across the display types that depend on the natural visual references available to the PF. All scenarios featured significant gusting winds with variable direction, which made flightpath and energy management challenging and elevated pilot workload. PF performance was evaluated in terms of flightpath tracking performance, airspeed management performance, deviation from the runway centerline at landing, and the aircraft's sink rate at landing. After each flight scenario, the workload of the PF was measured using the NASA-Task Load Index (TLX; Hart & Staveland, 1988). These data were used to evaluate PF performance and workload as a function of display type and flight visibility.

Some scenarios included non-normal events designed to make landing unsafe and require a missed approach after detection of the event. The first type of non-normal event was a runway incursion, which served to evaluate whether the type of display in use impacts how effectively the pilot can scan the runway environment to determine whether a landing is safe. The second type of non-normal event involved a failure of the Flight Director (FD) in the HUD and HWD symbology during a CAT III approach, where the pilot must use flight guidance on the HUD or HWD to touchdown. When the FD failed, the pilot no longer had the flight guidance needed to continue the approach safely, so a missed approach was required. The role of pilot eye dominance when flying with a monocular HWD was determined by carrying out subjective testing during familiarization scenarios, where pilots completed approaches and landings with the monocular HWD on the left and right eye and provided feedback on whether they experienced differences in performance, workload, and usability as a function of left or right eye configuration of the monocular HWD.

Method

Throughout the development and execution of this research, the authors worked with multiple Boeing 737 Type-Rated Pilot subject matter experts with HUD and HWD experience, as well as Engineer SMEs with experience in the design and implementation of HUD and HWD technology in transport category aircraft to ensure that the simulated flight scenarios were ecologically valid, and the Dependent Variables (DVs) appropriately represent real-world pilot performance and workload outcomes. A pilot study was conducted prior to executing the main study as a design review to ensure that data were reliably collected and processed, that the study procedure was suitable, and that the simulator scenarios were optimally designed. A combination of quantitative data and qualitative feedback from the pilot study participants was collected and used to make these determinations.

Participants

Forty-eight current ATP Captains participated in the study. The average age of the participants, all of whom were male, was 56.67 years (SD_{age} = 7.24 years). For the study, participants were paired into two-person flight crews, with each crew consisting of Captains from the same operator (i.e., airline) and matched based on availability. Each participant (a) possessed a Boeing 737 Type Rating; (b) was qualified and current according to their operator and FAA requirements to fly a low-visibility approach and landing with a HUD; (c) had experience using a HUD within 30 days of study participation; and (d) had at least 100 flight hours of HUD use. While a Captain-Captain pairing represents a departure from the typical ATP crew that is commonly comprised of a Captain-First Officer pairing, it enabled a regular rotation of seat and PF/PM roles during the study to mitigate participant fatigue and increase data collection efficiency, and also mitigated any differences in training and experience that could introduce underlying variability in PF performance and workload. All participants received monetary compensation for their participation in the study and were reimbursed for travel expenses.

All participants completed the hole-in-the-card eye dominance test during the study session, so the monocular HWD could be configured to present symbology to each participant's dominant sighting eye. Forty-two of the 48 participants (87.5%) were right-eye dominant; the remaining six participants (12.5%) were left-eye dominant. The participants reported an average of 19,507.55 total flight hours (*SD* = 17,642.24 hours), with an average of 94.72 of those hours (*SD* = 136.67 hours) occurring within one month of participating in the study. Seven (14.6%) of the participants indicated they had previous experience using an HWD system.

The reported total number of approaches and landings that the participants had flown while using a HUD ranged from 30 to 5400 (*M* = 731.48; *SD* = 1170.94). On average, 47.85 of those approaches and landings (*SD* = 151.62) had been conducted below CAT I minimums. In the 30 days prior to participation, the average number of approaches and landings that the participants reported having flown while using a HUD was 8.73 (*SD* = 9.50). Thirty (62.5%) of the participants indicated that their respective operators require the use of a HUD on approaches and/or landings beyond the minimum FAA requirements. Participants indicated the frequency with which they deploy the HUD during an approach and landing on a 10-point scale (1 = *use HUD as little as possible*, 10 = *always have HUD deployed*), with an average response of 8.29 (*SD* =2.78).

Research Design and Independent Variables

This study was designed to evaluate whether PF performance and workload, as well as the ability to respond to non-normal events as PF, differ depending on whether flight symbology is presented on a HUD, a binocular HWD, or a monocular HWD. Additionally, this study evaluated whether reducing the visibility of the runway environment (i.e., RVR and ceiling) produces differential impacts on PF performance and workload among these display types. Toward that end, each participant flew 15 instrument landing system (ILS) approach and landing scenarios using three types of displays that superimpose flight symbology, and across three degrees of RVR and ceiling (see Table 1). Nine of these scenarios involved normal operations, where the participants completed a routine approach and landing. The remaining six scenarios incorporated a non-normal event that made landing unsafe, requiring the participant to detect that event and respond by executing a missed approach.

Table 1

Research Design with Independent Variables and Dependent Variables

<table>
<tr><th></th><th colspan="3">IV_2: Runway Visual Range</th></tr>
<tr><th>IV_1: Display Type</th><th>4800 ft</th><th>1200 ft</th><th>600 ft</th></tr>
<tr><td>Head-Up Display</td><td colspan="3" rowspan="3">Dependent Variables
Instrument Segment Performance
• DV_1: Flightpath Deviation during Instrument Segment
• DV_2: Airspeed Deviation during Instrument Segment
Threshold Crossing Performance
• DV_3: Flightpath Deviation at Threshold Crossing
• DV_4: Airspeed Deviation at Threshold Crossing
Landing and Rollout Performance
• DV_5: Deviation from Runway Centerline at Touchdown
• DV_6: Distance from Runway Threshold at Touchdown
• DV_7: Sink Rate at Touchdown
• DV_8: Root Mean Square Deviation from Localizer during Rollout
PF Workload
• DV_9: NASA-TLX Total Weighted Score
• DV_{10}: NASA-TLX Subscale Scores
Response to Non-Normal Events
• DV_{11}: Response to Runway Incursions
• DV_{12}: Response to FD Failures
Supplemental Measures
• Usability Questionnaire and Open-Ended Feedback
• Subjective Assessment of Eye Dominance with Monocular HWD</td></tr>
<tr><td>Binocular Head-Worn Display</td></tr>
<tr><td>Monocular Head-Worn Display</td></tr>
</table>

IV_1: Display Type

Pilot performance and workload were evaluated across three types of displays that superimpose flight symbology: HUD, binocular HWD, and monocular HWD. The HUD and binocular HWD presented the symbology to both eyes. The monocular HWD presented the symbology to the left or right eye only. As stated elsewhere in this report, the HUD in this study was a production-quality, collimated display that presented symbology at a focal distance of optical infinity, whereas the HWD was a Microsoft HoloLens 2 that could be configured as a monocular or binocular display and presented the symbology at a focal distance of six feet ahead of the PF. Display collimation is a major design difference between the two devices, and it was not possible to control for this factor between HUD and HWD conditions. Because display collimation is a foundational characteristic of flight deck HUDs and HWDs, caution should be exercised when generalizing the findings to collimated HWD systems developed for use on the flight deck.

During the pre-experiment training and familiarization, participants used the monocular HWD on both the left and right eye while subjectively assessing whether there was a difference between flying with the monocular HWD on the left or right eye. Before beginning the experimental session, participants' sighting eye dominance was tested using the Distance Hole-In-The-Card test. Forty-two of the 48 participants (87.5%) were right-eye dominant; the remaining six participants (12.5%) were left-eye dominant. For the remainder of the study, the monocular HWD was configured to present symbology to the dominant eye.

IV_2: Runway Visual Range

In this study, pilot performance and workload were evaluated across three flight visibility conditions by implementing three different RVR and cloud ceiling values across the scenarios. The FAA defines RVR as the horizontal distance a pilot can see down the runway, based on sighting either the High-Intensity Runway Lights (HIRLs) or the visual contrast of other targets (FAA, 2011a). Three degrees of opacity of a ground-level fog layer and overcast cloud ceiling height were employed to create three corresponding levels of RVR and ceiling, respectively. These three degrees of RVR and ceiling are described in the following sections. Figure 4 presents depictions of these three conditions taken from the simulator.

4800 ft RVR

In the 4800 ft RVR condition, the cloud ceiling was set at 250 feet above the runway TDZ, and the Decision Altitude (DA) was 231 ft mean sea level (MSL) altitude, which corresponded with 200 ft Radio Altimeter (RA) altitude. This RVR level was chosen for this study as a baseline condition to evaluate HUD and HWD use in Standard Category (CAT) I ILS approach, landing, and rollout operations, where FAA regulations currently do not require the use of a HUD (FAA, 2018). Additionally, 4800 ft RVR was selected to provide adequate visibility during the runway incursion scenarios for the participant to detect the incursion and initiate a missed approach.

1200 ft RVR

In the 1200 ft RVR condition, the cloud ceiling was set at 150 feet above the runway TDZ, and the Decision Height (DH) was 100 ft RA altitude. 1200 ft RVR scenarios were carried out using a Standard CAT II approach procedure, where FAA regulations currently require the use of flight guidance on a HUD to the DA, at which point the flight crew transitions to visual flight references (FAA, 2018). FAA guidance states that the benefit of the HUD during a CAT II approach is that it reduces the time needed to acquire runway visual information when approaching the DA (FAA, 2018).

600 ft RVR

In the 600 ft RVR condition, the cloud ceiling was set at 100 feet above the runway TDZ, and the DH was 50 ft RA altitude. This condition represents the lowest RVR in which pilots are authorized to manually fly an approach and landing using flight guidance on a HUD in the National Airspace System. For manual flight operations at this visibility level, FAA regulations currently require the use of flightpath guidance and flare guidance to touchdown on a HUD or equivalent display (FAA, 1999, 2018). Such operations do not employ a visual segment, as the natural visual cues are not considered adequate to support an unaided landing.

Figure 4
Simulator Images Showing Runway Visual Range Levels Used in the Study

Dependent Variables

The DVs in this study included a combination of objective and subjective measures that evaluate the impacts of display type and RVR on pilot performance and workload, as well as the impact of display type on response to non-normal events. The objective measures included pilot performance during the instrument segment of flight (i.e., from the start of the approach until 300 ft above ground level (AGL) as measured by flightpath and airspeed deviation; flightpath and airspeed deviation at runway threshold crossing; and landing performance as measured by deviation from the runway centerline and distance from runway threshold at touchdown. The NASA-TLX was used to collect participants' workload ratings after each scenario they completed as the PF (Hart & Staveland, 1988). These DVs are described in detail herein.

Instrument Segment Performance

Root Mean Square Flightpath Deviation

During the instrument segment of the approach, the pilot's primary task is to keep the aircraft aligned with the ILS vertical and lateral guidance using the flightpath guidance symbology on the HUD or HWD. This research evaluated whether the display type used affects the ability of the PF to follow this guidance. From the start of the approach to 300 ft AGL, flightpath deviation in feet was measured relative to the center of the ILS localizer (lateral guidance) and ILS glideslope (vertical guidance) signals. The separate deviation values for lateral and vertical guidance were combined to create a single deviation value, calculated as the hypotenuse of the deviation from the localizer and glideslope signals (see Equation 1). Before conducting inferential analyses, flightpath deviation was transformed into root mean square flightpath deviation (Flightpath $\text{Deviation}_{\text{RMS}}$; see Equation 2).

Equation 1
Flightpath Deviation

$$d = \sqrt{Localizer_{Dev}^2 + Glideslope_{Dev}^2}$$

Where:

$Localizer_{Dev}$ = deviation from the center of the localizer signal
$Glideslope_{Dev}$ = deviation from the center of the glideslope signal

Equation 2
Root Mean Square Flightpath Deviation

$$Flightpath\ Deviation_{RMS} = \sqrt{\frac{\sum_{i=1}^{n}(d_i)^2}{n}}$$

Where:

d = hypotenuse of instantaneous vertical and lateral deviation in feet from the center of the flightpath

n = number of data points capturing instantaneous flightpath deviation

Root Mean Square Airspeed Deviation

Along with following the flightpath guidance during the approach, the PF must also maintain awareness of the aircraft's current airspeed and acceleration and adjust the throttles to maintain the target airspeed using the speed deviation tape, acceleration cue, and speed tape in the HUD or HWD symbology. The goal is to maintain the target airspeed from the start of the approach until the throttles are retarded to idle for landing. Because this task is performed in tandem with maintaining the flightpath defined by the approach, the pilot must effectively divide attention between the two information sources. This research evaluated pilots' ability to use the airspeed management symbology to maintain the target airspeed as a function of display type. Specifically, airspeed deviation was evaluated in terms of actual IAS relative to target IAS (i.e., 152 knots) from the start of the approach to 300 ft AGL. Prior to analysis, raw Airspeed Deviation data were transformed into Root Mean Square Airspeed Deviation (Airspeed $Deviation_{RMS}$; see Equation 3).

Equation 3
Airspeed Deviation

$$Airspeed\ Deviation_{RMS} = \sqrt{\frac{\sum_{i=1}^{n}(d_i)^2}{n}}$$

Where:

d = instantaneous deviation between the IAS and the target airspeed in knots

n = number of data points capturing the IAS

Threshold Crossing Performance

Localizer Deviation at Threshold Crossing

The degree of lateral flightpath deviation when the aircraft crosses the runway threshold was evaluated by measuring the distance, in feet, between the ILS localizer antenna on the aircraft and the center of the ILS localizer signal when the aircraft's center of gravity crossed over the runway threshold.

Glideslope Deviation at Threshold Crossing

The degree of vertical flightpath deviation when the aircraft crosses the runway threshold was evaluated by measuring the distance, in feet, between the ILS glideslope antenna on the aircraft and the center of the ILS glideslope signal when the aircraft's center of gravity crossed over the runway threshold.

Airspeed Deviation at Threshold Crossing

The degree of airspeed deviation when the aircraft crossed the runway threshold was evaluated by measuring the difference, in knots, between the observed IAS when the aircraft's center of gravity crossed over the runway threshold and the target IAS of the approach (i.e., 152 knots IAS).

Landing and Rollout Performance

Lateral Deviation from Runway Centerline at Touchdown

When landing, the goal of the PF is to land the aircraft on the runway centerline and within the runway TDZ. The functionality of the flare guidance in the flight symbology also guides the pilot to the runway centerline during landing. As such, pilots' ability to land the aircraft on the center of the runway was evaluated by measuring deviation in feet from the runway centerline at touchdown.

Longitudinal Distance from Runway Threshold at Touchdown

When landing, the goal of the PF is to land within the TDZ by following the flare guidance and natural visual cues to the touchdown aiming point on the runway, which was located 996 ft beyond the runway threshold in each scenario. Before touchdown, the PF must initiate the flare sequence by following the flare guidance on the HUD or HWD, along with judging the aircraft's trajectory and height using natural visual cues, to reduce the aircraft's descent rate before touchdown. As such, pilots' ability to land the aircraft within the TDZ using flare guidance on the HUD and HWD in tandem with natural visual cues from the runway environment was evaluated by measuring the longitudinal distance, in feet, from the runway threshold at touchdown.

Sink Rate at Touchdown

Immediately prior to landing, the pilot's goal is to manage sink rate and avoid a hard landing. The HUD and HWD CAT III flare guidance (i.e., AIII mode) also communicates when to execute the flare to reduce the sink rate prior to touchdown. As such, pilots' ability to manage sink rate, initiating the landing flare at the appropriate time, was evaluated by measuring the aircraft's vertical velocity in ft/second immediately prior to touchdown, with a value closer to zero indicating that the pilot was better able to manage the sink rate of the aircraft to avoid a hard landing.

Root Mean Square Deviation from Localizer During Rollout

The ability of the PF to track the runway centerline during rollout and HUD or HWD CAT III rollout guidance was evaluated by measuring localizer deviation, in feet, from touchdown to 25 knots ground speed. Before conducting inferential analyses, raw deviation data were transformed into root mean square deviation (see Equation 4).

Equation 4
Root Mean Square Localizer Deviation

$$Localizer\ Deviation_{RMS} = \sqrt{\frac{\sum_{i=1}^{n}(d_i)^2}{n}}$$

Where:

d = instantaneous lateral deviation from the localizer guidance

n = number of data points capturing localizer deviation

Pilot Flying Workload

This research evaluated pilot workload of the PF to determine whether task demands are impacted as a function of display type and RVR using the NASA-TLX (Hart & Staveland, 1988). The NASA-TLX assesses the degree of mental, temporal, and physical demands of the task, as well as the effort, performance, and frustration during each flight scenario. The participant acting as PF completed the NASA-TLX at the end of each scenario (participants did not complete the NASA-TLX when they acted as PM). After their final scenario as PF, each participant completed pairwise comparisons of NASA-TLX factors to evaluate the dimensions of workload that contributed the most to their workload, specifically when they were the PF. Responses to the pairwise comparisons were used to calculate the total weighted workload score for each scenario based on the procedure described in Hart and Staveland (1988).

Pilot Response to Non-Normal Events

This research evaluated the ability of the PF to detect and respond to non-normal events during the scenario to determine whether pilot attention distribution between symbology and runway visual information differs as a function of display type. The non-normal events in this study make landing unsafe, requiring the PF to execute a missed approach, which is a formalized procedure that is followed when a landing cannot be completed safely (FAA, 2017). In a Boeing 737, the missed approach procedure involves first pressing the takeoff/go-around (TO/GA) switch on the throttle lever, which activates go-around (GA) mode. The pilot then pitches the aircraft upward to climb to a higher altitude (Boeing Aircraft Co., 2023). Study participants experienced two types of non-normal events during the experimental session. The participants were instructed in the pre-experiment briefing that some of the scenarios in the study would include non-normal events; however, the specific nature of those events was withheld to minimize the expectancy of these events. The two types of non-normal events in this study are described below.

Runway Incursion

In the runway incursion scenarios, a stationary Boeing 767 was located on the taxiway, positioned so that it was partially extending out onto the runway surface (see Figure 5). This type of non-normal event evaluated the ability of the PF to switch attention from HUD or HWD flight symbology to runway visual information before the DA to determine the safety of the runway. Response to the runway incursion was assessed as the distance from the runway TDZ at which the PF pressed the TO/GA button, where pressing the TO/GA button at a farther distance from the runway TDZ indicated a faster response to the runway incursion. Each participant experienced this non-normal event three times—once for each display type—during an approach with 4800 ft RVR. This RVR was chosen for the runway incursion scenarios after evaluating a variety of RVRs during beta testing of runway incursion scenarios with Pilot SMEs. 4800 ft RVR provided an adequate response for the PF to reliably detect the incursion; if the RVR was lower, the incursion would not appear out of the fog until the aircraft was flying over it. Across all participants, half of the incursions were positioned on the left side of the runway, and the other half were positioned on the right side of the runway. Additionally, because participants were paired into two-person flight crews in the study session and the PM was able to see the first incursion for the PF, the positioning of the incursions was counterbalanced so that the first incursion for the first participant to act as PF was always on the opposite side of the runway from the first incursion for the second participant to act as PF. The sequence of runway incursion scenarios across each display type was counterbalanced using the Balanced Latin Square method.

Figure 5

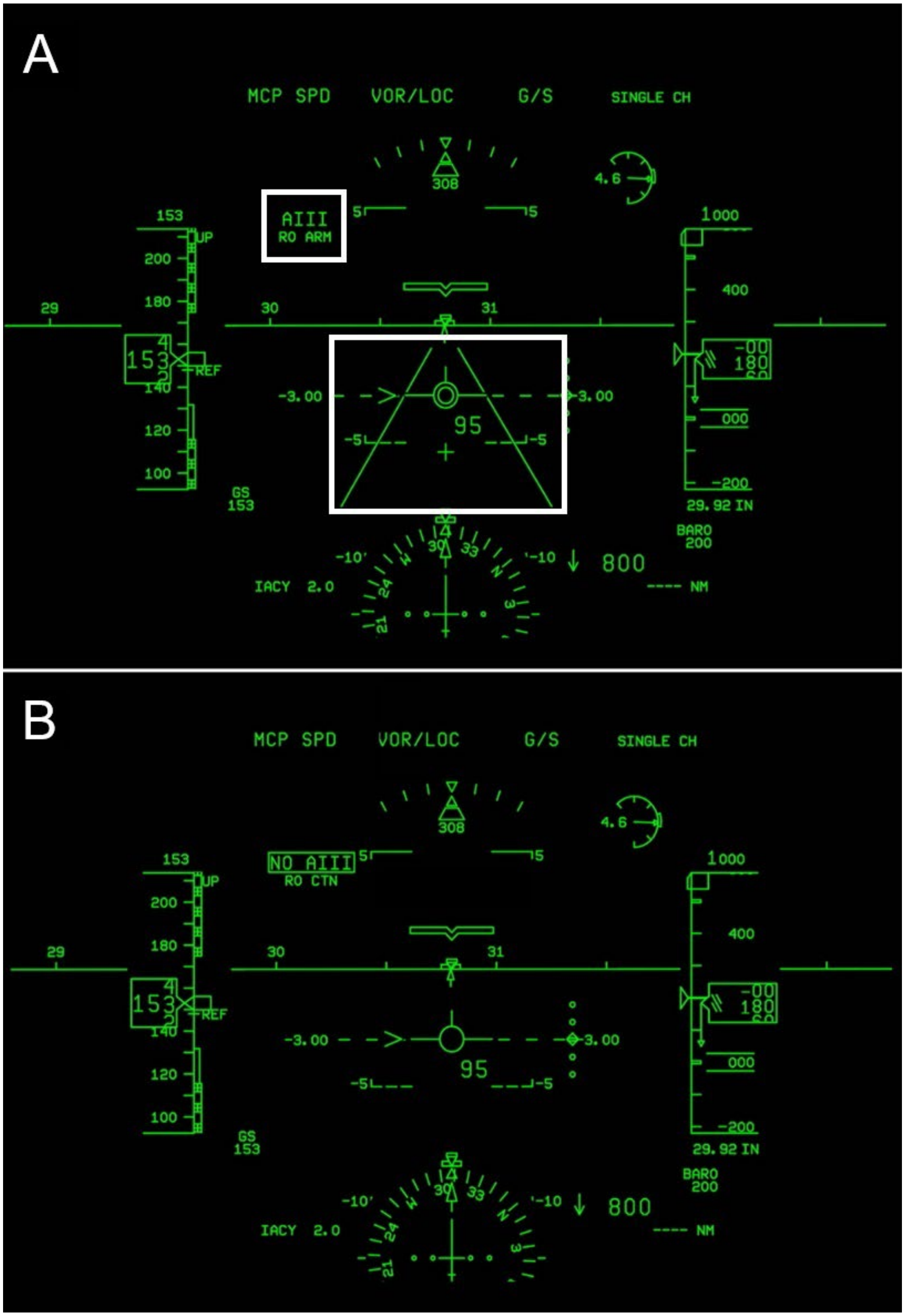
Images of Left and Right Runway Incursions Viewed Through the Head-Up Display

Flight Director Failure

The second type of non-normal event was a failure of the FD in the flight symbology during the approach. When this occurs, the FD disappears from the symbology when the aircraft descends to 100 ft AGL, resulting in a loss of flightpath guidance information (see Figure 6). The "APPROACH WARN" annunciation was suppressed for all FD failures to reduce the salience of this event to the PF. This type of non-normal event evaluated the ability of the PF to detect and respond to failures occurring in the flight symbology as a function of display type. Response to the FD failure was assessed as response time, in seconds, measured from the point when the simulator triggers the FD failure to the point when the PF presses the TO/GA button. Each participant experienced this non-normal event three times—once for each display type—during an approach with 600 ft RVR. Scenario beta testing with SME pilots led to the decision to employ a CAT III procedure with 600 ft RVR for these scenarios. FAA guidance for low-visibility flight operations communicates that a missed approach must be executed if the flight crew experiences a failure of flightpath guidance during a manually flown CAT III approach with 600 ft RVR (FAA, 1999, 2018). Because a missed approach was the desired response to this scenario in this study, triggering the FD failure in this scenario was determined to more reliably result in a missed approach. The sequence of FD failure scenarios was counterbalanced across display type using the Balanced Latin Square method.

Figure 6

Depiction of Normal FD and Failed FD



Symbology with no failures (A) versus symbology after the FD failure (B). White boxes in (A) contain features that were affected by the FD failure. The FD and runway edge lines disappeared, and "AIII" and "RO ARM" annunciations changed to flashing "NO AIII" and "RO CTN."

Supplemental Measures

Usability Questionnaire

Each participant rated the usability of HUD relative to binocular HWD —and binocular HWD relative to monocular HWD—on a 5-point Likert scale ranging from 1 (“Much worse”) to 5 (“Much better”). Display usability was evaluated based on the participants’ own perception of their ability to:

1. Follow flightpath guidance symbology
2. Maintain target airspeed of the approach
3. Transition from instrument to visual flight references
4. Land the aircraft and follow rollout guidance
5. Evaluate the safety of the runway environment
6. Detect failures in the HUD or HWD symbology

Subjective Evaluation of Dominant Versus Non-Dominant Eye with Monocular Head-Worn Display

After the eight familiarization session scenarios, the experimenter conducted a brief open-ended interview in the simulator with the participants, asking whether they experienced any differences in comfort, symbology FOV, and subjective workload between using the monocular HWD on the left eye and on the right eye. Each participant’s responses to this prompt were recorded on the simulator audio recording system and documented in notes by the experimenter.

Testing Environment and Flight Scenarios

Flight Simulator

All study procedures were carried out at the FAA Mike Monroney Aeronautical Center in Oklahoma City, Oklahoma. The flight scenarios were conducted in a CAE Boeing 737-800 Level D-equivalent flight simulator (see Figure 7). In addition to simulating the flight deck and handling qualities of a Boeing 737-800 aircraft, the simulator included a six-axis motion system, a night/dusk/day out-the-window visual model, a comprehensive weather and wind modeling system, and dynamic loading of flight controls. The out-the-window visual system in the simulator featured collimated projectors, so visual information out the windscreen appeared at optical infinity. The simulator was also equipped with an iPad mounted below the left outboard window, which was used to collect NASA-TLX responses during the study session. Data output from the simulator was sampled at a rate of 10 Hz. To prevent interference with the HWD head tracking system, simulator motion was disabled for this study.

Figure 7
Flight Deck of CAE Boeing 737-800 Simulator

Head-Up and Head-Worn Displays

The simulator was equipped with a Collins Aerospace HGS-6700 collimated, production-quality HUD (see Figure 8), and a Microsoft HoloLens 2 non-collimated, commercial off-the-shelf augmented reality headset was used as the HWD (see Figure 9). The symbology in the HUD and HWD was set to AIII mode in all scenarios, which displayed FD guidance, flare guidance, and rollout guidance.

The HoloLens 2 featured a 78% binocular overlap, and the flight symbology fit fully within the binocular region of the total FOV. The center of gravity of the HoloLens 2 is centered on the user's head when it is worn. The HoloLens 2 could be toggled across binocular, monocular left-eye, and monocular right-eye presentation of flight symbology. Regardless of presentation mode, the focal distance, image location, and FOV remained constant to avoid confounding differences in perceived symbology distance, size, or location between monocular and binocular modes. When in binocular mode, the HoloLens was programmed to position the separate monocular images so that, when fused into a binocular image, the vergence angle matched the focal distance of each image to ensure the absence of a vergence-accommodation conflict. The HoloLens 2 included a Moving Platform Mode, which allowed for the device to be used in moving environments such as transportation vehicles. However, this mode was not compatible with the motion system in the simulator. As a result, simulator motion and the HoloLens 2 moving platform mode were disabled in all scenarios.

During the HWD selection process for this research, the Microsoft HoloLens 2 was considered alongside several alternative HWD systems. The HoloLens 2 was chosen because it was the most compatible with the simulator testing environment, was able to present the symbology exactly the way it was presented on the HUD, and featured independently controllable image generators. These characteristics enabled a more direct comparison of pilot performance and workload among the display types. This selection process included an evaluation of the disparity in focal distance between the HoloLens 2 and the collimated HUD. The focal distance disparity was determined to be acceptable for the purposes of this research while also being an important factor to evaluate from a pilot performance and usability standpoint.

Figure 8

Collins Aerospace HGS-6700 Head-Up Display in Simulator

Figure 9
Microsoft HoloLens 2 in Simulator

The HoloLens 2 displayed the symbology image from the HUD using an Epiphan DVI2USB 3.0 video grabber. Use of the video grabber method resulted in an identical appearance of the symbology between the HUD and HWD, which ensured that there were no confounding effects caused by differences in symbology appearance and function between HUD and HWD. The latency between system input and display response of the HUD was measured at approximately 70 milliseconds, and the video grabber introduced an additional 110 ms of latency to the symbology, resulting in approximately 180 ms of latency between control input and display response with the HWD. This measurement is 30 ms higher than the 150 ms requirement for HUD symbology outlined by the FAA National Simulator Program (2012). As a result, subjective testing of HWD latency was carried out in technical working sessions with two Boeing 737 Type-Rated Pilot SMEs with real-world HUD experience and one Pilot SME with HUD and HWD experience. These SMEs agreed that the additional latency of the HWD was not noticeable and did not impact their performance or workload during the approach and landing scenarios used in this study. This conclusion aligns with past research suggesting that latency of basic HUD flight symbology does not begin to impact pilot performance and workload during straight-in approach and landing operations in a fixed-wing aircraft using basic flight symbology until it extends beyond 250 ms, indicating that the HWD latency would probably not negatively impact pilot performance and workload in this study (Bailey et al., 1989, 2004, 2005; Johns & Funk, 1991).

Simulator Scenarios

Each flight scenario involved an ILS approach and landing under Instrument Flight Rules beginning six miles from the runway threshold. At the beginning of each scenario, the aircraft was positioned on the final approach fix (i.e., aligned with the lateral and vertical guidance of the approach), traveling at the target IAS of 152 knots, and configured for landing with the landing gear down and flaps set at 30° before the simulation was started and control was transferred to the PF. All scenarios were flown manually without automated flight control systems (e.g., autopilot, autothrottle, and autoland).

Weather

All experimental flight scenarios involved daytime lighting and overcast clouds, whereby runway visual information was unavailable until it was within the visual range specified for the scenario. All scenarios occurred during daytime to maintain consistency in the brightness of the airfield lighting systems relative to the ambient light. For this study, daytime flight was preferable to nighttime because, when combined with overcast weather, there is reduced contrast between the runway lighting infrastructure and the ambient environment, which reduces the salience of runway visual information, maximizing the impact of the RVR manipulation (Paprocki & Gates, 1966). Overcast weather conditions were chosen to control the point at which runway visual information becomes available, controlling the window to detect and respond to the non-normal events and requiring the PF to focus exclusively on HUD or HWD symbology until runway visual information is within range. Collectively, the daytime, overcast conditions maximize the amount of experimental control over runway visibility so that variation in salience is solely based on the RVR and cloud ceiling employed in the scenario.

Winds

Gusting winds with variable direction and velocity were present from the start of each scenario until the aircraft reached approximately 375 ft AGL, the purpose of which was to increase the difficulty of maintaining a stabilized approach to determine if flying with a binocular or monocular HWD in lieu of a HUD results in an increased likelihood of a missed approach. By introducing variable lateral (i.e., crosswind component) and longitudinal (i.e., headwind component) aerodynamic drag forces, the PF was required to continuously manipulate the flight controls and throttles to maintain a stabilized approach and continue to a landing.

At the beginning of each scenario, the baseline winds were 12 knots from 90° offset from the runway heading with gusts up to 24 knots. As the scenario progressed, the wind direction continuously shifted to be parallel with the runway (i.e., 0° offset, heading 103°), increasing in velocity to 24 knots with gusts up to 48 knots. The wind direction then shifted back so that when the aircraft descended below 375 ft AGL, the gusting winds with variable wind heading dissipated, resulting in sustained 16 knot winds offset 60° left or right of the runway heading (13.86 knots crosswind component). Bringing the winds to a constant direction and velocity by this point of the approach allowed the PF to better stabilize the flightpath and airspeed before reaching the DA/DH. Doing so also controlled for variability in wind direction and velocity across scenarios that would confound the measurements of pilot performance at threshold crossing and during landing. Wind direction was counterbalanced so that in half of the scenarios, the winds began at 90° and ended at 60° offset from the left of the runway, and in the other half of the scenarios, the winds began at 90° and ended at 60° offset from the right of the runway

Runway Environment

All scenarios involved an ILS approach and landing into Portland International Airport (PDX) runway (RWY) 10R. The PDX RWY 10R visual model and approach procedure used in this study featured a runway width of 150 feet, a 3° glideslope angle, 0° runway slope, and a coincident touchdown aiming point and glide path intercept point (GPIP) located 996 feet beyond the runway threshold. Lighting and runway markings remained consistent across all scenarios, with the airfield being equipped with high-intensity runway lighting (HIRL), TDZ lighting, centerline lighting (CLL), and a Medium Intensity Approach Lighting System with Runway Alignment Indicator Lights (MALSR).

Procedure

The FAA Institutional Review Board (IRB) approved all study procedures prior to data collection. Each study session lasted approximately five hours, including breaks. During this time, participants completed an initial briefing, a simulator familiarization session, an alignment eye dominance test, an experimental session with 30 ILS approach and landing scenarios, and a post-experiment debriefing (see Table 2).

Table 2

Approximate Schedule of Study Session

Activity	Pilot Flying	Start Time	End Time	Duration
Initial Briefing		7:00 AM	7:30 AM	30 min
Familiarization Session		7:30 AM	8:30 AM	60 min
Break		*8:30 AM*	*8:40 AM*	*10 min*
Experimental Block 1	Participant A	8:40 AM	8:55 AM	15 min
Experimental Block 2	Participant B	8:55 AM	9:10 AM	15 min
Break		*9:10 AM*	*9:20 AM*	*10 min*
Experimental Block 3	Participant A	9:20 AM	9:35 AM	15 min
Experimental Block 4	Participant B	9:35 AM	9:50 AM	15 min
Break		*9:50 AM*	*10:00 AM*	*10 min*
Experimental Block 5	Participant A	10:00 AM	10:15 AM	15 min
Experimental Block 6	Participant B	10:15 AM	10:30 AM	15 min
Break		*10:30 AM*	*10:40 AM*	*10 min*
Experimental Block 7	Participant A	10:40 AM	10:55 AM	15 min
Experimental Block 8	Participant B	10:55 AM	11:10 AM	15 min
Break		*11:10 AM*	*11:20 AM*	*10 min*
Experimental Block 9	Participant A	11:20 AM	11:35 AM	15 min
Experimental Block 10	Participant B	11:35 AM	11:50 AM	15 min
Debriefing		11:50 AM	12:00 PM	10 min
			Total Time	5 hours

Initial Briefing

During the initial briefing, each participant was provided an iPad showing the Informed Consent form, which outlined their rights, responsibilities, and the purpose of the study. Following their written consent to participate, participants completed a demographics and pilot experience questionnaire. This questionnaire asked participants to report basic demographic information, aviation experience (e.g., total and recent flight hours, total and recent flight hours with a HUD, experience with low-visibility approach and landing operations, prior HWD experience), as well as information about their vision (e.g., type of corrective lenses used).

During the briefing, the experimenter and a Boeing 737 Type-Rated FAA Research Pilot gave a PowerPoint presentation describing the purpose of the study, the approach and landing scenarios, and the standard operating procedures to be followed during the scenarios. Participants were instructed to follow those procedures as outlined by the FAA and their operator. Specifically, when acting as PF, their responsibilities included controlling the flightpath and airspeed from the left seat of the simulator, adjusting the trajectory of the airplane using manual inputs to the yoke and throttles without any automated flight control systems, and using the symbology presented on the HWD or HUD. When acting as PM, participant responsibilities involved carrying out monitoring duties from the right seat of the simulator without a HUD or HWD (FAA, 2022). Participants were briefed on one deviation from standard operating procedures for the study session: They were instructed to refrain from all verbal callouts as PM throughout the scenario, including the missed approach callout. Participants were given this instruction to (a) prevent them from alerting the PF to flightpath and airspeed deviations or any non-normal events during the scenarios, and (b) prevent the PM from calling a missed approach. This placed all responsibility for maintaining a stabilized approach and detecting non-normal events on the PF, enabling better investigation of the impact of display type and RVR on these processes.

Part of this briefing included an explanation of the performance expectations during the scenario, including (a) maintaining the vertical and lateral profiles of the approach as well as possible, (b) maintaining the target airspeed of the approach as well as possible, and (c) landing as close to the runway touchdown point as possible with minimal sink rate. The experimenter provided each participant with a description of the NASA-TLX, including definitions of the subscales, instructions for completing the subscale portion, and instructions for completing the pairwise comparison portion at the end of the experimental session. Participants were given hands-on familiarization with the Microsoft HoloLens 2 headset during the briefing and were given instructions on how to wear and adjust it, as well as how to perform the alignment process before each HWD scenario. The participants then viewed a training video that described and demonstrated the HUD and HWD specific to approach and landing. Cumulatively, the initial briefing lasted approximately 30 minutes.

Familiarization Session

After the initial briefing, the participants entered the simulator for a familiarization session. A Boeing 737 type-rated FAA pilot gave each participant a flight deck walk-through to acquaint them with the simulator. For each participant, the Microsoft HoloLens calibration process was carried out to account for individual differences in interpupillary distance and vertical offset. The participants then completed a practice session designed to familiarize them with the ILS approach and landing scenarios in the experimental session, the HUD and HWD, and their PF and PM tasks. In this session, each participant completed four approach and landing practice scenarios as PF, and a second set of four as PM (see Table 3). Before each HWD scenario, the Microsoft HoloLens calibration process was carried out to check and correct symbology alignment. Cumulatively, the familiarization session lasted approximately 60 minutes.

Table 3

Familiarization Session Scenarios

Scenario	Display Type	RVR/Ceiling (ft)	Pilot Flying	Wind Direction
1	HUD	CAVOK	Participant A	None
2	B. HWD	4800/250	Participant A	Left
3	M. HWD Left Eye	4800/250	Participant A	Left
4	M. HWD Right Eye	4800/250	Participant A	Left
		Participants Switch Seats		
5	HUD	CAVOK	Participant B	None
6	B. HWD	4800/250	Participant B	Left
7	M. HWD Left Eye	4800/250	Participant B	Left
8	M. HWD Right Eye	4800/250	Participant B	Left

All crews completed the familiarization session scenarios in sequential order. CAVOK refers to clear weather, where the cloud ceiling is greater than 5,000 ft MSL altitude and the prevailing visibility is greater than 10 miles.

Pilot Feedback on Eye Dominance with Monocular HWD

At the end of the familiarization session, the experimenter conducted a brief open-ended interview in the simulator with the participants, asking whether they experienced any differences in comfort, symbology FOV, and subjective workload between using the monocular HWD on the left eye and on the right eye. This feedback was collected before participants were made aware that their sighting eye dominance would be measured to minimize any bias in their responses. Each participant's responses to this prompt were recorded on the simulator audio recording system and documented in notes by the experimenter.

Sighting Eye Dominance Test

After the familiarization session, each participant was tested for sighting eye dominance using the Distance Hole-In-The-Card test (Johannson et al., 2015). In this test, the participant was given a piece of black cardstock with a 3-centimeter-diameter circular hole cut in the center of the card. The participant was instructed to hold the cardstock at arm's length while viewing a single letter with both eyes open, positioned 10 feet in front of the participant. The participant then drew the card toward their face. The eye with which the participant viewed the letter after drawing the card toward their face was recorded as the dominant eye. The experimenter recorded the results of this test for each participant.

Experimental Session

Flight Scenario Procedure

Following the eye dominance test, each participant completed 15 ILS approach and landing scenarios as PF, for a total of 30 scenarios completed by each crew. Before each scenario, the experimenter briefed the crew on the display type in use and whether the scenario utilized a CAT I (i.e., 4800 ft RVR), CAT II (i.e., 1200 ft RVR), or CAT III (i.e., 600 ft RVR) approach procedure. The PM then completed the approach checklist. If the scenario involved the use of the HUD, the PF removed the HWD, transferred it to the experimenter, and deployed the HUD combiner. If the scenario involved the use of the HWD, the PF put on the HWD, checked and adjusted symbology alignment, and stowed the HUD combiner. When the PM announced that the checklist was complete and the PF announced that the HUD or HWD was set up and HWD alignment was verified, the experimenter announced, “cleared to land,” and started the scenario from inside the simulator.

After touchdown, the PF decelerated the aircraft to a stop on the runway during rollout with autobrakes set to MAX. The simulator automatically ended the scenario and began loading the next scenario after the aircraft had decelerated to 20 knots on the runway surface. At this point, the PF completed the NASA-TLX rating scale on the iPad. After each participant’s final scenario as PF, they completed the pairwise comparison portion of the NASA-TLX on the iPad. Each scenario, which includes the approach briefing and checklist, HWD alignment, flight scenario, and NASA-TLX rating scale, lasted approximately five minutes.

Session Structure and Scenario Sequence

Each independent variable (IV) and each normal and non-normal scenario were fully permutated within each participant, so that each participant experienced every level of each IV and all normal and non-normal scenarios. Eighteen of the scenarios (nine per participant) were normal, routine operations, with no non-normal or non-routine events that may have prevented the crew from fully completing the approach and landing. In addition to the normal operations scenarios, crews flew an additional six scenarios with 4800 ft RVR (three per participant as PF) that included a runway incursion and an additional six scenarios with 600 ft RVR (three per participant as PF) that included an FD failure. Display type order was counterbalanced within each block so that each block contained one HUD scenario, one binocular HWD scenario, and one monocular HWD scenario. RVR order was pseudo-randomized across each study session.

To minimize the expectancy of the non-normal events, scenarios that included a runway incursion or FD failure were randomly distributed among the normal scenarios, with the exception that the first scenario for each crew was always a normal scenario. The non-normal scenarios were not randomly sequenced; rather, they were counterbalanced among the normal scenarios using the Balanced Latin Square method. Not only did this account for order and sequence effects among the non-normal scenarios, but it also allowed for between-subjects comparisons of responses to the first, unexpected occurrence of each runway incursion and FD failure as a function of display type. The 30 scenarios were structured into 10 blocks of three scenarios each, with each block lasting approximately 15 minutes. After each block, the crew swapped seats in the simulator so that the participant who acted as PF in the previous block became the PM, and vice versa. After blocks 2, 4, 6, and 8, the crew took a 10-minute break (see Table 2). Cumulatively, the experimental session lasted approximately 3 hours and 20 minutes, including breaks.

Post-Experiment Procedures

After the final block of the experimental session, the participants exited the simulator and were taken to the briefing room, where they completed a usability questionnaire. The experimenter and FAA research pilot then conducted a brief unstructured interview to gather additional feedback on the display types and flight scenarios, debriefed the participants, and closed out the study session.

Results

The data analysis plan for this research was designed to determine whether pilot performance and workload during approach and landing differ among the three display types, as well as among three degrees of flight visibility (i.e., RVR and cloud ceiling). The data analysis plan was also designed to determine whether the ability to detect and respond to non-normal events during the approach differs as a function of display type. Prior to analyses, the data were inspected to determine the presence of outliers. Unless otherwise specified, outliers were retained in the data for all analyses. This determination was made after comparing results from analyses where outliers were retained against results from those same analyses with outliers removed, with the finding that removing outliers did not meaningfully impact the model. Furthermore, the outliers were retained because, in many cases, they may represent infrequent yet critical events that occur as a function of the display type. This is especially important for operationally focused data, such as those analyzed herein, where outliers often represent performance anomalies that could result in an incident or accident, and removing those outliers may mask important risk factors (Achour et al., 2023; Rey et al., 2021).

Prior to analyses, the data were inspected for normality and homogeneity of variance. In cases where the normality assumption was not met, a Box-Cox transformation was applied to the raw values (Malik et al., 2018). The Box-Cox transformation parameter, lambda (λ), was determined using the Shapiro-Wilk normality test method from the Box-Cox transformation function in the AID package of R (v2.9; Dag et al., 2017). This method was used to estimate an optimal lambda from the data, with the aim of transforming the distribution into an approximately normal one (Osborne, 2010). This approach was taken for all analyses involving data transformation using the Box-Cox method. All main effects and interaction effects involving repeated measures were adjusted using the Greenhouse-Geisser method to correct for any violations of sphericity. The Bonferroni correction was applied to the p-value for all post hoc multiple comparisons. Unless otherwise indicated, $\alpha = .05$ for all comparisons. Throughout the Results section, box-and-whisker plots are provided to visualize the data. For all box-and-whisker plots, the median is indicated by a black line within each box. The box represents the interquartile range (IQR), encompassing the middle 50% of the data. The whiskers extend to 1.5 × IQR, visualizing the variability beyond the upper and lower quantiles. Data points that lie above or below the ends of the whiskers are outliers.

Frequency of Missed Approaches versus Landings

Each normal scenario was evaluated to determine whether the PF continued to a landing or conducted a missed approach to evaluate the likelihood of the PF conducting a missed approach as a function of display type and RVR. As described in the Method section of this report, the PM was asked to refrain from calling out missed approaches, placing all responsibility of making that determination on the PF to better evaluate any effects of display type on that decision-making process.

Among all 432 normal scenarios in the complete dataset, there were seven scenarios (1.62%) in which the PF executed a missed approach. As shown in Table 4, these missed approaches were not isolated to any one condition in the study. They appeared to be distributed somewhat evenly across the display type and RVR conditions, indicating that the ability to maintain a stabilized approach and continue to a landing is roughly consistent regardless of whether the PF uses symbology on a HUD, binocular HWD, or monocular HWD to manually fly instrument approaches with significant gusting winds and visibilities as low as 600 ft RVR.

Table 4

Frequency of Landings versus Missed Approaches in Normal Scenarios

		Display Type			
	RVR	HUD	B. HWD	M. HWD	**Total**
Landings	*600 ft*	*47*	*47*	*47*	141
	1200 ft	*48*	*48*	*47*	143
	4800 ft	*47*	*48*	*46*	141
	Total	142	143	140	**425**
Missed Approaches	*600 ft*	*1*	*1*	*1*	3
	1200 ft	*0*	*0*	*1*	1
	4800 ft	*1*	*0*	*2*	3
	Total	2	1	4	**7**
Total	*600 ft*	*48*	*48*	*48*	144
	1200 ft	*48*	*48*	*48*	144
	4800 ft	*48*	*48*	*48*	144
	Total	144	144	144	**432**

Instrument Segment Performance

The ability of the PF to manage the flightpath and aircraft energy as a function of display type was measured from release of the simulator to 300 ft AGL. A simulator issue occurred during the instrument segment for 3 of the 24 crews (12.5% of all data). This simulator issue caused the FD in the HUD and HWD symbology to disappear for approximately 10 seconds at the beginning of each scenario, causing the flight crew to deviate significantly from the flightpath and requiring significant flightpath correction once the FD reappeared. Inspection of the data revealed that this issue impacted flightpath tracking and airspeed management performance during the first 60 seconds of the instrument segment of the approach (approximately the first 27% of the instrument segment). Therefore, the data from these crews were excluded from analyses of instrument segment performance, resulting in data from 21 of the 24 crews (87%)

being retained for analyses. Among the data for 21 crews that were retained for analyses, six missed approaches occurred in 378 normal scenarios (1.59%). However, no instrument segment data were excluded because of a missed approach.

Root Mean Square Flightpath Deviation

The ability of the PF to maintain the lateral and vertical flightpath of the approach (i.e., Flight Technical Error [FTE]; Levy, Som, & Greenhaw, 2003) during the instrument segment was evaluated within normal scenarios (i.e., without a runway incursion or FD failure) as a function of display type using a one-way, repeated measures Analysis of Variance (ANOVA). Because there was no operational impact of RVR or cloud ceiling during the measurement period for instrument segment performance, data were combined across RVR conditions for this analysis. The distribution of the raw data was positively skewed, so the data were subjected to a Box-Cox transformation ($\lambda = -0.54$), which successfully normalized the distribution. Analysis of the transformed data revealed no significant effect of display type on RMS flightpath deviation ($p = .156$). As shown in Figure 10, several outliers were present in the flightpath deviation data across each display type, demonstrating that the significant gusting winds with variable direction present in the scenarios made it challenging to maintain the flightpath. Table 5 presents descriptive statistics for RMS flightpath deviation in the instrument segment.

Figure 10

Box-and-Whisker Plot of Root Mean Square Flightpath Deviation with Each Display Type

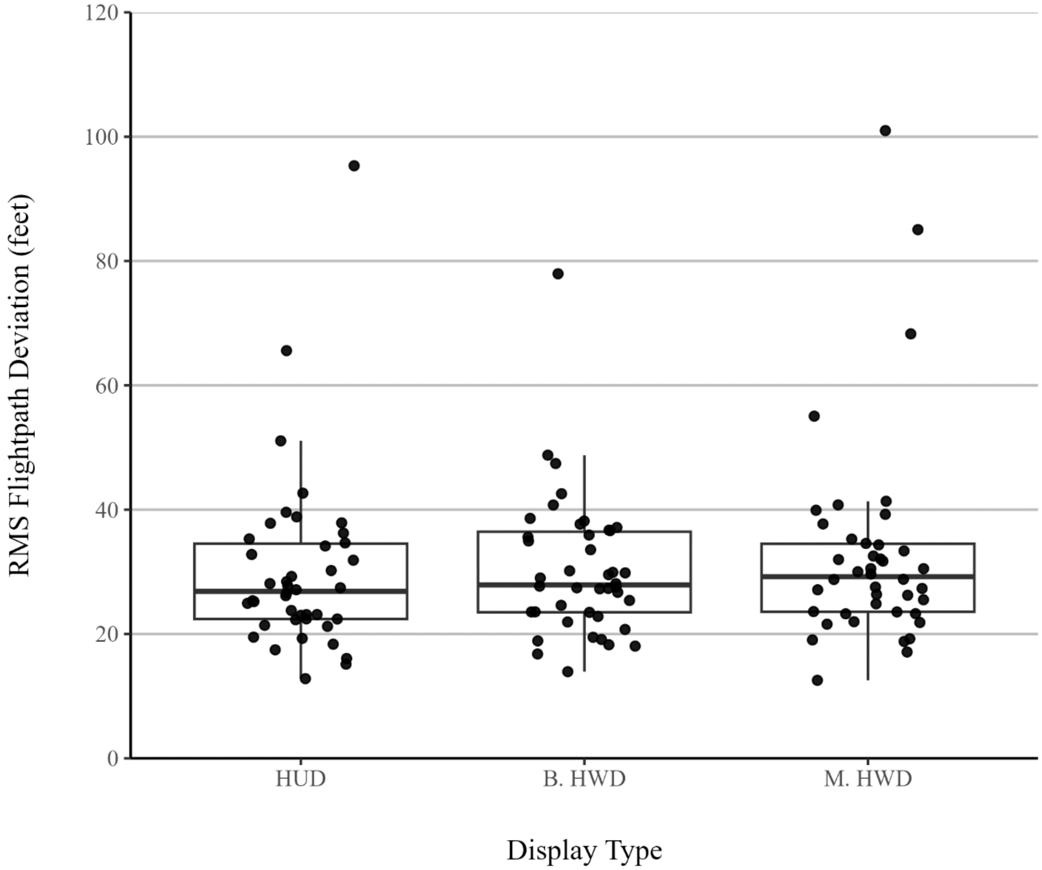


Table 5

Adjusted Means, Standard Errors, and Confidence Intervals for Root Mean Square Flightpath Deviation with Each Display Type

Display Type	M_{adj} (ft)	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
HUD	30.04	2.21	25.58	34.50
B. HWD	30.32	1.73	26.83	33.82
M. HWD	32.93	2.61	27.67	38.19

Root Mean Square Airspeed Deviation

The ability of the PF to maintain the target airspeed of the approach, in knots IAS, during the instrument segment was evaluated within normal scenarios as a function of display type using a one-way, repeated-measures ANOVA. Because there was no operational impact of RVR or cloud ceiling during the measurement period for instrument segment performance, data were combined across RVR conditions for this analysis. The distribution of the raw data was positively skewed, so the data were subjected to a Box-Cox transformation ($\lambda = 0.05$), which successfully normalized the distribution. Analysis of the transformed data revealed that display type had a significant effect on RMS deviation from the target IAS during the instrument segment, $F(1.673, 68.582) = 13.062$, $p < .001$, $\eta_p^2 = .242$. Post-hoc pairwise comparisons between each experimental condition revealed that RMS airspeed deviation was significantly higher when participants flew with a monocular HWD than when they flew with a binocular HWD ($p = .023$, $d = 0.381$) or HUD ($p < .001$, $d = 0.610$). RMS airspeed deviation was also significantly higher when participants flew with a binocular HWD than when they flew with a HUD ($p = .014$, $d = 0.447$; see Figure 11). However, as demonstrated by the mean airspeed deviation values presented in Table 6, these effects were small in magnitude and are unlikely to translate into operationally significant differences in pilot performance.

Figure 11

Box-and-Whisker Plot of Root Mean Square Airspeed Deviation with Each Display Type

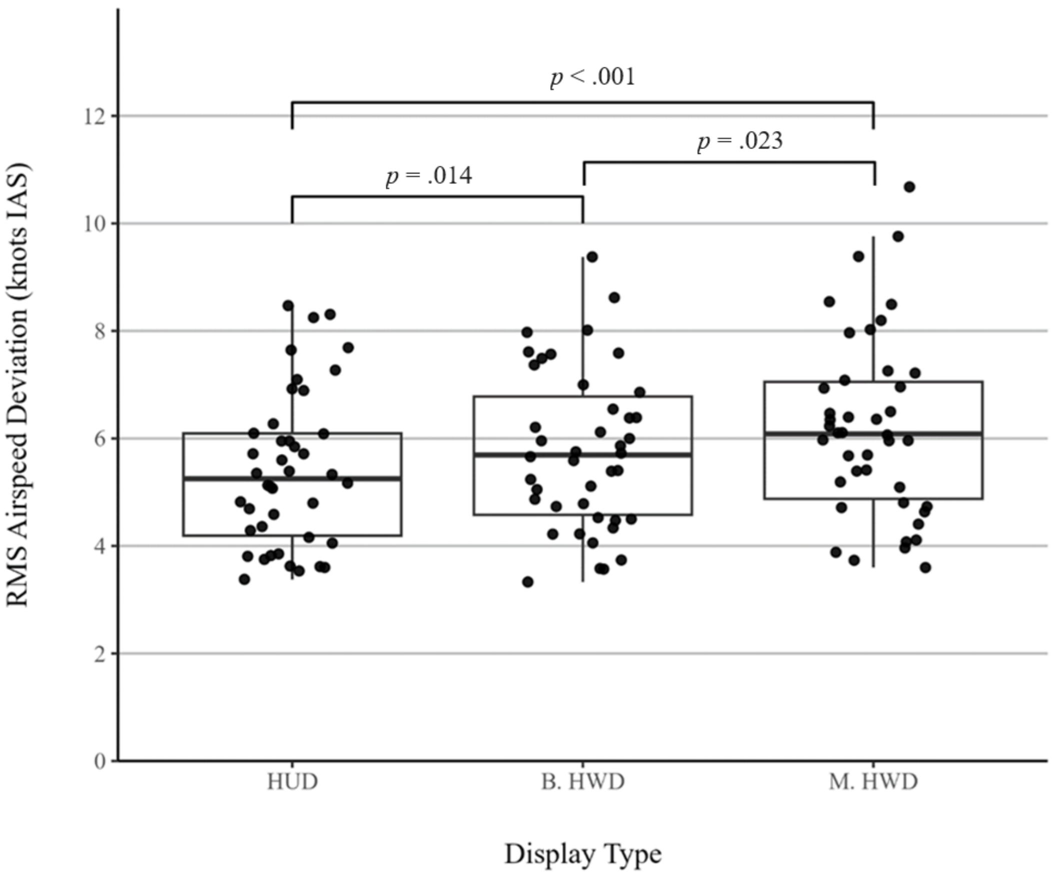


Table 6
Adjusted Means, Standard Errors, and Confidence Intervals for Root Mean Square Airspeed Deviation with Each Display Type

Display Type	M_{adj} (knots)	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
HUD	5.41	0.22	4.96	5.86
B. HWD	5.78	0.23	5.32	6.24
M. HWD	6.19	0.26	5.67	6.72

Threshold Crossing Performance

The ability of the PF to maintain the flightpath and target airspeed of the approach while crossing the runway threshold was measured for all normal scenarios. Normal scenarios in which participants conducted a missed approach while they were flying were excluded from analyses of threshold crossing performance because the missed approach procedure did not involve maintaining the flightpath and target airspeed of the approach at the point when the aircraft crossed the runway threshold. Among all 432 normal scenarios in the complete dataset, there were seven scenarios where the PF initiated a missed approach prior to crossing the runway threshold (1.62% of all scenarios; see Table 4). As a result, data from 425 (98.38%) normal scenarios were retained for analyses.

Localizer Deviation at Threshold Crossing

The ability of the PF to follow the lateral guidance of the approach defined by the ILS localizer while crossing the runway threshold was evaluated using a 3 (display type) x 3 (RVR) repeated measures ANOVA. Because the simulator outputted localizer deviation values to the left of the localizer as negative values and values to the right of the localizer as positive values—and because crosswind direction can influence whether the aircraft tends to deviate to the left or to the right of the localizer—raw localizer deviation values were converted to absolute values by multiplying all values by -1. The distribution of the raw data was positively skewed, so the data were subjected to a Box-Cox transformation ($\lambda = 0.53$), which successfully normalized the distribution. Analysis of the transformed data revealed that there was not a significant effect of display type on localizer deviation at threshold crossing ($p = .733$). However, there was a significant effect of RVR on localizer deviation at threshold crossing, $F(1.946, 77.844) = 5.379$, $p = .007$, $\eta_p^2 = .119$.

Post-hoc pairwise comparisons between RVR conditions revealed that localizer deviation at threshold crossing was significantly higher with 600 ft RVR than it was with 4800 ft RVR ($p = .010$, $d = 0.329$). However, as demonstrated by the mean values presented in Table 7, this statistically significant effect likely does not translate to operational significance: Mean deviation values in the condition where localizer deviation at threshold crossing was lowest (i.e., HUD, 4800 ft RVR) were only 3.86 ft lower than in the condition where deviation was highest (i.e., binocular HWD, 600 ft RVR). Patterns among the means also indicate that localizer deviation at threshold crossing was higher with 600 ft RVR than it was with 1200 ft RVR and was higher with 1200 ft RVR than with 4800 ft RVR. However, these effects did not reach statistical significance, nor do these differences suggest any operational significance ($p > .05$, in each case). Figure 12 presents a box-and-whisker plot of localizer deviation at threshold crossing in each experimental condition. As shown in the figure, notable variability is present in the data, and several conditions contain outliers, demonstrating that the 13.86-knot crosswind component present at threshold crossing scenarios made it challenging to manage the flightpath leading up to landing.

Figure 12

Box-and-Whisker Plot of Localizer Deviation at Threshold Crossing in Each Experimental Condition

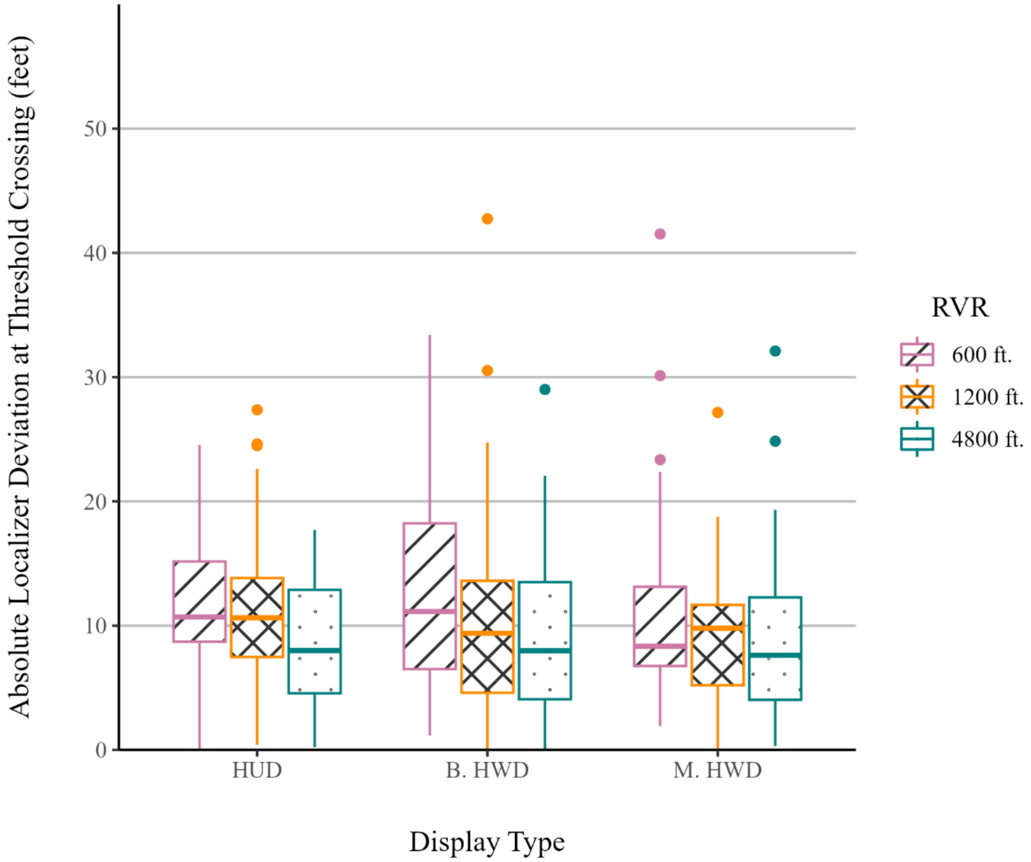


Table 7

Adjusted Means, Standard Errors, and Confidence Intervals for Localizer Deviation at Threshold Crossing in Each Experimental Condition

Display Type	RVR (ft)	M_{adj} (ft)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	600	11.42	0.88	9.64	13.20
	1200	10.73	1.00	8.70	12.75
	4800	8.61	0.79	7.01	10.21
B. HWD	600	12.47	1.17	10.11	14.84
	1200	10.66	1.35	7.94	13.38
	4800	9.20	1.05	7.08	11.33
M. HWD	600	11.31	1.26	8.76	13.87
	1200	8.96	0.81	7.31	10.60
	4800	9.30	1.08	7.12	11.48

Glideslope Deviation at Threshold Crossing

The ability of the PF to follow the vertical guidance of the approach defined by the ILS glideslope while crossing the runway threshold was evaluated using a 3 (display type) x 3 (RVR) repeated measures ANOVA. The ANOVA revealed that there was a statistically significant effect of display type on glideslope deviation at threshold crossing, $F(1.809, 72.377) = 7.277$, $p = .002$, $\eta_p^2 = .154$. Post-hoc comparisons between display type conditions revealed that glideslope deviation was greater with the HUD than it was with the binocular HWD ($p = .004$, $d = 0.226$) or monocular HWD ($p = .004$, $d = 0.225$), whereas there was no significant difference between binocular HWD and monocular HWD conditions ($p > .999$). Patterns among the means indicate that the difference in glideslope deviation between the HUD and binocular HWD conditions is not operationally significant: On average, the aircraft crossed the runway threshold 3.57 feet above the glideslope when the PF was flying with the HUD—1.86 feet higher than when flying with a binocular HWD, and 2.34 feet higher than when flying with a monocular HWD (see Figure 13 and Table 8).

There was also a statistically significant effect of RVR on glideslope deviation at threshold crossing, $F(1.936, 77.446) = 14.685$, $p < .001$, $\eta_p^2 = .269$. Post-hoc pairwise comparisons between RVR conditions revealed that glideslope deviation at threshold crossing was significantly less with 4800 ft RVR than it was with 600 ft RVR ($p < .001$, $d = 0.454$) and with 1200 ft RVR ($p < .001$, $d = 0.445$); there was not a significant difference between 600 ft RVR and 1200 ft RVR conditions ($p = .915$). Again, these statistically significant differences do not appear to translate to operational significance. Patterns among the means indicate that, on average, the aircraft tended to cross the runway threshold 3.75 feet above the glideslope in scenarios with 600 ft RVR and 2.92 feet above the glideslope with 1200 ft RVR, whereas with 4800 ft RVR the aircraft crossed the runway threshold 0.16 feet below the glideslope (see Figure 13 and Table 8). Figure 13 also illustrates several outliers and notable variability in each condition, demonstrating that the significant winds present in the scenarios made it challenging for the PF to manage the flightpath leading up to the landing.

Figure 13

Box-and-Whisker Plot of Glideslope Deviation at Threshold Crossing in Each Experimental Condition

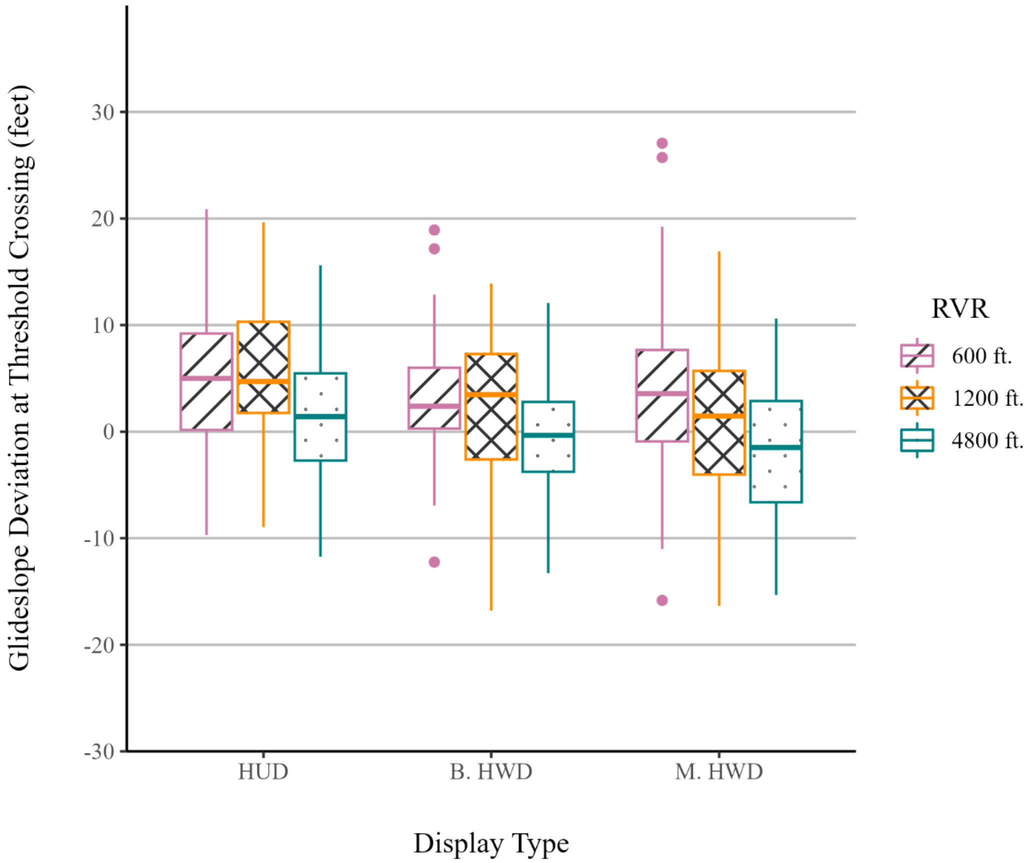


Table 8

Adjusted Means, Standard Errors, and Confidence Intervals for Glideslope Deviation at Threshold Crossing in Each Experimental Condition

Display Type	RVR (ft)	M_{adj} (ft)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	600	4.27	1.14	1.97	6.57
	1200	4.87	0.96	2.93	6.82
	4800	1.56	0.93	-0.33	3.45
B. HWD	600	3.20	0.99	1.21	5.20
	1200	2.43	1.08	0.24	4.61
	4800	-0.50	0.86	-2.24	1.25
M. HWD	600	3.78	1.42	0.90	6.65
	1200	1.45	1.13	-0.84	3.74
	4800	-1.55	1.04	-3.65	0.56

Absolute Airspeed Deviation at Threshold Crossing

The ability of the PF to maintain the target IAS of the approach while crossing the runway threshold was evaluated using a 3 (display type) × 3 (RVR) repeated measures ANOVA. Prior to analyses, absolute airspeed deviation at threshold crossing was computed in each normal scenario by (1) calculating the difference between the raw IAS and target airspeed (i.e., 152 knots) as the aircraft center of gravity crossed the runway threshold and (2) converting the result to an absolute (i.e., non-negative) value. There was a statistically significant main effect of display type on airspeed deviation at threshold crossing, $F(1.978, 79.121) = 7.11$, $p = .002$, $\eta_p^2 = .151$. Post-hoc pairwise comparisons between display type conditions revealed that airspeed deviation at threshold crossing was higher when the PF flew with a monocular HWD than with a HUD ($p = .005$, $d = 0.346$) or binocular HWD ($p = .009$, $d = 0.310$). However, as demonstrated by the mean values presented in Table 9, these differences were small in magnitude and are unlikely to represent an operationally significant difference.

There was not a significant difference in airspeed deviation at threshold crossing between the HUD and the binocular HWD conditions ($p > .999$). The ANOVA revealed that there was not a significant main effect of RVR on airspeed deviation at threshold crossing, nor was there a significant interaction between display type and RVR ($p > .05$, in each case). Figure 14 presents airspeed deviation at threshold crossing in each experimental condition. As shown in the figure, several outliers are present, and the overall variability across conditions is high, indicating that the winds present in each scenario made it challenging to manage the energy of the aircraft, even after the wind gusts had dissipated.

Figure 14

Box-and-Whisker Plot of Airspeed Deviation at Threshold Crossing in Each Experimental Condition

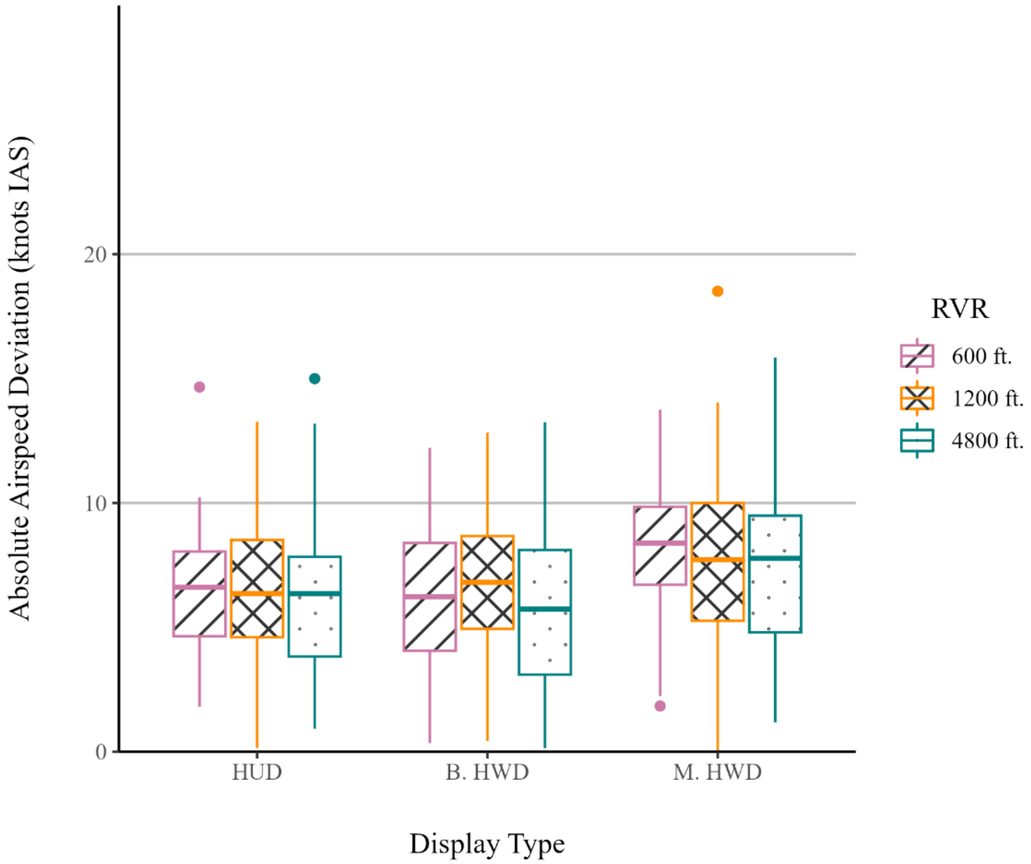


Table 9

Adjusted Means, Standard Errors, and Confidence Intervals for Airspeed Deviation at Threshold Crossing in Each Experimental Condition

Display Type	RVR (ft)	M_{adj} (knots)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	600	6.41	0.45	5.51	7.31
	1200	6.57	0.49	5.57	7.57
	4800	6.37	0.53	5.29	7.45
B. HWD	600	6.34	0.47	5.40	7.28
	1200	6.85	0.44	5.96	7.73
	4800	6.10	0.56	4.98	7.23
M. HWD	600	8.09	0.47	7.15	9.04
	1200	7.41	0.60	6.20	8.63
	4800	7.32	0.52	6.28	8.37

Landing and Rollout Performance

Landing and rollout performance was measured for all normal scenarios, and scenarios where a missed approach occurred were excluded from analyses. Among all 432 normal scenarios in the complete dataset, seven missed approaches occurred (1.62% of all scenarios; see Table 4). Thus, as was the case for threshold crossing performance analyses, data from 425 (98.38%) normal scenarios were retained for landing and rollout performance analyses.

Lateral Deviation from Runway Centerline at Touchdown

The lateral position of the aircraft center of gravity relative to the runway centerline, in feet, at the point when the main landing gear contacted the runway surface was evaluated using a 3 (display type) × 3 (RVR) repeated measures ANOVA. Because the simulator output data recorded deviation to the left of the runway centerline as a negative value and deviation to the right of the runway centerline as a positive value, raw deviation values were converted to absolute values prior to analyses. The ANOVA revealed that there was not a significant main effect of display type on lateral deviation from the runway centerline at touchdown (p = .847). Conversely, there was a significant main effect of RVR on lateral deviation from the runway centerline at touchdown, $F(1.939, 77.571) = 17.49$, $p < .001$, $\eta_p^2 = .304$.

Post-hoc comparisons between RVR conditions indicated that deviation was highest in scenarios with 600 ft RVR (M = 15.06; SE = 0.73 feet), second highest in scenarios with 1200 ft RVR (M = 12.23, SE = 0.71 feet), and lowest in scenarios with 4800 ft RVR (M = 10.46, SE = 0.54 feet). Deviation in scenarios with 600 ft RVR was significantly higher than in scenarios with 4800 ft RVR (p < .001, d = 0.508), as well as in scenarios with 1200 ft RVR (p = .005, d = 0.310). Deviation in scenarios with 1200 ft RVR was marginally higher than deviation in scenarios with 4800 ft RVR (p = .054, d = 0.214).

While these differences were statistically significant, the small differences in mean values presented in Table 10 suggest that they are not operationally significant. Figure 15 presents lateral deviation from the runway centerline at touchdown in each experimental condition. As shown in the plot, there were numerous outliers in the data, particularly in 600 ft RVR scenarios flown with the monocular or binocular HWD, which may have been a product of the 16 knot winds present at landing, compounded by the non-collimated HWD image that may have led to greater difficulty with following the AIII mode flare guidance to touchdown.

Figure 15

Box-and-Whisker Plot of Lateral Deviation from the Runway Centerline at Touchdown in Each Experimental Condition

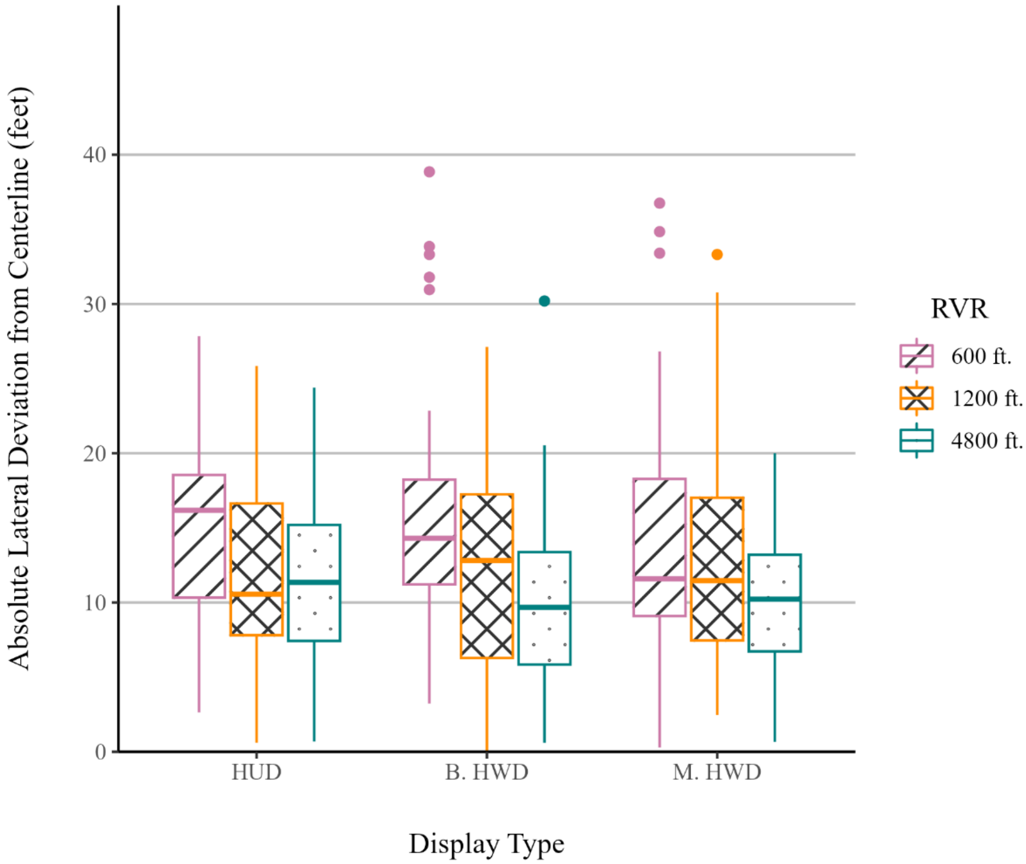


Table 10

Adjusted Means, Standard Errors, and Confidence Intervals for Lateral Deviation from Runway Centerline at Touchdown in Each Experimental Condition

Display Type	RVR (ft)	M_{adj} (ft)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	600	15.27	0.97	13.30	17.23
	1200	11.55	0.95	9.64	13.47
	4800	11.29	0.92	9.44	13.14
B. HWD	600	15.84	1.27	13.28	18.40
	1200	12.23	1.10	10.01	14.46
	4800	10.21	0.98	8.23	12.19
M. HWD	600	14.06	1.30	11.44	16.69
	1200	12.91	1.12	10.65	15.16
	4800	9.89	0.76	8.35	11.42

Distance from Runway Threshold at Touchdown

The longitudinal position of the aircraft center of gravity relative to the runway threshold, in feet, at the point when the main landing gear contacted the runway surface was evaluated using a 3 (display type) × 3 (RVR) repeated measures ANOVA. The ANOVA revealed that there was a significant main effect of display type on distance from the runway threshold at touchdown, $F(1.779, 71.169) = 13.505$, $p < .001$, $\eta_p^2 = .252$.

Post-hoc pairwise comparisons between display type conditions revealed that pilots landed farther from the runway threshold when they flew with a HUD than they did when they flew with a binocular HWD ($p < .001$, $d = 0.340$) or monocular HWD ($p = .004$, $d = 0.338$;). There was not a significant difference in distance from runway threshold at touchdown between binocular HWD and monocular HWD conditions ($p = .815$). There was not a significant effect of RVR on distance from runway threshold at touchdown ($p = .675$), nor was there a significant interaction effect between display type and RVR ($p = .909$).

When considered alongside the data for glideslope deviation at threshold crossing, it appears that the PF tended to fly higher and land later when using symbology on the HUD, whereas they tended to fly lower and land earlier when using symbology on the monocular and binocular HWD. While these differences were statistically significant, the small difference in mean values shown in Table 11 indicate that the differences are likely not operationally significant. Figure 16 presents the distance from the runway threshold at touchdown in each experimental condition. As shown in the figure, notable variability is present in the data, and several outliers are present, regardless of whether the PF flew with a HUD, binocular HWD, or monocular HWD, and regardless of the RVR. It is possible that the gusting winds aloft and high sustained surface wind at landing made it challenging for the PF to land consistently on the same point within the runway TDZ.

Figure 16

Box-and-Whisker Plot of Longitudinal Distance from the Runway Threshold at Touchdown in Each Experimental Condition

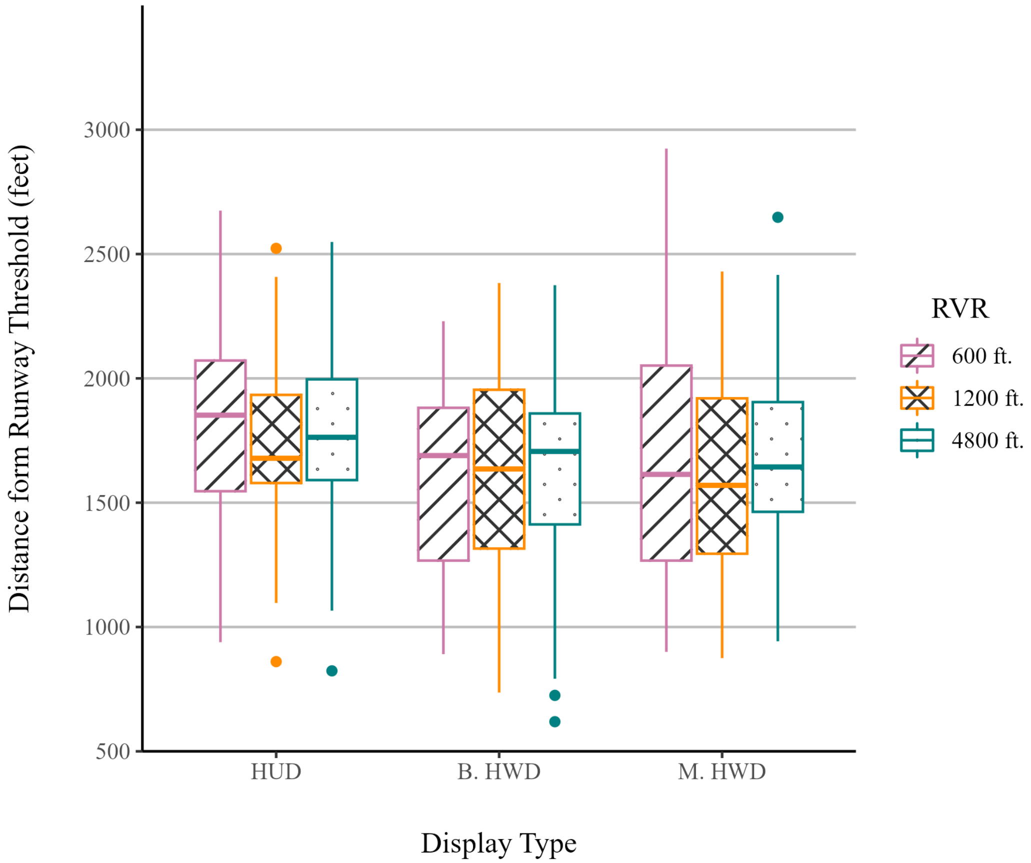


Table 11

Adjusted Means, Standard Errors, and Confidence Intervals for Longitudinal Distance from Runway Threshold at Touchdown in Each Experimental Condition

Display Type	RVR	M_{adj} (ft)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	600	1778.31	66.88	1643.14	1913.49
	1200	1737.62	55.24	1625.97	1849.27
	4800	1782.94	61.44	1658.76	1907.13
B. HWD	600	1590.44	61.98	1465.17	1715.72
	1200	1611.87	68.58	1473.28	1750.47
	4800	1614.81	64.91	1483.61	1746.01
M. HWD	600	1676.01	78.02	1518.33	1833.70
	1200	1606.01	62.14	1480.43	1731.60
	4800	1651.89	59.14	1532.36	1771.42

Sink Rate at Touchdown

The mean vertical velocity, in ft/second, during the two seconds leading up to the aircraft's main landing gear contacting the runway surface was evaluated using a 3 (display type) × 3 (RVR) repeated measures ANOVA. Initial inspection of the data revealed that one of the 47 data points (2.1%) for sink rate with the monocular HWD at 1200 ft RVR was a positive value (0.54 ft/second). Because the aircraft was determined to be descending immediately prior to touching down on the runway surface in this scenario, this data point was determined to be a result of simulator measurement error and was excluded from the ANOVA. The ANOVA revealed that there was a significant main effect of display type on sink rate at touchdown, $F(1.972, 79.923) = 6.631$, $p = .002$, $\eta_p^2 = .145$.

Post-hoc pairwise comparisons between display type conditions revealed that sink rate at touchdown was lower when pilots flew with a HUD than when they flew with a binocular HWD ($p = .007$, $d = 0.213$) or monocular HWD ($p = .008$, $d = 0.295$). There was not a significant difference in sink rate at touchdown between the binocular HWD and monocular HWD conditions ($p > .999$). There was also not a significant main effect of RVR on sink rate at touchdown ($p = .314$), nor was there a significant interaction effect between display type and RVR ($p = .831$).

While there was a statistically significant effect of display type on sink rate at touchdown, the small differences in mean sink rate values presented in Table 12 suggest that this effect is not operationally significant: In the condition where the mean sink rate at touchdown was greatest (binocular HWD, 600 ft RVR), the mean sink rate was 1.23 ft/second greater than in the condition where the mean sink rate was lowest (HUD, 1200 ft RVR). Figure 17 presents the sink rate at touchdown in each experimental condition. As shown in the figure, several outliers are present in the data, with several outliers in the monocular and binocular HWD conditions exceeding a sink rate of 10 ft/second. Several contributing factors may have led to these hard landings, including (1) high overall approach speeds to account for gusting winds during the scenarios; (2) simulator motion disabled to avoid interference with the HWD head tracking system; and (3) a non-collimated image on the HWD.

Figure 17

Box-and-Whisker Plot of Sink Rate at Touchdown in Each Experimental Condition

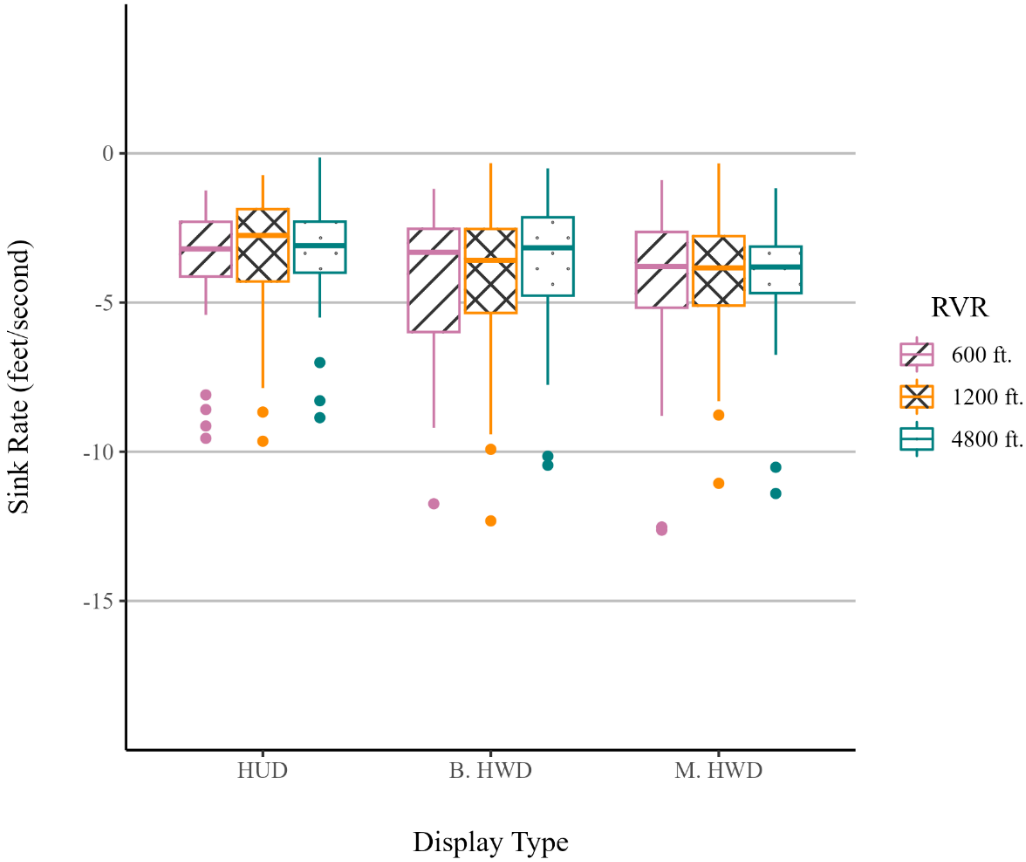


Table 12

Adjusted Means, Standard Errors, and Confidence Intervals for Sink Rate at Touchdown in Each Experimental Condition

Display Type	RVR (ft)	M_{adj} (ft/second)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	600	-3.69	0.33	-4.35	-3.03
	1200	-3.28	0.33	-3.95	-2.60
	4800	-3.35	0.29	-3.94	-2.77
B. HWD	600	-4.51	0.40	-5.31	-3.71
	1200	-4.22	0.43	-5.09	-3.35
	4800	-3.71	0.36	-4.43	-2.99
M. HWD	600	-4.29	0.41	-5.13	-3.46
	1200	-4.22	0.39	-5.01	-3.43
	4800	-4.06	0.34	-4.74	-3.38

Deviation from Localizer During Rollout

The ability of the PF to follow localizer guidance on the HUD and HWD and track the runway centerline during rollout was evaluated using a 3 (display type) x 3 (RVR) repeated measures ANOVA. The ANOVA revealed no significant effect of display type on RMS deviation from the localizer during rollout (p = .402). Conversely, there was a significant effect of RVR on deviation from the localizer during rollout, $F(1.925, 76.993) = 6.314$, p = .003, η_p^2 = .136).

Post-hoc pairwise comparisons between RVR conditions revealed that deviation from the localizer during rollout was significantly higher in 600 ft RVR scenarios than in 1200 ft RVR scenarios (p = .049, d = 0.145) and 4800 ft RVR scenarios (p = .009, d = 0.277). There was not a significant difference in deviation from the localizer during rollout between 1200 ft RVR and 4800 ft RVR conditions (p = .407). There was no significant interaction effect on RMS deviation from the localizer during rollout between display type and RVR (p = .677).

While there was a statistically significant increase in localizer deviation during rollout during 600 ft RVR scenarios, the small difference in mean values shown in Table 13 suggests that this effect is unlikely to be operationally significant. Figure 18 plots RMS deviation from the localizer during rollout in each experimental condition. As shown in the figures, outliers are present in the data; however, they appear to be spread consistently across the experimental conditions. Figure 19 plots the maximum deviation from the localizer during rollout in each experimental condition, demonstrating that trends in maximum deviation were roughly consistent in each condition; however, there is a visible pattern of greater maximum deviation values in scenarios with 600 ft RVR.

Figure 18
Box-and-Whisker Plot of Root Mean Square Deviation from Localizer During Rollout

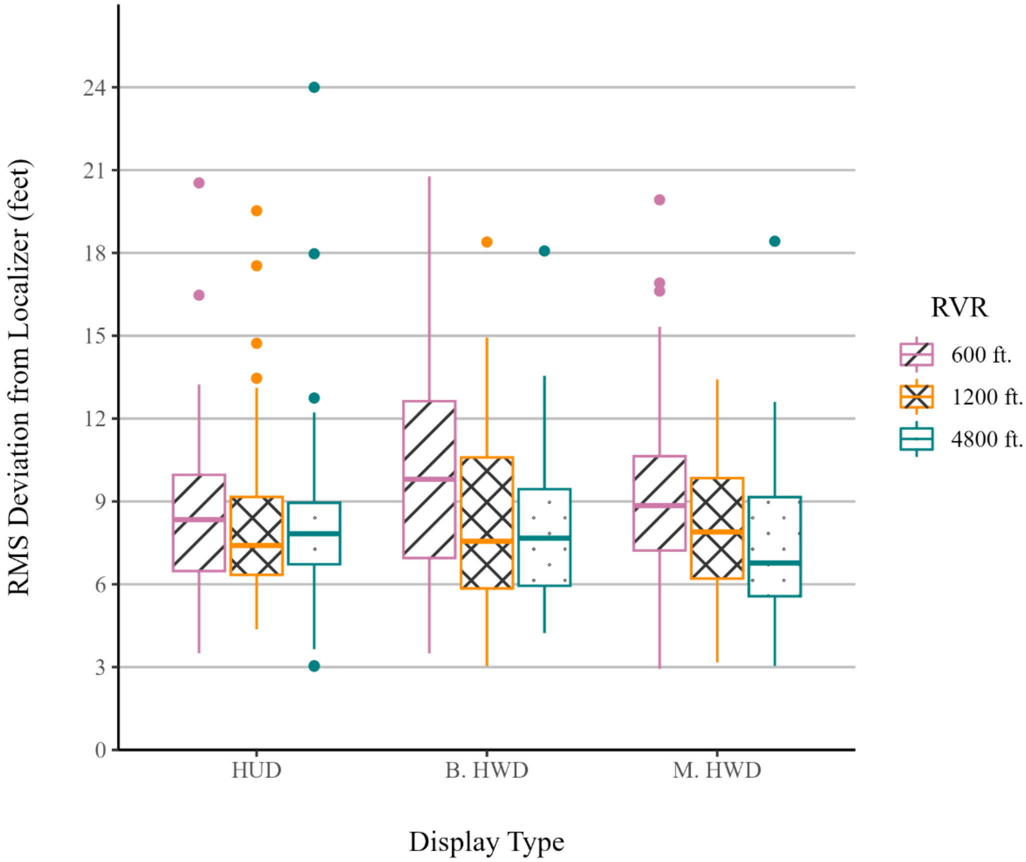


Figure 19

Frequency Bar Plot of Maximum Deviation from Localizer During Rollout

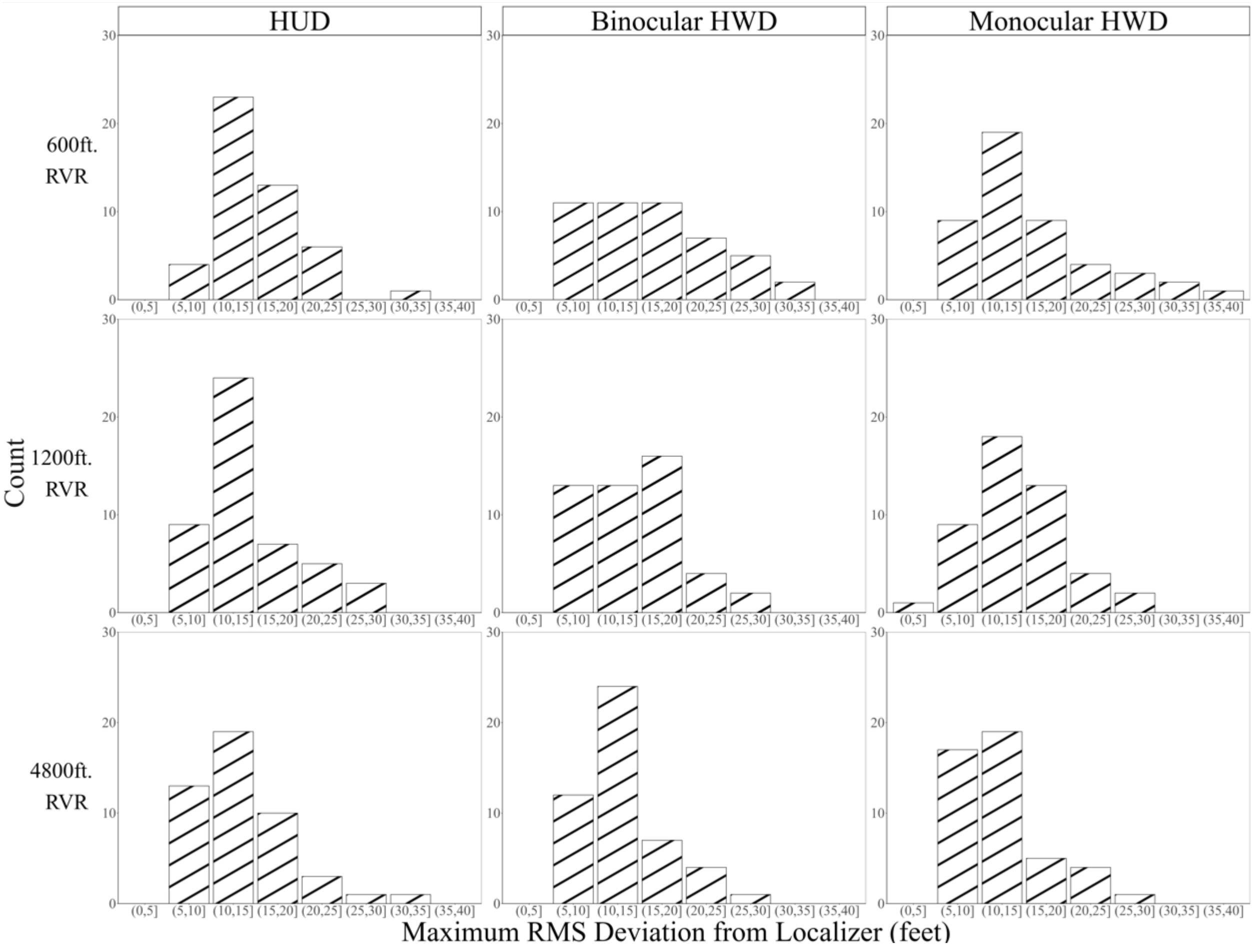


Table 13

Adjusted Means, Standard Errors, and Confidence Intervals for Root Mean Square Deviation from Localizer During Rollout in Each Experimental Condition

Display Type	RVR	M_{adj} (ft)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	600	8.74	0.52	7.68	9.80
	1200	8.30	0.52	7.24	9.35
	4800	8.25	0.60	7.04	9.46
B. HWD	600	9.78	0.67	8.43	11.14
	1200	8.45	0.55	7.33	9.57
	4800	8.16	0.45	7.26	9.07
M. HWD	600	9.26	0.56	8.14	10.39
	1200	8.18	0.44	7.30	9.06
	4800	7.49	0.49	6.51	8.48

Response to Non-Normal Events

Runway Incursion Scenarios

The speed with which the PF responded to runway incursions was evaluated as the distance of the aircraft center of gravity, in feet, from the runway touchdown aiming point at which the PF pressed the TO/GA button, marking the beginning of the missed approach procedure. Responses made at greater distances from the touchdown aiming point represented faster responses to the runway incursion.

Data from runway incursion scenarios were excluded from analysis if they met the following conditions: (1) the PF did not conduct a missed approach, continuing to land; and (2) the PM verbally called out the runway incursion to the PF, resulting in a missed approach. Data from these occurrences were excluded from analyses because (a) participants were instructed to not call out non-normal events when they were the PM, so these instances represented a failure to follow instructions and do not serve the research question at hand; and (b) if the PF did not conduct a missed approach, there is not a data point available to assess when they may have detected the runway incursion. Instances where the PF did not conduct a missed approach could also be due to a failure to follow the standard operating procedures outlined in the briefing, one of which was to conduct a missed approach if a non-normal event occurred.

Inspection of the data revealed that, among the 144 runway incursion scenarios, 13 (9.0%) involved a PF who did not conduct a missed approach. Inspection of the audio data and experimenter notes from these scenarios revealed that the PF did not detect the runway incursion in all but one of these scenarios. Among the 131 runway incursion scenarios in which a missed approach was conducted, the PM called out the incursion to the PF in 17 scenarios (13.0%). Therefore, data from 114 runway incursion scenarios (79.2%) were retained for analysis (see Table 13).

Table 13

Frequency of Landings versus Missed Approaches During Runway Incursion Scenarios for Each Display Type

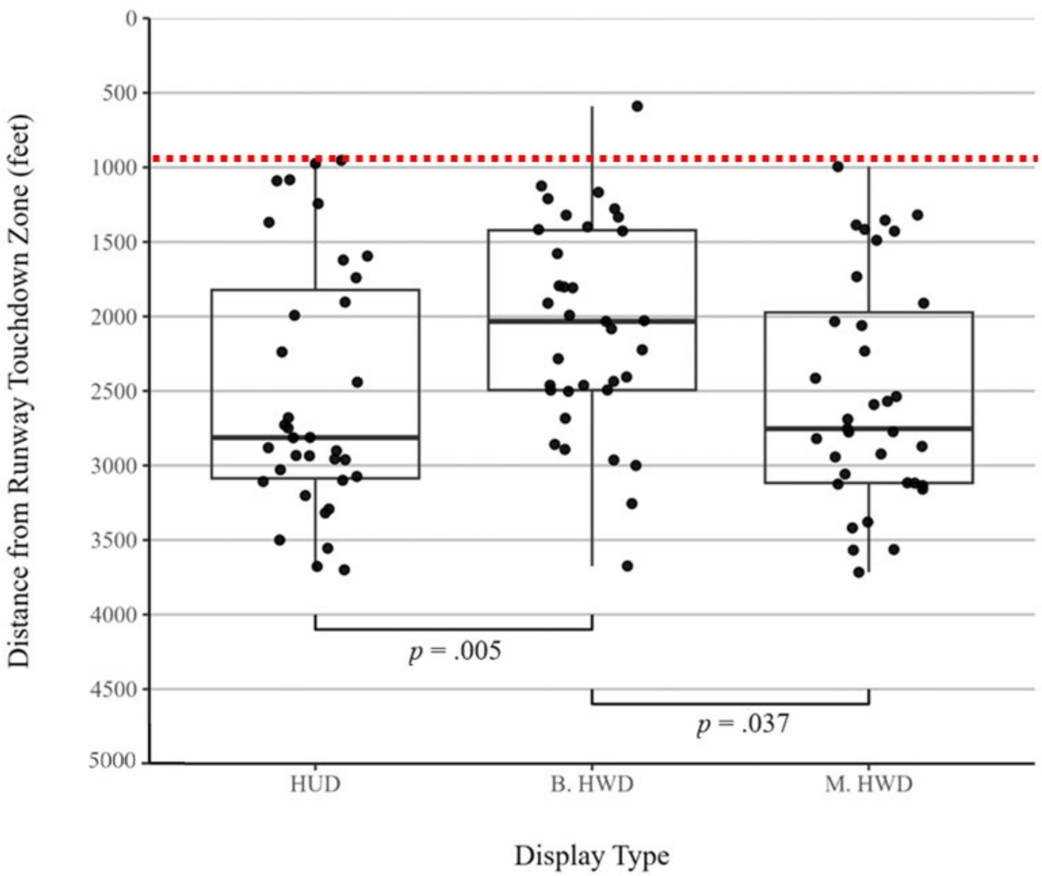
		Display Type			
		HUD	B. HWD	M. HWD	**Total**
All Runway Incursion Scenarios	*Landings*	*3*	*5*	*5*	13
	Missed Approaches	*45*	*43*	*43*	131
	Total	48	48	48	**144**
First Runway Incursion per Crew	*Landings*	*1*	*3*	*3*	7
	Missed Approaches	*7*	*5*	*5*	17
	Total	8	8	8	**24**

Missed approaches where the PM called out the incursion to the PF are included in the frequencies reported here.

Runway incursions only occurred in 4800 ft RVR scenarios, so response to runway incursions was evaluated as a function of display type using a one-way, repeated measures ANOVA. The ANOVA revealed a significant effect of display type on response to runway incursions, $F(1.806, 61.412) = 5.250$, $p = .010$, $\eta_p^2 = .134$. Post-hoc pairwise comparisons between display type conditions indicated that, on average, the PF was significantly delayed in responding to runway incursions when flying with a binocular HWD relative to when flying with a monocular HWD ($p = .037$, $d = 0.499$) and a HUD ($p = .005$, $d = 0.474$; see Table 14). There was not a significant difference in response to runway incursions between HUD and monocular HWD conditions ($p > .999$). Figure 20 presents the distance from the TDZ at which pilots pressed the TO/GA button in response to runway incursions. These patterns in the data are corroborated by responses to the usability questionnaire, in which participants reported that their ability to evaluate the safety of the runway environment was worse when flying with the binocular HWD than when flying with the monocular HWD.

Figure 20

Box-and-Whisker Plot of Aircraft Distance from Runway TDZ when the Pilot Flying Pressed the TO/GA Button in Response to Runway Incursions for Each Display Type



Data points that are plotted at greater distances from the runway touchdown aiming point represent earlier (i.e., faster) responses to the runway incursion. The horizontal dotted line represents the longitudinal location of the runway threshold (i.e., 996 feet before the runway touchdown aiming point).

Table 14

Adjusted Means, Standard Errors, and Confidence Intervals for Distance from Runway TDZ when the Pilot Flying Pressed the TO/GA Button in Response to Runway Incursions for Each Display Type

Display Type	M_{adj} (ft)	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
HUD	2518.54	141.78	2806.68	2230.40
B. HWD	2068.36	118.23	2308.62	1828.09
M. HWD	2525.08	128.01	2785.22	2264.94

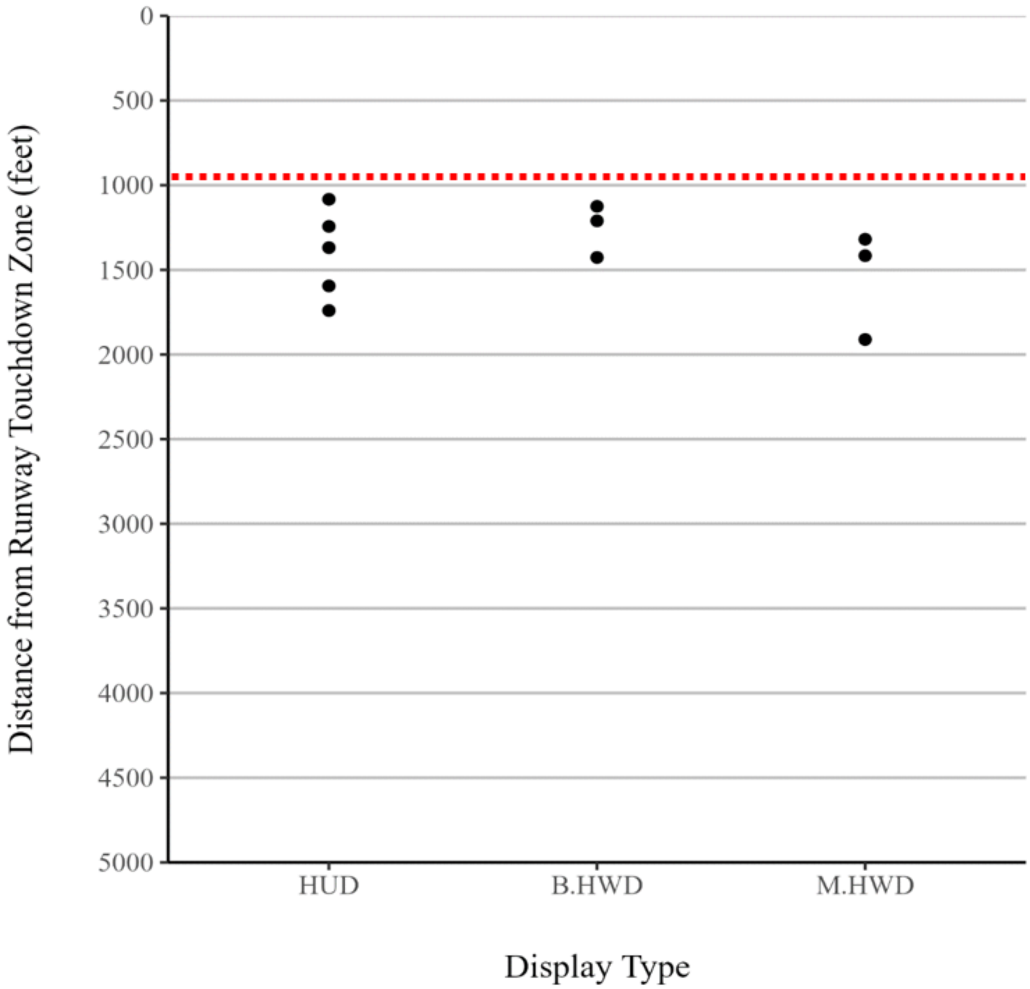
First Occurrence of a Runway Incursion per Crew

Because the first occurrence of the runway incursion was unexpected, and because event expectancy may influence responses to subsequent non-normal events (e.g., see Fadden et al., 1998), the ability of the PF to respond to the first, unexpected occurrence of the runway incursion for each crew was evaluated separately from subsequent occurrences. As documented in Table 13, there were numerous instances in each display type condition in which the PF continued to land during a runway incursion scenario. Analysis of simulator video and audio recordings confirmed that all but one of these cases were due to the PF failing to notice the runway incursion. Additionally, among the scenarios in which the PF conducted a missed approach during the first occurrence of a runway incursion, there were numerous instances in each display type condition where the PM called out the runway incursion to the PF.

Filtering these instances out from the data, leaving only missed approaches that the PF initiated themselves without a callout from the PM, leaves five data points in the HUD condition, three data points in the binocular HWD condition, and three data points in the monocular HWD condition. Due to the limited number of observations per experimental condition, it was not possible to analyze responses to the first occurrence of the runway incursion per crew using an inferential analysis. Figure 21 plots the distances from the runway TDZ at which the PF pressed the TO/GA button in response to the first runway incursion in each experimental session, due to the PF detecting the runway incursion and the PM, as instructed, refraining from calling out the incursion.

Figure 21

Strip Chart of Aircraft Distance from Runway TDZ when the PF Pressed the TO/GA Button in Response to the First Runway Incursion for Each Display Type



Data points that are plotted at greater distances from the runway touchdown aiming point represent earlier (i.e., faster) responses to the runway incursion. The horizontal dotted line represents the longitudinal location of the runway threshold (i.e., 996 feet before the runway touchdown aiming point).

Flight Director Failure Scenarios

The ability of the PF to respond to an FD failure was evaluated in terms of the time, in seconds, that elapsed between the point at which the simulator triggered the FD failure and the point at which the PF pressed the TO/GA button, marking the beginning of the missed approach procedure. It is important to note that the “APPROACH WARN” annunciation was suppressed during the FD failure. Data from FD failure scenarios were excluded from analysis if they met the following conditions: (1) the PF did not conduct a missed approach, continuing to land; or (2) the PM verbally called out the FD failure to the PF. Data from these occurrences were excluded from analyses because (a) participants were instructed to not call out non-normal events when they were the PM, so these instances would represent a failure to follow instructions and do not serve the research questions at hand; and (b) if the PF did not conduct a missed approach, there is not a data point available to assess when they may have detected the FD failure.

Inspection of the data revealed that, among the 144 FD failure scenarios, there were no cases in which the PM verbally called out the FD failure to the PF. There were 40 scenarios (27.8%) in which pilots did not conduct a missed approach, which were spread relatively evenly across the display type conditions (see Table 15). In the majority of these 40 scenarios, the PF detected the FD failure but decided to continue to a landing. When prompted to describe the decision-making process behind continuing to a landing, most pilots explained that they continued because the approach and threshold lighting became visible shortly after the failure occurred, and they could verify that the aircraft would land in the TDZ. Cases where a landing was completed after an FD failure were excluded prior to analysis of the PF response to the FD failure. After filtering out the data from these cases, data from 104 FD failure scenarios (72.2%) were retained for analysis

Table 15

Frequency of Landings versus Missed Approaches During FD Failure Scenarios for Each Display Type

		Display Type			
		HUD	B. HWD	M. HWD	**Total**
All FD Failure Scenarios	*Landings*	*13*	*16*	*11*	40
	Missed Approaches	*35*	*32*	*37*	104
	Total	48	48	48	**144**
First FD Failure per Participant	*Landings*	*6*	*6*	*4*	16
	Missed Approaches	*11*	*10*	*11*	32
	Total	17	16	15	**48**

FD failures occurred only in 600 ft RVR scenarios, so response time was evaluated as a function of display type using a one-way, repeated-measures ANOVA. The distribution of the raw data was positively skewed, so the data were subjected to a Box-Cox transformation ($\lambda = -0.15$), which successfully normalized the distribution. Analysis of the transformed data revealed no significant effect of display type on response to FD failures ($p = .354$). On average, the PF pressed the TO/GA button in response to a FD failure at roughly the same time, regardless of whether they were flying with a HUD, binocular HWD, or monocular HWD (see Table 16). Figure 22 presents the response to FD failures in each experimental condition.

Figure 22

Box-and-Whisker Plot of Elapsed Time from the Onset of a FD Failure to when the PF Pressed the TO/GA Button for Each Display Type

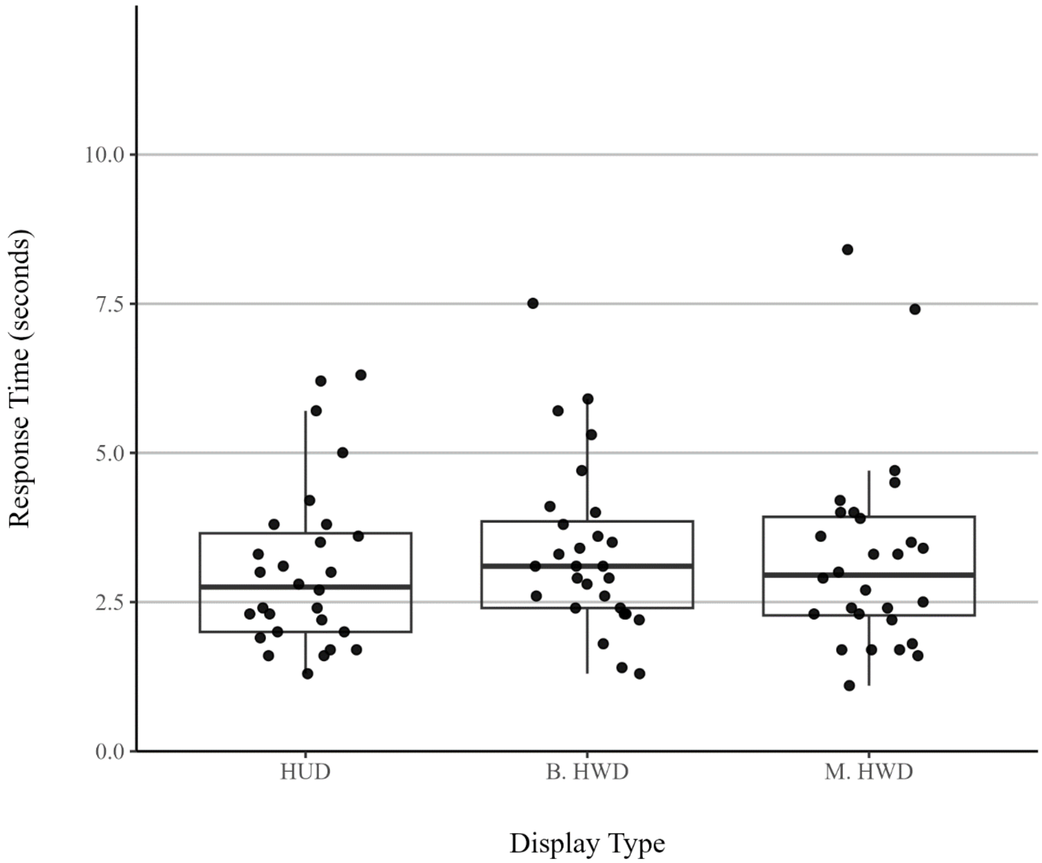


Table 16

Adjusted Means, Standard Errors, and Confidence Intervals for Elapsed Time from the Onset of a FD Failure to when the Pilot Flying Pressed the TO/GA Button for Each Display Type

Display Type	M_{adj} (seconds)	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
HUD	3.05	0.26	2.52	3.59
B. HWD	3.36	0.27	2.81	3.91
M. HWD	3.24	0.31	2.60	3.87

First Occurrence of a Flight Director Failure per Pilot

Response to the first, unexpected occurrence of the FD failure as a function of display type was evaluated separately from all occurrences using a one-way, between-subjects ANOVA. The distribution of the raw data was positively skewed, so the data were subjected to a Box-Cox transformation ($\lambda = 0.04$), which successfully normalized the distribution. Analysis of the transformed data revealed that there was no significant effect of display type on response to the first, unexpected FD failure ($p = .340$; see Table 17). Figure 23 presents the response to each pilot's first FD failure in each experimental condition.

Figure 23

Box-and-Whisker Plot of Elapsed Time from the Onset of the First FD Failure to when the Pilot Flying Pressed the TO/GA Button for Each Display Type

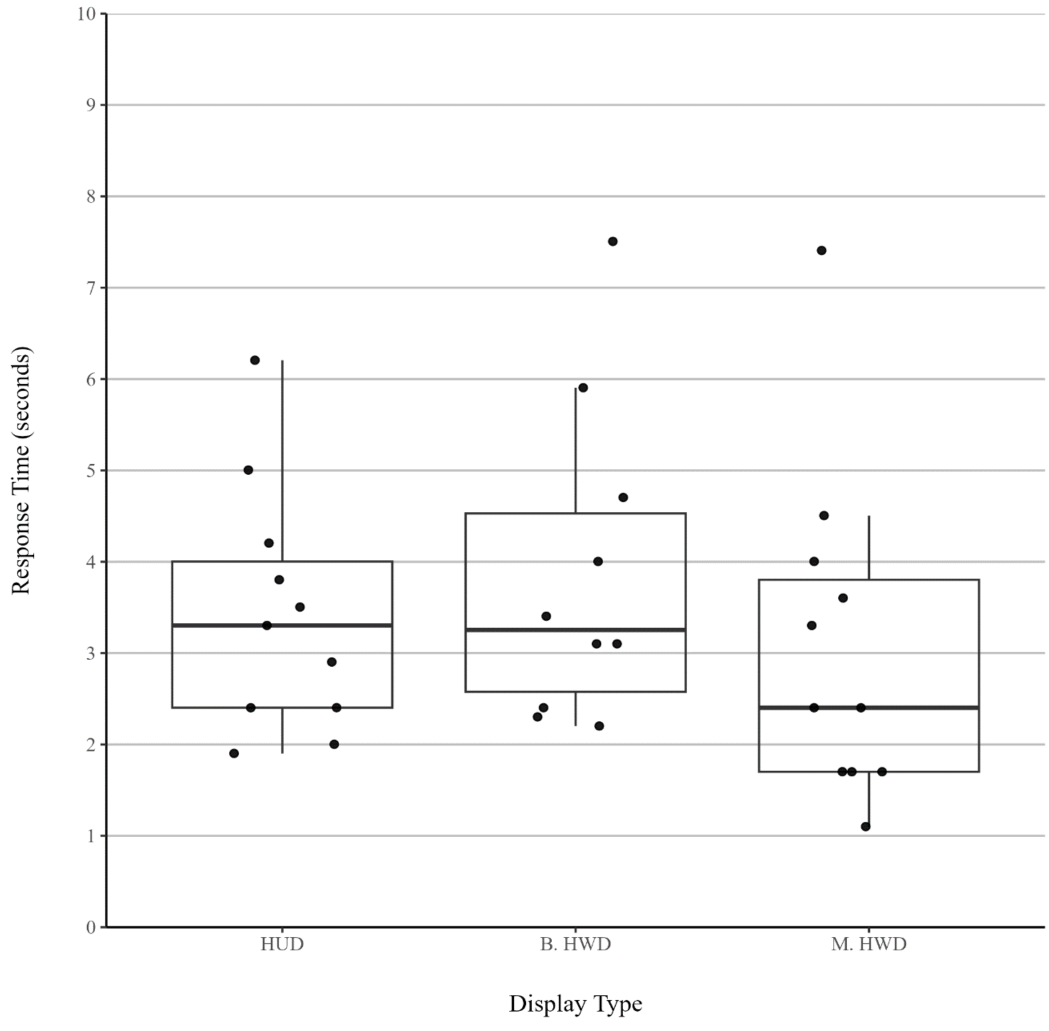


Table 17

Adjusted Means, Standard Errors, and Confidence Intervals for Response to First FD Failure for Each Display Type

Display Type	M_{adj} (seconds)	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
HUD	3.42	0.49	2.42	4.43
B. HWD	3.86	0.52	2.81	4.92
M. HWD	3.08	0.49	2.07	4.08

Pilot Flying Workload

As noted earlier in this report, a simulator issue occurred during data collection for 3 of the 24 crews (12.5% of all data). Inspection of the data revealed that this issue impacted flightpath symbology during the first 10 seconds of each scenario for these crews and negatively impacted their flightpath tracking and airspeed management performance for the first 60 seconds of the approach. Because of the negative impact observed with pilot performance due to this issue, it is possible that the issue may have influenced pilot workload in these scenarios. As a result, the NASA-TLX data for these three crews were excluded from analyses, resulting in data from 21 of the 24 crews (87.5%) being retained for analyses of NASA-TLX data. Prior to analyses, a NASA-TLX total weighted score was computed for each scenario using the procedure described by Hart and Staveland (1988).

NASA-TLX Total Weighted Score During Normal Scenarios

NASA-TLX total weighted score for the PF during normal scenarios was analyzed using a 3 (display type) × 3 (RVR) repeated-measures ANOVA. Normal scenarios in which participants conducted a missed approach while they were flying were excluded from analyses of NASA-TLX scores. Among the data for 21 crews that were retained for analyses, six missed approaches occurred in 378 normal scenarios (1.59%). Thus, NASA-TLX data from 371 normal scenarios (98.41%) were retained for analyses.

The ANOVA revealed that there was a significant main effect of display type on NASA-TLX total weighted score, $F(1.434, 50.174) = 34.120$, $p < .001$, $\eta_p^2 = .494$. Post-hoc pairwise comparisons between display type conditions revealed that the NASA-TLX total weighted score was significantly higher with a monocular HWD than with a binocular HWD ($p < .001$, $d = 0.377$) or HUD ($p < .001$, $d = 0.606$). The NASA-TLX total weighted score was also significantly higher with a binocular HWD than with a HUD ($p < .001$, $d = 0.410$; see Table 18). There was not a significant main effect of RVR on NASA-TLX total weighted score ($p = .169$), nor was there a significant interaction between display type and RVR on NASA-TLX total weighted score ($p = .518$). Figure 24 presents the NASA-TLX total weighted score in each experimental condition. The NASA-TLX data were corroborated by participants' written feedback on the usability questionnaire, where many participants reported that flying with the HWD—particularly the monocular HWD—required more effort and concentration than flying with the HUD (see Table 29).

Figure 24

Box-and-Whisker Plot of NASA-TLX Total Weighted Score during Normal Scenarios in Each Experimental Condition

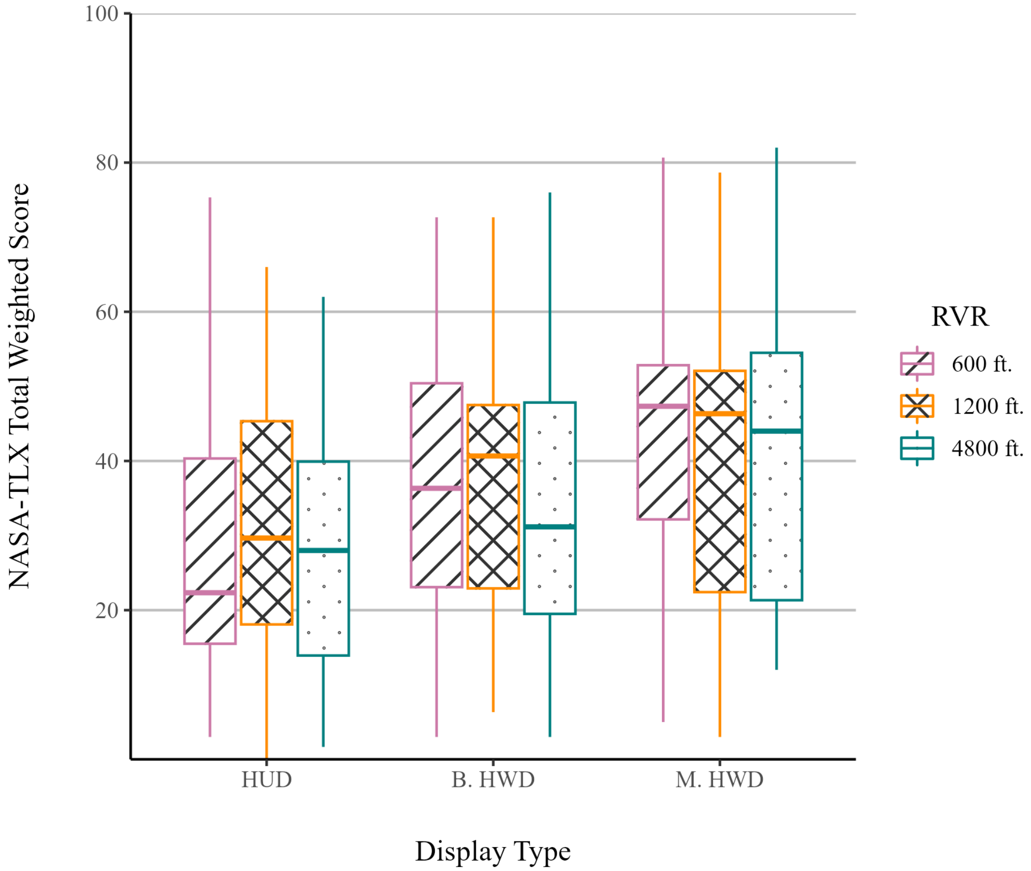


Table 18

Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Total Weighted Score during Normal Scenarios in Each Experimental Condition

Display Type	RVR (ft)	M_{adj} (scale = 0-100)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	600	28.45	2.81	22.76	34.15
	1200	30.33	2.94	24.36	36.31
	4800	28.44	2.60	23.16	33.73
B. HWD	600	36.68	2.93	30.72	42.63
	1200	34.96	2.81	29.25	40.67
	4800	33.64	3.16	27.23	40.05
M. HWD	600	43.10	3.19	36.62	49.58
	1200	40.39	3.22	33.85	46.92
	4800	39.19	3.22	32.65	45.72

NASA-TLX Subscale Scores During Normal Scenarios

NASA-TLX subscale scores for the PF during normal scenarios were analyzed using a 3 (display type) x 3 (RVR) repeated-measures Multivariate Analysis of Variance (MANOVA). The DVs of the MANOVA were the six NASA-TLX subscales: (1) Mental Demand, (2) Physical Demand, (3) Temporal Demand, (4) Performance, (5) Effort, and (6) Frustration. The distribution of the raw data was positively skewed, so the data were subjected to a Box-Cox transformation ($\lambda = 0.45$), which successfully normalized the distribution. Analysis of the transformed data revealed a statistically significant main effect of display type on all six NASA-TLX subscales, Wilks's $\Lambda = .421$, $F(12, 24) = 5.859$, $p < .001$, $\eta_p^2 = .351$. There was not a significant main effect of RVR on the six NASA-TLX subscales ($p = .362$), nor was there a significant interaction effect of display type and RVR ($p = .816$).

Bonferroni-corrected post-hoc ANOVAs were conducted as follow-up tests to the MANOVA to further break down the effect of display type and RVR on NASA-TLX subscale scores. The ANOVAs revealed that there was a significant main effect of display type on all of the NASA-TLX subscale scores (see Table 19). Mean NASA-TLX subscale scores for each display type shown in Table 20 indicate that the monocular HWD resulted in the highest scores and the HUD resulted in the lowest scores for all of the NASA-TLX subscales. Table 21 presents descriptive statistics for NASA-TLX subscale scores in each RVR condition, which indicate that NASA-TLX subscale scores were consistent across all subscales, with small, non-significant trends in the means toward higher scores in the 600 ft RVR scenarios.

Table 19

Results from Post-Hoc ANOVAs

Within Subjects Effect	Dependent Variable	DF	*F*	*p*	η_p^2
Display Type	Mental Demand	1.545, 54.060	30.123	< .001*	.463
	Physical Demand	1.844, 64.546	23.152	< .001*	.398
	Temporal Demand	1.622, 56.767	19.533	< .001*	.398
	Performance	1.594, 55.794	7.049	.022*	.168
	Effort	1.696, 59.374	27.261	< .001*	.418
	Frustration	1.363, 52.458	25.711	< .001*	.438
RVR	Mental Demand	1.763, 61.695	2.004	.892	.054
	Physical Demand	1.633, 57.143	0.287	> .999	.008
	Temporal Demand	1.983, 69.401	2.409	.587	.064
	Performance	1.877, 65.697	0.077	> .999	.002
	Effort	1.696, 59.368	2.140	.803	.058
	Frustration	1.715, 60.036	0.528	> .999	.015
Display Type x RVR	Mental Demand	3.410, 119.336	0.721	> .999	.020
	Physical Demand	3.406, 119.215	0.302	> .999	.009
	Temporal Demand	3.638, 127.335	0.805	> .999	.022
	Performance	2.919, 102.179	0.413	> .999	.012
	Effort	3.793, 132.745	0.041	> .999	.001
	Frustration	3.206, 112.218	0.910	> .999	.025

** = significant at $\alpha = .05$ after Bonferroni correction.*

Table 20

Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Subscale Scores during Normal Scenarios in Each Display Type Condition

Subscale	Display Type	M_{adj} (scale: 0-20)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
Mental Demand	HUD	6.27	0.60	5.04	7.49
	B. HWD	7.70	0.65	6.39	9.02
	M. HWD	8.88	0.67	7.52	10.24
Physical Demand	HUD	5.56	0.56	4.42	6.69
	B. HWD	6.68	0.62	5.42	7.93
	M. HWD	7.92	0.67	6.55	9.28
Temporal Demand	HUD	4.90	0.57	3.74	6.06
	B. HWD	6.19	0.67	4.82	7.55
	M. HWD	7.21	0.67	5.85	8.58
Performance	HUD	5.11	0.60	3.9	6.32
	B. HWD	5.47	0.55	4.37	6.58
	M. HWD	5.90	0.56	4.77	7.03
Effort	HUD	5.63	0.53	4.55	6.71
	B. HWD	7.00	0.55	5.89	8.11
	M. HWD	8.45	0.62	7.20	9.71
Frustration	HUD	3.40	0.54	2.30	4.50
	B. HWD	4.82	0.59	3.62	6.01
	M. HWD	5.62	0.65	4.31	6.93

Table 21

Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Subscale Scores During Normal Scenarios in Each Runway Visual Range Condition

Subscale	RVR (ft)	M_{adj} (scale: 0-20)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
Mental Demand	600	7.93	0.64	6.63	9.23
	1200	7.62	0.63	6.35	8.90
	4800	7.31	0.65	5.98	8.63
Physical Demand	600	6.82	0.64	5.54	8.11
	1200	6.71	0.60	5.50	7.92
	4800	6.61	0.60	5.40	7.83
Temporal Demand	600	6.40	0.68	5.02	7.78
	1200	6.19	0.61	4.96	7.42
	4800	5.71	0.58	4.54	6.89
Performance	600	5.47	0.56	4.33	6.61
	1200	5.59	0.59	4.40	6.79
	4800	5.42	0.56	4.28	6.56
Effort	600	7.40	0.55	6.28	8.52
	1200	7.03	0.55	5.91	8.15
	4800	6.66	0.57	5.51	7.81
Frustration	600	4.46	0.58	3.29	5.64
	1200	4.76	0.57	3.59	5.92
	4800	4.61	0.62	3.36	5.87

NASA-TLX Total Weighted Score During Runway Incursion Scenarios

NASA-TLX total weighted score for the PF during runway incursion scenarios was analyzed as a function of display type using a one-way, repeated-measures ANOVA. Runway incursion scenarios in which participants did not conduct a missed approach while they were flying, and continued to land, were excluded from analyses of NASA-TLX data. Among the data for 21 crews that were retained for all NASA-TLX analyses, 12 landings occurred in the 126 runway incursion scenarios (9.52% of all scenarios). Thus, NASA-TLX data from 114 runway incursion scenarios (90.48%) were retained for analyses. The distribution of the raw data was positively skewed, so the data were subjected to a Box-Cox transformation ($\lambda = 0.54$), which successfully normalized the distribution.

Analysis of the transformed data revealed that there was a significant effect of display type on NASA-TLX total weighted score, $F(1.696, 54.270) = 8.315$, $p < .001$, $\eta_p^2 = .206$. Post-hoc pairwise comparisons reveal that the NASA-TLX total weighted score was significantly lower during runway incursion scenarios when the PF used a HUD compared to when they used a binocular HWD ($p = .003$, $d = 0.567$; see Table 22) or monocular HWD during those scenarios ($p = .020$, $d = 0.501$). There was not a significant difference in the NASA-TLX score during runway incursion scenarios between binocular HWD and monocular HWD conditions ($p > .999$).

Figure 25 presents the NASA-TLX total weighted score during runway incursion scenarios in each experimental condition.

Figure 25

Box-and-Whisker Plot of NASA-TLX Total Weighted Score During Runway Incursion Scenarios for Each Display Type

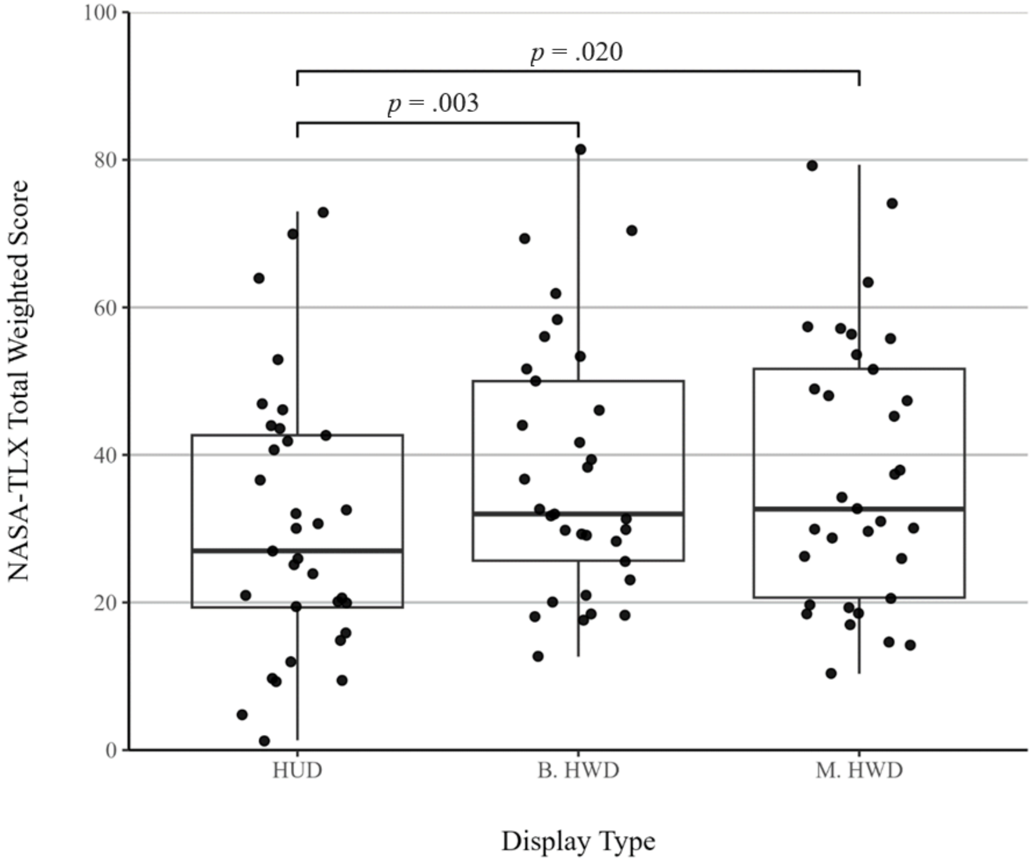


Table 22

Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Total Weighted Score During Runway Incursion Scenarios for Each Display Type

Display Type	M_{adj} (scale: 0-100)	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
HUD	30.56	3.16	24.11	37.00
B. HWD	37.79	3.02	31.63	43.95
M. HWD	37.42	3.16	30.99	43.86

NASA-TLX Subscale Scores During Runway Incursion Scenarios

NASA-TLX subscale scores for the PF during runway incursion scenarios were analyzed as a function of display type using a one-way, repeated-measures MANOVA. The DVs of the MANOVA were the six NASA-TLX subscales: (1) Mental Demand, (2) Physical Demand, (3) Temporal Demand, (4) Performance, (5) Effort, and (6) Frustration. The distribution of the raw data was positively skewed, so the data were subjected to a Box-Cox transformation ($\lambda = 0.37$), which successfully normalized the distribution. Analysis of the transformed data revealed that there was a significant main effect of display type on the six NASA-TLX subscales, Wilks's Λ = .639, $F(12, 118) = 2.473$, $p = .006$, $\eta_p^2 = .201$.

Bonferroni-corrected post-hoc ANOVAs were conducted as follow-up tests to the MANOVA to further break down the effect of display type on NASA-TLX subscale scores during runway incursion scenarios. The ANOVAs revealed that there was a significant effect of display type on Performance and Effort ($p = .003$ and $p = .004$, respectively). There was not a significant effect of display type on any of the other subscales ($p > .05$, in each case; see Table 23).

Table 24 presents mean subscale scores during runway incursion scenarios for each display type. Trends among the means indicate that scores on the Performance subscale were higher (i.e., pilots rated their own performance as poorer) when pilots flew with a binocular HWD than when they flew with a monocular HWD or HUD during runway incursion scenarios. Mean subscale scores indicate that scores on the Performance subscale were roughly equivalent between the monocular HWD and HUD conditions but were elevated during the binocular HWD condition. Trends among means indicate that Effort was highest when pilots flew with a monocular HWD, second highest when they flew with a binocular HWD, and lowest when they flew with a HUD during runway incursion scenarios.

Table 23

Results from Post-Hoc ANOVAs

Within Subjects Effect	DV	DF	*F*	*p*	η_p^2
Display Type	Mental Demand	1.841, 58.916	4.351	.118	.120
	Physical Demand	1.859, 59.488	3.168	.316	.090
	Temporal Demand	1.752, 56.063	2.293	.700	.067
	Performance	1.970, 63.024	8.793	.003*	.216
	Effort	1.864, 59.663	8.790	.004*	.215
	Frustration	1.777, 56.859	2.205	.750	.064

** = significant at α = .05 after Bonferroni correction.*

Table 24

Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Subscale Score during Runway Incursion Scenarios in Each Experimental Condition

Subscale	Display Type	M_{adj} (scale: 0-20)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
Mental Demand	HUD	6.73	0.71	5.29	8.17
	B. HWD	7.73	0.71	6.28	9.17
	M. HWD	7.97	0.77	6.39	9.55
Physical Demand	HUD	5.46	0.63	4.17	6.74
	B. HWD	6.46	0.69	5.05	7.86
	M. HWD	6.55	0.71	5.10	7.99
Temporal Demand	HUD	5.82	0.77	4.25	7.39
	B. HWD	6.49	0.77	4.92	8.05
	M. HWD	6.49	0.68	5.11	7.86
Performance	HUD	5.24	0.71	3.79	6.70
	B. HWD	7.27	0.76	5.72	8.83
	M. HWD	6.55	0.73	5.07	8.03
Effort	HUD	5.61	0.66	4.27	6.95
	B. HWD	6.79	0.67	5.43	8.14
	M. HWD	7.79	0.79	6.18	9.40
Frustration	HUD	4.42	0.79	2.81	6.04
	B. HWD	5.36	0.82	3.70	7.03
	M. HWD	4.46	0.57	3.30	5.61

NASA-TLX Total Weighted Score During Flight Director Failure Scenarios

NASA-TLX total weighted score for the PF during FD failure scenarios was analyzed as a function of display type using a one-way, repeated-measures ANOVA. FD failure scenarios in which participants did not conduct a missed approach while they were flying, and continued to land, were excluded from analyses of NASA-TLX data. Among the data for 21 crews that were retained for all NASA-TLX analyses, 40 landings occurred in 126 FD failure scenarios (31.75% of all scenarios). Thus, NASA-TLX data from 86 FD failure scenarios (68.25%) were retained for analyses. The distribution of the raw data was positively skewed, so the data were subjected to a Box-Cox transformation ($\lambda = 0.69$), which successfully normalized the distribution.

Analysis of the transformed data revealed that there was a significant effect of display type on NASA-TLX total weighted score, $F(1.410, 29.617) = 8.346$, $p = .003$, $\eta_p^2 = .284$. Post-hoc pairwise comparisons reveal that the NASA-TLX total weighted score was significantly lower during FD failure scenarios when the PF used a HUD compared to when they used a binocular HWD ($p = .025$, $d = 0.688$) or monocular HWD ($p = .009$, $d = 0.558$) during those scenarios (see Table 25). There was not a significant difference in NASA-TLX total weighted score during FD failure scenarios between binocular HWD and monocular HWD conditions ($p > .999$). Figure 26 presents the NASA-TLX total weighted score during FD failure scenarios in each experimental condition.

Figure 26

Box-and-Whisker Plot of NASA-TLX Total Weighted Score During FD Failure Scenarios for Each Display Type

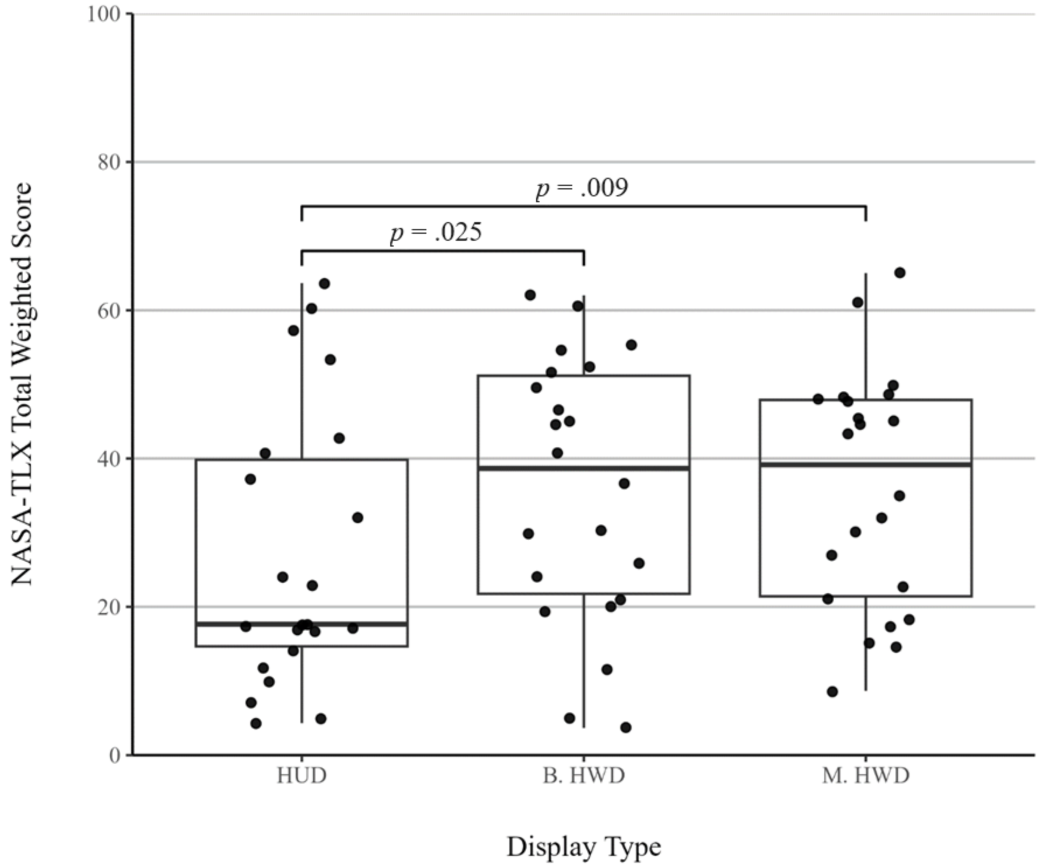


Table 25

Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Total Weighted Score During FD Failure Scenarios for Each Display Type

Display Type	M_{adj} (scale: 0-100)	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
HUD	26.80	3.98	18.52	35.08
B. HWD	35.96	3.80	28.05	43.86
M. HWD	35.85	3.42	28.74	42.95

NASA-TLX Subscale Scores During Flight Director Failure Scenarios

NASA-TLX subscale scores for the PF during FD failure scenarios were analyzed as a function of display type using a one-way, repeated-measures MANOVA. The DVs of the MANOVA were the six NASA-TLX subscales: (1) Mental Demand, (2) Physical Demand, (3) Temporal Demand, (4) Performance, (5) Effort, and (6) Frustration. The distribution of the raw data was positively skewed, so the data were subjected to a Box-Cox transformation ($\lambda = 0.45$), which successfully normalized the distribution. Analysis of the transformed data revealed that there was not a significant main effect of display type on the six NASA-TLX subscale scores, Wilks's $\Lambda = .599$, $F(12, 72) = 1.801$, $p = .064$, $\eta_p^2 = .226$.

Bonferroni-corrected post-hoc ANOVAs were conducted as follow-up tests to the MANOVA to further break down the effect of display type on NASA-TLX subscale scores during FD failure scenarios. The ANOVAs revealed that there was a significant effect of display type on Mental Demand, Physical Demand, Effort, and Frustration; there was not a significant effect of display type on any of the other subscales ($p > .05$, in each case; see Table 26).

Table 27 presents mean subscale scores during FD failure scenarios for each display type. Trends among the means indicate that Mental Demand, Physical Demand, Effort, and Frustration were higher when pilots used a binocular or monocular HWD than when they used a HUD during FD Failures scenarios. Effort was highest when pilots used a monocular HWD, second highest when they used a binocular HWD, and lowest when they used a HUD during FD failure scenarios.

Table 26

Results from Post-Hoc ANOVAs

Within Subjects Effect	DV	DF	*F*	*p*	η_p^2
Display Type	Mental Demand	1.867, 39.201	7.864	.010*	.272
	Physical Demand	1.395, 29.296	8.121	.024*	.279
	Temporal Demand	1.982, 41.616	5.129	.062	.196
	Performance	1.670, 35.072	1.512	> .999	.067
	Effort	1.865, 39.161	7.913	.010*	.274
	Frustration	1.960, 41.167	7.144	.014*	.254

** = significant at α = .05 after Bonferroni correction.*

Table 27

Adjusted Means, Standard Errors, and Confidence Intervals for NASA-TLX Subscale Scores during Runway Incursion Scenarios in Each Experimental Condition

Subscale	Display Type	M_{adj} (scale: 0-20)	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
Mental Demand	HUD	5.73	0.92	3.82	7.63
	B. HWD	8.18	1.03	6.05	10.32
	M. HWD	7.82	0.85	6.04	9.60
Physical Demand	HUD	5.14	0.87	3.32	6.95
	B. HWD	6.96	0.89	5.10	8.81
	M. HWD	7.09	0.89	5.24	8.94
Temporal Demand	HUD	4.73	0.84	2.97	6.48
	B. HWD	6.23	0.82	4.51	7.94
	M. HWD	6.14	0.72	4.64	7.64
Performance	HUD	4.73	0.80	3.06	6.40
	B. HWD	5.46	0.71	3.97	6.94
	M. HWD	5.55	0.62	4.26	6.83
Effort	HUD	5.68	0.92	3.78	7.59
	B. HWD	7.32	0.84	5.58	9.06
	M. HWD	8.27	0.89	6.42	10.13
Frustration	HUD	3.41	0.73	1.90	4.92
	B. HWD	5.14	0.97	3.12	7.15
	M. HWD	5.46	0.86	3.66	7.25

Usability Questionnaire

Each participant rated the usability of HUD relative to binocular HWD —and binocular HWD relative to monocular HWD—on a 5-point Likert scale ranging from 1 (“Much worse”) to 5 (“Much better”). Display usability was evaluated based on the participants’ own perception of their ability to:

1. Follow flightpath guidance symbology
2. Maintain target airspeed of the approach
3. Transition from instrument to visual flight references
4. Land the aircraft and follow rollout guidance
5. Evaluate the safety of the runway environment
6. Detect failures in the HUD or HWD symbology

The distribution of participants’ responses to each item on this questionnaire is detailed in Table 28 and Figure 27. In addition to the Likert scale items, the usability questionnaire gave participants the option to provide written feedback. Thirty-two participants (66.7%) elected to provide additional written feedback. The experimenter reviewed the responses and grouped them into categories based on themes. The most common themes that the experimenter identified among the responses included (a) visual perception, (b) workload, and (c) practice and learning effects. Highlights among the responses that fell into these thematic categories are listed in Table 28. Table 29 presents highlights among the written feedback that participants provided in the usability questionnaire.

Table 28

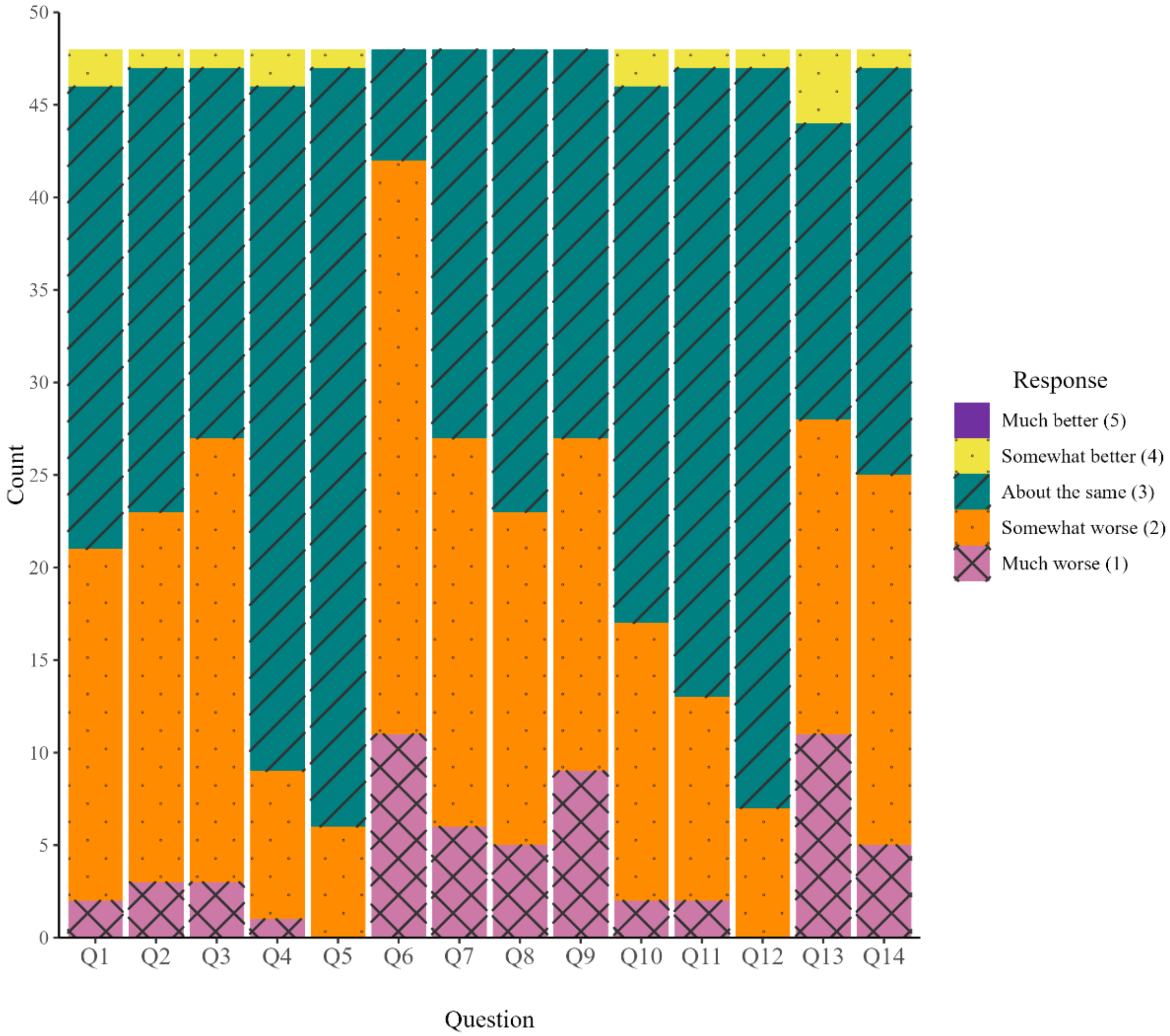
Descriptive Statistics for Usability Questionnaire Items

		Question	*M*	Median	*SD*	Min	Max
Compared to with the HUD	Q1	Ability to follow flightpath guidance symbology while flying with the binocular HWD:	2.563	3	0.649	1	4
	Q2	Ability to maintain target airspeed while flying with the binocular HWD:	2.479	3	0.652	1	4
	Q3	Ability to transition to visual references while flying with the binocular HWD:	2.396	2	0.644	1	4
	Q4	Ability to land the aircraft while flying with the binocular HWD:	2.833	3	0.519	1	4
	Q5	Ability to follow rollout guidance while flying with the binocular HWD:	2.896	3	0.371	2	4
	Q6	Ability to evaluate the safety of the runway environment while flying with the binocular HWD:	1.896	2	0.592	1	3
	Q7	Ability to detect failures in the symbology while flying with the binocular HWD:	2.313	2	0.689	1	3
Compared to with the binocular HWD	Q8	Ability to follow flightpath guidance symbology while flying with the monocular HWD:	2.417	3	0.679	1	3
	Q9	Ability to maintain target airspeed while flying with the monocular HWD:	2.250	2	0.758	1	3
	Q10	Ability to transition to visual references while flying with the monocular HWD:	2.646	3	0.635	1	4
	Q11	Ability to land the aircraft while flying with the monocular HWD:	2.708	3	0.582	1	4
	Q12	Ability to follow rollout guidance while flying with the monocular HWD:	2.875	3	0.393	2	4
	Q13	Ability to evaluate the safety of the runway environment while flying with the monocular HWD:	2.271	2	0.917	1	4
	Q14	Ability to detect failures in the symbology while flying with the monocular HWD:	2.396	2	0.707	1	4

1 = Much Worse; 2 = Somewhat Worse; 3 = About the Same; 4 = Somewhat Better; 5 = Much Better

Figure 27

Stacked Bar Chart of Usability Questionnaire Responses



Refer to Table 28 for the prompt question associated with each item.

Table 29

Highlighted Written Feedback on Participants' Experiences with the Head-Up Display and Head-Worn Display

Visual Perception
I experienced symbology blanking with monocular that I did not experience with binocular.
My field of vision and focus seemed to be diminished [with the monocular HWD].
The "soda straw" effect was more apparent monocular vs binocular.
While focusing on reference symbol and acceleration cue, depending what kind of wind correction is in, much of the peripheral info on the screen is not clearly visible [with the monocular HWD].
The ability to detect runway incursions and other unsafe runway conditions seemed to be much more difficult using the [HWD] due to the depth of focus and lack of a wide field of view while wearing the device.
In general, I felt like I was looking through a soda straw [with the HWD]. I had less SA, less perception of airspeed deviations, and a harder time cross checking cockpit instruments and the runway environment.
Monocular presents a significant deterioration of cues.
Workload
The effort required to maintain aircraft control was increased [with monocular HWD] which made it more difficult keep the workload manageable.
The monocular HWD required more concentration, mental effort, and eye movement to get the information I needed.
The monocular HWD was more physically demanding versus the binocular or the HUD. It created additional workload when transitioning from instrument to visual exactly at the most demanding time in the approach and landing.
Overall, monocular was harder to use. I felt like my brain had to work harder to interpret information.
I personally found using the [binocular HWD] is much easier than [using the monocular HWD]. This could be due to focal length or just the inability of my brain to process information that is only presented in one eye vs two.
Harder to see everything. Really had to work to fly those approaches with the [monocular HWD].
Both HWD[s] caused higher workload and distraction due to the frame[s] also blocked importan[t] field of view.
Practice and Learning Effects
With practice, the binocular HWD was about the same overall experience as using a HUD.
Overall flying off the HWD took some getting used to, however, I'm fairly confident given a much better HWD, and a few hours in a simulator, I could perform about as equal as I do on a HUD.
I started off and still do favor the binocular version of the HWD, however, at the end of the study, I was getting a touch more comfortable with the mono[cular] version. Practice makes perfect.
There was a learning curve, but near the end of the session my experience between the HUD and binocular HWD was nearly identical.

Subjective Evaluation of Dominant Versus Non-Dominant Eye with Monocular Head-Worn Display

Simulator audio data and experimenter notes from the familiarization session were compiled and reviewed to assess participants' subjective feedback on the use of the monocular HWD and whether they experienced noticeable differences in the symbology appearance, FOV, or functionality between left and right eye presentation of the HWD symbology. The results of this evaluation indicate that alignment eye dominance may not be a primary factor that determines participants' eye preference when flying with a monocular HWD. When asked whether they preferred the monocular HWD on one eye over the other, nearly all participants indicated that they had no preference.

Of the participants who did prefer one eye over the other, they elaborated by describing that their preference was driven by other factors besides eye dominance, including interocular differences in visual acuity. For example, some participants reported possessing better far vision in one eye over the other, and that they would prefer to view the symbology on that eye. Some participants indicated that they were more easily able to see information on the head-down instrument panel when the monocular HWD was configured for the left eye. The basis for this preference was that the lack of visual interference from the symbology on the right eye made it easier to scan the instruments. However, this may have been a byproduct of the specific optical characteristics of the Microsoft HoloLens 2 and may not be representative of all HWDs designed specifically for use on the flight deck.

Discussion

The present study was carried out to determine whether pilot performance and workload during manually flown, low-visibility approach and landing operations are impacted if flight guidance is presented to the PF on a monocular or binocular HWD in lieu of a HUD. The basis for this investigation is that the HWD is worn by the pilot, possesses unique physical and optical characteristics such as a near-to-eye combiner, and can be configured as a monocular or binocular display. These characteristics are different from those of a HUD that is installed on the flight deck and thus may impact pilot performance and workload during low-visibility approach and landing operations. These characteristics may also impact a pilot's ability to respond to non-normal events, including flight guidance failures and runway incursions. In addition, the HWD used in this study was a non-collimated device that presented symbology at a focal distance of six feet, whereas the HUD was a collimated, production-quality system that presented symbology at optical infinity. The non-collimated HWD used in this research may impact a pilot's flying performance during segments of the flight where external visual references are used, and the symbology is superimposed onto those references. The following section summarizes the findings from this research as they relate to the initial research questions and presents operational implications of the findings.

Findings

RQ_1: Do pilots demonstrate differences in flightpath tracking, airspeed management, landing, and rollout performance depending on whether they fly with a monocular HWD, binocular HWD, or HUD?

Across the analysis of pilot performance in this study, there were some statistically significant effects of display type on performance. However, the disparity in performance between conditions was very small, suggesting that differences across the HUD, binocular HWD, and monocular HWD in this research do not reach operational significance. During the approach, there was no difference in pilots' flightpath tracking performance as a function of display type; however, pilots were somewhat less able to maintain the target airspeed of the approach when they flew with a monocular HWD compared to when they flew with a binocular HWD or a HUD. When crossing the runway threshold, airspeed deviation in scenarios with the monocular HWD was 1.18 knots higher than it was with the binocular HWD and 1.16 knots higher than it was with the HUD, indicating that these effects are not operationally significant. These patterns in performance data are reflected in participants' written feedback on the usability questionnaire, which included comments about increased difficulty in viewing peripheral symbology, such as airspeed information, when flying with the monocular HWD.

Pilots also tended to cross the threshold somewhat lower across the threshold and land sooner and harder when flying with the binocular and monocular HWD compared to when flying with the HUD; however, these differences were small and, as with the differences in airspeed deviation, did not suggest an operationally significant performance concern. Furthermore, there was no difference in lateral or longitudinal touchdown location, sink rate at touchdown, or localizer deviation at rollout between monocular HWD and binocular HWD scenarios.

In conclusion, there are findings in this study that are consistent with laboratory research demonstrating that binocular rivalry occurs when using a monocular display, subsequently hindering the ability to use task-critical visual information presented on the display (Patterson et al., 2006; Winterbottom et al., 2006b). However, it appears that these effects do not translate to operationally significant performance concerns when pilots manually fly approach and landing operations using symbology on a monocular or binocular HWD.

RQ_2: Does flight visibility impact pilots' flightpath tracking, airspeed management, landing, and rollout performance when flying with a HUD, binocular HWD, and monocular HWD?

This research found that decreases in flight visibility resulted in statistically significant increases in localizer deviation at threshold crossing, lateral deviation from the runway centerline at landing, and localizer deviation during rollout. This effect was largely consistent regardless of whether the pilot flew with a monocular HWD, a binocular HWD, or a HUD. That is, reductions in runway visibility had corresponding negative impacts on pilot performance immediately before, during, and after landing, regardless of the display type in use. This finding is supported by previous research, which demonstrates that the ability to land on the runway centerline degrades as RVR decreases, even when pilots use flight symbology on a HUD (Boucek et al., 1983; Newton et al., forthcoming).

The findings herein align with this previous research and build on it by demonstrating that this effect is consistent regardless of whether the pilot flies using symbology on a HUD, binocular HWD, or monocular HWD. This effect may primarily be driven by a decreased ability to see runway visual information that communicates the aircraft position relative to the runway TDZ, such as the approach lighting system, TDZ lighting, runway edge lighting, and runway centerline paint markings and lighting. Taken together, this suggests that pilots rely on natural visual cues about the runway environment to laterally align the aircraft with the runway as they land, and that the flight guidance symbology is not the sole information source during this segment of the operation.

Whereas RVR affected lateral deviation from the runway centerline at landing, it did not appear to impact the longitudinal touchdown point or the pilot's ability to manage the sink rate of the aircraft prior to touchdown; these two factors were instead driven as a function of whether the pilot flew with a HUD or an HWD. When flying with an HWD, pilots tended to land closer to the runway threshold and with a higher sink rate than they did when flying with a HUD. Because pilots landed sooner and with a higher vertical velocity with the HWD than they did with the HUD, and because these outcomes were consistent between the monocular and binocular HWD conditions, the physical and optical characteristics of the HWD in this study—as well as increased pilot workload when flying with an HWD—appear to be contributing factors in determining a pilot's ability to initiate the flare sequence at the optimal time to reduce the vertical velocity of the aircraft prior to touchdown.

RQ_3: Do pilots experience different workload levels during an instrument approach and landing depending on whether they fly with a monocular HWD, binocular HWD, or HUD?

One of the primary takeaways from this research is that pilot workload was significantly higher when pilots flew with the monocular HWD compared to when they flew with the binocular HWD or HUD during normal scenarios. Pilot workload was also higher when flying with the binocular HWD than when flying with the HUD. The elevated workload with both HWD configurations relative to the HUD is consistent with earlier research on aviation HWDs and suggests that increased effort is required to use flight symbology presented on a near-to-eye display relative to when using flight symbology on a HUD (Thomas, 2010). Analysis of NASA-TLX subscale ratings indicates that this increase in workload as a function of display type was driven by all NASA-TLX subscale ratings, with the primary drivers being Mental Demand, Physical Demand, Temporal Demand, Effort, and Frustration. The trends in NASA-TLX ratings are corroborated by written feedback on the usability survey, in which participants thematically reported that more cognitive effort was required to use the symbology on the monocular HWD compared to the other display types. Responses to the usability questionnaire also referenced increased workload for both HWD configurations relative to the HUD because the near-to-eye-combiner obscured their view of the head-down instrumentation. These findings suggest that pilot workload should be a primary consideration when implementing an HWD in lieu of a HUD, particularly for workload that arises from a monocular HWD configuration and interference between the near-to-eye HWD combiner and other displays and information sources on the flight deck.

NASA-TLX workload ratings remained largely consistent across the degrees of RVR used in this study. One possible explanation for the lack of an effect of RVR is that the range of RVR levels employed in this research was not of a magnitude that would elicit differences in pilot workload. It is also possible that pilots' subjective workload ratings were biased by the large influence of display type on workload. Indeed, there is evidence that subjective workload assessments tend to be biased such that ratings are driven by the greatest contributor to workload during a task and less so by factors that have a lower influence on workload (i.e., peak-end effect; Peterson & Kozhokar, 2017; Qiao et al., 2022). In this case, pilots may have been less sensitive to workload variations driven by RVR than to those driven by display type.

In non-normal scenarios, PF workload was elevated when flying with the binocular and monocular HWD compared to when flying with the HUD, but no differences were found between the binocular and monocular HWD—a different pattern compared to workload patterns during normal scenarios. The standard operating procedures during a missed approach may interact with the unique optical and physical characteristics of the HWD to contribute to this pattern. As a point, when the flight crew executes a missed approach in a Boeing 737 aircraft, the PF must divert attention from the HUD or HWD symbology to scan aircraft and navigation information in the head-down displays, as well as coordinate with the PM to carry out aircraft configuration changes (e.g., flaps and landing gear configuration; Boeing Aircraft Co., 2023; United Airlines, 2022). During this process, the physical and optical differences between the HUD and HWD, including the near-to-eye combiner of the HWD obscuring the pilot's view of the instrument panel, may have been larger contributors to workload than the perceptual effects of binocular rivalry when flying with a monocular HWD. This finding indicates that future evaluations of HWD systems and operations should include assessments of pilot workload during missed approach procedures in addition to routine instrument approaches and landings to ensure that the near-to-eye optical hardware of the HWD does not obscure the pilot's view of the instrument panel.

RQ$_4$: Is there a difference in pilots' ability to detect symbology failures or hazards on the runway depending on whether they fly with a monocular HWD, binocular HWD, or HUD?

While the small number of data points for the first unexpected runway incursion for each crew precluded any inferential analyses, patterns among the data indicate that pilots initiated a missed approach at roughly the same point in response to the first runway incursion regardless of the display type in use. As pilots experienced repeated runway incursion scenarios, there was an increase in the speed with which they detected the incursion and pressed the TO/GA button to initiate the missed approach, indicating that increased event expectancy led to faster response times. Patterns among these data begin to develop when pilots experience repeated runway incursion scenarios and develop an expectancy for the event. Across all pilot response data in the runway incursion scenarios, the PF initiated a missed approach in response to the incursion significantly sooner when flying with a monocular HWD than they did when flying with a binocular HWD. On average, the aircraft was 456 feet closer, longitudinally, to the runway TDZ when the PF initiated a missed approach in response to the incursion when flying with the binocular HWD compared to when they were flying with the monocular HWD. This finding suggests that pilots are better able to see the runway incursion through the symbology when flying with a monocular HWD.

It is possible that this occurred because the monocular HWD provided one eye with an unobscured view of the runway environment, making it easier for the pilot to look through the symbology and see the runway incursion. Past research demonstrates that an image presented on a monocular HWD interferes less with the background scenery than an image presented on a binocular HWD, which can enhance performance on tasks requiring the use of information in the background (Laramee & Ware, 2002). Binocular rivalry may also cause monocular symbology to be less visually compelling than binocular symbology; that is, because binocular rivalry disrupts visual perception of the symbology, the monocular symbology may be less “attractive” to attention, and the pilot may be less inclined to tunnel into the symbology at the expense of scanning the runway environment (Wickens & Yeh, 2018).

The PF also responded to runway incursions much sooner when flying with the HUD than when flying with the binocular HWD: On average, the aircraft was 450 feet closer, longitudinally, to the runway TDZ when the PF initiated a missed approach in response to the incursion when flying with the binocular HWD compared to when they were flying with the HUD. These patterns in the data are reflected in participants' written feedback, which indicates that it was more difficult to see runway incursions and other unsafe runway conditions when flying with the HWD. This finding highlights the impact of display collimation on the pilot’s ability to switch attention from symbology to the runway when reaching the DA of the approach. Display collimation is a foundational aspect of HUD design, and its importance is supported by findings from this research. While display collimation was not a fully controlled experimental variable in this study, patterns in the data support the conclusion that pilot response to the runway incursion scenarios when flying with the HWD might have occurred sooner if the HWD in this study had been collimated.

In terms of PF response to the FD failures, the ability to detect and respond to the failure was not compromised when pilots flew with a monocular or binocular HWD in lieu of a HUD. Responses to the first occurrence of the failure were delayed compared to responses to subsequent failures, suggesting that pilots were better able to anticipate and respond to future instances of the failure; however, the response time was not different across the display types. Previous research indicates that performance on simple reaction time tasks can be poorer when targets are presented on a monocular display (e.g., see Winterbottom et al., 2006b); however, this study suggests that this may not translate to delays in detecting failures of flight guidance symbology in an operational setting with a monocular HWD.

It is probable that the more complex nature of detecting and interpreting a guidance symbology failure, determining whether it is safe to continue to a landing after the failure, and executing the appropriate response adds additional variability to the time taken to press the TO/GA button, potentially masking any effect of binocular rivalry on response time (e.g., see Fercho, Beringer, & Donovan, 2024).

RQ$_5$: Do the physical and optical differences between the HUD and HWD used in this study impact pilot performance, workload, and non-normal event detection during an instrument approach and landing?

Several of the pilot performance and workload outcomes in this research suggest that the unique optical and physical characteristics of the HWD in this study, compared to a production-quality HUD, influenced several aspects of pilot performance and workload during instrument approach and landing operations. In particular, while pilots were able to maintain the flightpath of the approach and land on the runway centerline just as well when flying with the binocular HWD as when flying with the HUD, they were less able to maintain awareness of and correct for deviations from the airspeed target, and they were less accurate in managing sink rate leading up to touchdown, as evidenced by a higher vertical velocity at touchdown, on average, when flying with an HWD. When encountering a runway incursion, pilots responded significantly more slowly using the binocular HWD than using the HUD. There was also a significant difference in pilot response to the incursions between the HUD and the HWD during the runway incursion scenarios. Whereas pilots appeared to respond to the first runway incursion event of the study session at roughly the same point in the approach regardless of whether they used a HUD, binocular HWD, or monocular HWD, as they encountered subsequent incursions—and learned to expect them—they were faster on average in responding to them when using a HUD compared to when using a binocular HWD. This finding indicates that there may be substantial safety implications regarding whether an HWD is collimated.

These results may largely be attributed to the non-collimated display of the HoloLens 2 compared to the collimated display of the production-quality HUD. The HUD, with a focal distance of optical infinity, minimizes the need for the pilot's eyes to re-accommodate when dividing attention between the symbology and the runway environment. HUD collimation has been shown to lead to more rapid detection of runway visual information when transitioning to visual flight references and more effective integration of conformal elements of the symbology with their real-world counterparts (Peterson, 2006; Weintraub et al., 1985; Weintraub & Endsing, 1992). A collimated display is a foundational HUD feature, which enables efficient division of attention between the symbology and the runway environment (Kimchi et al., 2007; Wickens & Long, 1995). The results of the present study provide corroborating evidence to support this claim. When landing, pilots appeared to be better able to judge the height of the aircraft above the runway and manage the sink rate of the aircraft prior to touchdown when using the collimated HUD compared to when using the non-collimated HWD. Based on these factors, although display collimation was not a fully controlled experimental variable in this study, the patterns in the data suggest that if pilots flew with an HWD that featured a collimated display in this study, pilot performance might be improved compared to the performance observed in this research.

RQ_6: Is pilot eye dominance a significant consideration when flying with a monocular HWD?

The findings of this research indicate that sighting eye dominance may not be a primary driver of pilots' experiences when flying with a monocular HWD. When pilots were asked if they preferred the monocular HWD on one eye over the other during the familiarization session, nearly all pilots indicated that they had no preference. Of those pilots who did prefer one eye over the other, they elaborated by describing that their preference was driven by other factors besides eye dominance, including interocular differences in visual acuity, where some pilots reported possessing better far vision in one eye over the other, and that they would prefer to view the symbology on that eye. These findings are similar to the results of the Apache AH-64 IHADSS HWD evaluation by Hiatt et al. (2004). In that evaluation, pilots' visual experiences while flying with the IHADSS were not dependent on which eye was their dominant sighting eye. These findings also corroborate empirical research showing that visual task performance and subjective comfort while using a monocular display do not differ as a function of eye dominance (Bayle et al., 2020).

The practical takeaway from this finding for monocular HWD implementation on the flight deck is that a test of sighting eye dominance may not be a critical aspect of monocular HWD implementation. Some pilots may prefer to use a monocular HWD on one eye rather than the other. However, this preference may be based on other individual factors. This suggests that a monocular HWD that can be configured to present on either the left or right eye may improve the user experience for pilots but may not be critical for ensuring optimal performance and safety. An important consideration alongside these findings is that the subjective evaluation of the left-versus-right eye configuration of the monocular HWD was based on a small number of scenarios during the familiarization session, which was the pilots' first experience flying with the HWD. Practice effects and more subtle differences between the left- and right-eye configurations of a monocular HWD may arise after extended use, which would be important to evaluate in future research.

Limitations

There are aspects of this research that may limit the generalizability of the findings to a broad range of operational use cases. A primary limitation was that the Microsoft HoloLens 2 HWD used in this study was not originally designed for use on a flight deck and may not best represent HWD technology that will be used in real flight operations. The use of the HoloLens 2 bolstered the internal validity of the study by enabling the presentation of identical flight symbology regardless of display type, thus avoiding a confound in the research design that would be present if fundamental aspects of the symbology differed between the HWD and the HUD. Some study participants provided feedback addressing this limitation, commenting that the HWD imagery appeared to be lower resolution than the HUD imagery, which could have been a source of distraction during the scenarios. Participants also commented that the prominent near-to-eye combiner of the HWD made it more difficult to clearly see the head-down displays on the flight deck. Finally, some participants commented that transitioning to visual flight references during the scenarios was made more challenging by the non-collimated image on the HWD compared to the collimated HUD. These characteristics may not be present in HWDs designed to be certified for use on a flight deck. Based on these limitations, it would be beneficial to verify results from this research using a production-quality aviation HWD that incorporates a collimated image display.

An additional limitation of this research, due to the use of a Microsoft HoloLens 2, is that while the device offers a moving platform mode that can be enabled for use in vehicles, this mode was not compatible with the motion system in the simulator. As a result, the simulator motion was not enabled for the study, which may have impacted pilots' performance and workload levels during the scenarios compared to if the simulator motion had been enabled. Several participants commented that there was an additional learning curve to overcome because of the difference in simulator handling qualities when the simulator motion was disabled. Based on this, it is recommended that future research investigate methods for implementing an HWD that is compatible with simulator motion so that motion can be enabled during the flight scenarios.

An aspect of the research design that may impact the external validity of this research is the protocol for the PM to refrain from all verbal callouts during each scenario, including any callouts about non-normal events. While this protocol represents a departure from established crew coordination practices for operations with a two-person flight crew, it was critical for ensuring that any patterns in the data were a function of the display type in use during the scenario. In real-world operational settings for a two-person flight crew, FAA guidance dictates that the role of the PM is to monitor the flightpath, aircraft state, and environment while the PF is controlling the aircraft or monitoring any automated flight control systems. They are responsible for verbally alerting the PF to any situations that may compromise the safety of the operation or necessitate a missed approach. This aspect of crew coordination is an important component of flight safety in real-world operations.

In research settings where non-normal events are introduced to evaluate the effect of technology used specifically by the PF on detecting those events, standard PM callouts compromise the internal validity of the study. For example, the PM may anticipate that the PF experiences increased workload and compromised performance and subsequently provide more frequent callouts to alert them to deviations from the flightpath or target airspeed. During a runway incursion scenario, the PM may see the incursion before the PF and alert them of the hazard and need for a missed approach. However, the protocol to refrain from callouts enables the findings of this study on the impacts of monocular and binocular HWD use compared to HUD use to be generalized beyond the operational setting for this study to a wider variety of settings, including those where there is a single pilot on the flight deck who is flying with an HWD or HUD in use.

Conclusions and Future Directions

Despite these limitations, the results from this study provide valuable evidence about how the implementation of an HWD in lieu of a HUD might impact performance and workload. The results from this study suggest that if pilots use flight symbology on a binocular or monocular HWD to fly low-visibility approach and landing operations, they would exhibit somewhat similar flightpath tracking and airspeed management accuracy to what they would exhibit if the symbology is presented on a HUD. However, pilots may experience elevated workload levels when flying with an HWD, especially if the HWD is monocular. Therefore, if pilots fly with an HWD, particularly a monocular HWD, it would be important to ensure that failsafe processes and technologies are in place to curtail the risk of workload-induced performance problems. This may include display design characteristics, crew resource management processes, pilot training, and operational authorizations that are designed to mitigate additional workload that may arise when pilots fly with an HWD.

Conversely, a pilot's ability to acquire visual references and evaluate the safety of the runway environment before landing may be enhanced when flying with a monocular HWD compared to when flying with a binocular HWD, as evidenced by faster response to runway incursions when flying with the monocular HWD. Regardless, there were significant delays in response by the PF to the initial, unexpected runway incursion, regardless of whether they flew with a HUD or HWD. To generalize this finding to situations outside the scope of the present study, this is a particular concern for flight operations that may be conducted by a single pilot flying with a HUD or HWD (e.g., 14 CFR § 91; 14 CFR § 135), where there is not a second pilot to independently verify the occupancy of a runway. This finding highlights the importance of employing additional mitigations for runway incursion events, such as Runway Incursion Devices, Surface Awareness Initiative Systems, and Approach Runway Verification technology.

Taken as a whole, the findings from this research suggest that the most critical human factors considerations for flight deck HWD implementations are pilot cognitive and physical workload, compatibility between HWD symbology and the runway environment, and visibility of flight deck displays and controls through the HWD combiner. These factors may be significantly impacted by whether the HWD is monocular or binocular, whether the HWD image is collimated, and the physical characteristics of the HWD, such as combiner transparency and shape, as well as headset weight and size. The findings from this research could be applied to the development of future operational authorizations for the use of an HWD for flight operations where flight guidance on a HUD is currently required—such as manually flown ILS approach and landing operations with RVR below 1800 ft—and could potentially inform HWD design guidance.

References

Achour, G., Silva, R. F., Payan, A. P., Johnson, C. C., & Mavris, D. N. (2023). Rotorcraft takeoff analysis and classification to detect outlier operations that could present a safety risk. In *AIAA AVIATION 2023 Forum* (p. 3578).

Arefin, M. S., Phillips, N., Plopski, A., Gabbard, J. L., & Swan, J. E. (2022). The effect of context switching, focal switching distance, binocular and monocular viewing, and transient focal blur on human performance in optical see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, *28*(5), 2014-2025.

Aswathappa, J., Kutty, K., & Annamalai, N. (2011). Relationship between handedness and ocular dominance in healthy young adults–A study. *Int J Pharm Biomed Res*, *2*(2), 76-78.

Bailey, R. E. (1989). Effect of head-up display dynamics on fighter flying qualities. *Journal of Guidance, Control, and Dynamics*, *12*(4), 514-520.

Bailey, R. E., Arthur III, J. J., & Williams, S. P. (2004, August). Latency requirements for head-worn display S/EVS applications. In *Enhanced and Synthetic Vision 2004* (Vol. 5424, pp. 98-109). SPIE.

Bailey, R. E., Arthur III, J. J., Williams, S. P., & Kramer, L. J. (2005, January). Latency in visionic systems: Test methods and requirements. In *Workshop on Toward Recommended Methods for Testing and Evaluation of EV and ESV Based Visionic Devices*.

Bayle, E., Hourlier, S., Lelandais, S., Salasc, C. A., Leroy, L., Plantier, J., & Neveu, P. (2020, April). Monocular augmented reality devices: Does the choice of eye matter? In *Situation Awareness in Degraded Environments 2020* (Vol. 11424, pp. 45-59). SPIE.

Bennett, C. T., & Schwirzke, M. (1992). Analysis of accidents during instrument approaches. *Aviation, Space, and Environmental Medicine, 63*(4), 253-261.

Beringer, D. B. (2020). *Synthetic vision applied to general aviation: An evaluation of pilot performance and preferences when using head-up, head-down, and head-mounted synthetic vision displays for SA CAT I approaches in flat terrain and missed approaches in challenging terrain* (Report No. DOT/FAA/AM-20/02). Washington, D.C.: Federal Aviation Administration.

Beringer, D. B., Domino, D. A., & Kamenski, J. (2018). *Pilot performance on a SA CAT I instrument approach using synthetic vision and a retrofit head-down display* (Report No. DOT/FAA/AM-19/9). Washington, D.C.: Federal Aviation Administration.

Beringer, D. B., & Drechsler, G. (2013). Enhancing helicopter-pilot obstacle avoidance using a binocular head-mounted display. *17th International Symposium on Aviation Psychology,* 500-505.

Blake, R. (2001). A primer on binocular rivalry, including current controversies. *Brain and Mind, 2*, 5-38.

Boeing Aircraft Co. (2023). *737 NG/MAX Flight Crew Operations Manual*. Arlington, VA: Boeing Aircraft Co.

Boeing Aircraft Co. (2025). *Statistical Summary of Commercial Jet Airplane Accidents* (1959-2024). Arlington, VA: Boeing Aircraft Co.

Boucek, G. P., Pfaff, T. A., & Smith, W. D. (1983). *The use of holographic head-up display of flightpath symbology in varying weather conditions.* (SAE Technical Paper No. 831445) Warrendale, PA: Society of Automotive Engineers.

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

Condino, S., Carbone, M., Piazza, R., Ferrari, M., & Ferrari, V. (2019). Perceptual limits of optical see-through visors for augmented reality guidance of manual tasks. *IEEE Transactions on Biomedical Engineering*, *67*(2), 411-419.

Dag, O., Asar, O., Ilk, O., & Dag, M. O. (2017). Package 'AID.'

de Waard, D. 1996. *The Measurement of Drivers' Mental Workload* [Doctoral dissertation]. Traffic Research Centre, University of Groningen, Haren.

Dieter, K. C., & Tadin, D. (2011). Understanding attentional modulation of binocular rivalry: A framework based on biased competition. *Frontiers in Human Neuroscience*, *5*, 155.

Eser, I., Durrie, D. S., Schwendeman, F., & Stahl, J. E. (2008). Association between ocular dominance and refraction. *Journal of Refractive Surgery*, *24*(9), 685-689.

Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDS: Are they viable? *Human Factors, 43*(2), 173-193.

Federal Aviation Administration. (1999). *Criteria for Approval of Category III Weather Minima for Takeoff, Landing, and Rollout* (Advisory Circular No. 120-28D). Washington, D.C.

Federal Aviation Administration. (2011a). *Runway Visual Range (RVR)* (Order No. 6560.10C). Washington, D.C.

Federal Aviation Administration. (2011b). *Spatial Disorientation: Visual Illusions* (Publication No. AM-400-00/1). Washington, D.C.

Federal Aviation Administration. (2017b). *Standard Operating Procedures and Pilot Monitoring Duties for Flight Deck Crewmembers* (Advisory Circular No. 120-71B). Washington, D.C.

Federal Aviation Administration. (2018). *Criteria for Approval/Authorization of All Weather Operations (AWO) for Takeoff, Landing, and Rollout* (Advisory Circular No. 120-118). Washington, D.C.

Federal Aviation Administration. (2022a). *Flightpath Management* (Advisory Circular No. 120-123). Washington, D.C.

Federal Aviation Administration Flight Technologies and Procedures Division. (2022b). *Report to Congress: Advanced Cockpit Displays*. Washington, D.C.

Federal Aviation Administration National Simulator Program. (2012). *Head-Up Display (HUD) Simulator Qualification* (FSTD Guidance Bulletin 03-02). Washington, D.C.

Fercho, K. A., Beringer, D. B., & Donovan, C. (2024, September). Factors that affect pilot response times to alerts: Findings from a literature review and Aviation Safety Reporting System (ASRS) reports. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 68*(1), 50-55.

Fischer, E., & Haines, R. F. (1980). *Cognitive issues in head-up displays* (NASA Technical Paper No. 1711). Moffett Field, CA: NASA.

Flight Safety Foundation. (2009). *Special report: head-up guidance system technology – a clear path to increasing flight safety*. Alexandria, VA: Flight Safety Foundation.

Fox, R., & Check, R. (1968). Detection of motion during binocular rivalry suppression. *Journal of Experimental Psychology*, *78*(3p1), 388.

Gibb, R., Schvaneveldt, R., & Gray, R. (2008). Visual misperception in aviation: Glide path performance in a black hole environment. *Human Factors*, *50*(4), 699-711.

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Goteman, Ö., Smith, K., & Dekker, S. (2007). HUD with a velocity (flight-path) vector reduces lateral error during landing in restricted visibility. *The International Journal of Aviation Psychology*, *17*(1), 91-108.

Gronwall, D. M. A., & Sampson, H. (1971). Ocular dominance: A test of two hypotheses. *British Journal of Psychology*, *62*(2), 175-185.

Hancock, P. A., & Ganey, H. C. N. (2003). From the inverted-U to the extended-U: The evolution of a law of psychology. *Journal of Human Performance in Extreme Environments, 7*(1), 5–14.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In A*dvances in Psychology* (Vol. 52, pp. 139-183). North-Holland.

Hiatt, K. L., Rash, C. E., Harris, E. S., & McGilberry, W. H. (2004, September). AH-64 IHADSS aviator vision experiences in operation Iraqi freedom. In *Helmet-and Head-Mounted Displays IX: Technologies and Applications* (Vol. 5442, pp. 179-194). SPIE.

Iaveccia, J. H., Iaveccia, H. P., & Roscoe, S. N. (1988). Eye accommodation to head-up virtual images. *Human Factors, 30*(6), 689-702.

Interaction Design Foundation. (n.d.). *Field of view (FOV) in extended reality.* https://www.interaction-design.org/literature/topics/field-of-view-fov-in-extended-reality?srsltid=AfmBOors5y8PERW6Ndk6_ehNc5m58GNxH8QSY2MCjpFE56iCYJyxBYU-

Jarmasz, J., Herdman, C. M., & Johannsdottir, K. R. (2005). Object-based attention and cognitive tunneling. *Journal of Experimental Psychology: Applied*, *11*(1), 3-12.

Johansson, J., Seimyr, G. Ö., & Pansell, T. (2015). Eye dominance in binocular viewing conditions. *Journal of Vision*, *15*(9), 21-21.

Johns, J. B., & Funk, J. D. (1991). Impact of V-22 display latency on flying qualities. *Naval Air Development Center Report NADC-91040-60*.

Kimchi, R., Yeshurun, Y., & Cohen-Savransky, A. (2007). Automatic, stimulus-driven attentional capture by objecthood. *Psychonomic Bulletin & Review*, *14*, 166-172.

Lack L. C. (1978). *Selective Attention and the Control of Binocular Rivalry*. Paris: Mouton

Laramee, R. S., & Ware, C. (Eds.). (2002). Rivalry and interference with a head-mounted display. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *9*(3), 238-251.

Larish, I., & Wickens, C. D. (1991). Attention and HUDs: Flying in the dark? In *Society for Information Display International Symposium Digest of Technical Papers XXII*.

Levy, B. S., Som, P., & Greenhaw, R. (2003). Analysis of flight technical error on straight, final approach segments. In *Proceedings of the 59th Annual Meeting of The Institute of Navigation and CIGTF 22nd Guidance Test Symposium (2003)* (pp. 456-467).

Lintern, G. (2000). An analysis of slant for guidance of landing approaches. *The International Journal of Aviation Psychology*, *10*(4), 363-376.

Lintern, G., & Liu, Y. T. (1991). Explicit and implicit horizons for simulated landing approaches. *Human Factors*, *33*(4), 401-417.

Malik, F., Khan, A., & Shah, M. T. A. (2018). Box-Cox Transformation approach for data normalization: A study of new product development in manufacturing sector of Pakistan. *IBT Journal of Business Studies, 1*(1).

Murphy, G., & Murphy, L. (2018). Perceptual load affects change blindness in a real-world interaction. *Applied Cognitive Psychology, 32*(5), 655–660

National Transportation Safety Board. (1973). *Aircraft Accident Report Eastern Air Lines, Inc. L-1011, N310EA Miami, Florida December 29, 1972* (Report No. NTSB-AAR-73-14). Washington, D.C.

Newton, D. C., Mofle, T. C., DiDomenica, R. L., Choi, I., & Hesselroth, J. (forthcoming). *Dual HUD use during CAT II and CAT III approach, landing, and rollout operations: Considerations for flight crew coordination, performance, and workload.* Washington, D.C.: Federal Aviation Administration.

Osborne, J. W. (2010). Improving your data transformations: Applying the Box-Cox Transformation. *Practical Assessment, Research, and Evaluation, 15*(12), 1-9.

Paprocki, T. H., & Gates, R. F. (1966). *Evaluation of minimum approach light system for lower activity airports* (Report No. RD-66-11). Washington, D.C.: Federal Aviation Agency.

Patterson, R., & Martin, W. L. (1992). Human stereopsis. *Human Factors*, *34*(6), 669-692.

Patterson, R., Winterbottom, M. D., & Pierce, B. J. (2006). Perceptual issues in the use of head-mounted visual displays. *Human Factors*, *48*(3), 555-573.

Patterson, R., Winterbottom, M., Pierce, B., & Fox, R. (2007). Binocular rivalry and head-worn displays. *Human Factors, 49*(6), 1083-1096.

Peterson, S. (2006). Collimation of airport tower HUDs. In *5th Eurocontrol Innovative Research Workshop and Exhibition in Bretigny sur Orge, France*.

Peterson, D. A., & Kozhokar, D. (2017). Peak-end effects for subjective mental workload ratings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*(1), 2052-2056.

Qiao, H., Zhang, J., Zhang, L., Li, Y., & Loft, S. (2022). Exploring the peak-end effects in air traffic controllers' mental workload ratings. *Human Factors*, *64*(8), 1292-1305.

Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., & McMahan, R. P. (2015). Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE Transactions on Visualization and Computer Graphics*, *21*(7), 794-807.

Rash, C. E. (1998). H*elmet mounted displays. Design issues for rotary-wing aircraft*. http://www.usaarl.army.mil/hmdbook/cp_0002_contents.htm

Rey, M., Aloise, D., Soumis, F., & Pieugueu, R. (2021). A data-driven model for safety risk identification from flight data analysis. *Transportation Engineering*, *5*, 100087.

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight: A decade of practical use*. Bedford, UK: Royal Aerospace Establishment.

Rosekind, M. R., Co, E. L., Gregory, K. B., & Miller, D. L. (2000). *Crew factors in flight operations XIII: A survey of fatigue factors in corporate/executive aviation operations* (Report No. NASA/TM-2000-209610). National Aeronautics and Space Administration.

SAE International. (2008, January). *Aerospace recommended practice: Transport category airplane head up display (HUD) systems* (SAE Standard ARP5288). Warrendale, PA: SAE International.

SAE International. (2023, July). *Aerospace recommended practice: Human engineering considerations associated with design and airborne implementation of head worn display (HWD) systems* (SAE Standard ARP6377). Warrendale, PA: SAE International.

Schall, J., Nawrot, M., Blake, R., & Yu, K. (1993). Visual guided attention is neutralized when informative cues are visible but unperceived. *Vision Research, 33*, 2057–2064.

Shneor, E., & Hochstein, S. (2005, September). Effects of eye dominance in visual perception. In *International Congress Series* (Vol. 1282, pp. 719-723). Elsevier.

Thomas, L. C. (2009, September). Head-worn displays on commercial flight decks. *SPIE Newsroom*. Retrieved from https://www.spie.org/news/1735-head-worn-displays-on-commercial-flight-decks#_=

Thomas, L. C. (2010). Evaluation of the use of a head worn display (HWD) for flight support in the commercial flight deck. *SAE International Journal of Aerospace, 2*(1), 109-114.

United Airlines. (2022, January). 737 *Flight Manual*. Denver, CO: United Airlines.

Velger, M. (1998). *Helmet-mounted displays and sights.* Boston, MA: Artech House.

Walker, P. (1975). The subliminal perception of movement and the ‘suppression’ in binocular rivalry. *British Journal of Psychology*, *66*(3), 347-356.

Weintraub, D. J., & Endsing, M. (1992). *Human factors issues in head-up display design: The book of HUD* (Report No. SOAR 92-2). Dayton, OH: University of Dayton Research Institute.

Weintraub, D. J., Haines, R. F., & Randle, R. J. (1984). The utility of head-up displays: Eye-focus vs decision times. *Proceedings of the Human Factors Society Annual Meeting, 28*(6), 529-533.

Weintraub, D. J., Haines, R. F., & Randle, R. J. (1985, October). Head-up display (HUD) utility, II: Runway to HUD transitions monitoring eye focus and decision times. *Proceedings of the Human Factors Society Annual Meeting, 29*(6), 615-619.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449-455.

Wickens, C. D., & Long, J. (1995). Object versus space-based models of visual attention: Implications for the design of head-up displays. *Journal of Experimental Psychology: Applied, 1*(3), 179.

Wickens, C. D., & Yeh, M. (2018). *Display compellingness: A literature review* (Report No. DOT/FAA/AM-19/13). Washington, D.C.: Federal Aviation Administration.

Wilson, G. F., & Hankins, T. (1994). EEG and subjective measures of private pilot workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 38*(19), 1322-1325. https://doi.org/10.1177%2F154193129403801916

Winterbottom, M., Patterson, R., Covas, C., Rogers, J., & Pierce, B. (2006a). Binocular rivalry and attention in helmet-mounted display applications. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*, 2061–2066.

Winterbottom, M., Patterson, R., Pierce, B., & Taylor, A. (2006b). Visual suppression of monocularly presented symbology against a fused background in simulation and training environment. *Proceedings of SPIE: Helmet- and Head-Mounted Displays XI: Technologies and Applications, 6224*, 2.1-2.10

Yeh, M., & Wickens, C. D. (1997). *Performance issues in helmet mounted displays* (Report No. ARL-97-9/ARMY-FED-LAB-97-1). Savoy, IL: University of Illinois Aviation Research Laboratory.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*(1), 1-17.

Young, M. S., & Stanton, N. A. (2005). Mental workload. In N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. W. Hendrick (Eds.), *Handbook of Human Factors and Ergonomics Methods*. London: Taylor & Francis.