DOT/FAA/AM- AM-26/03

Office of Aerospace Medicine

Washington, D.C. 20591

Pilot Performance, Workload, and Usability Considerations for the Use of a Monocular Head-Worn Display in Lieu of a Head-Up Display During SA CAT I Operations with an Enhanced Flight Vision System

David C. Newton, Ph.D.[1]
Inchul Choi, Ph.D.[2]
Hanna Forbat[2]
Rebecca DiDomenica[2]
Farzaneh Shahini, Ph.D.[2]
Theodore C. Mofle[1]
Ryan Weaver[3]

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

[2]Cherokee Federal
Oklahoma City, OK 73169

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

January 2026

Technical Report Documentation

1. Report No.	
DOT/FAA/AM- AM-26/03	
2. Title & Subtitle Pilot Performance, Workload, and Usability Considerations for the Use of a Monocular Head-Worn Display in Lieu of a Head-Up Display During SA CAT I Approach, Landing, and Rollout with an Enhanced Flight Vision System	**3. Report Date** January 2026 **4. Performing Organization Code** AAM-510
5. Author(s) David C. Newton (0000-0002-3112-3316), Inchul Choi (0000-0002-2048-7026), Hanna Forbat, Rebecca DiDomenica (0009-0009-4726-578X), Farzaneh Shahini (0000-0002-1514-5312), Theodore C. Mofle (0000-0003-2217-9083), and Ryan Weaver	**6. Performing Org Report Number** N/A
7. Performing Organization Name & Address [1]Federal Aviation Administration Civil Aerospace Medical Institute Oklahoma City, OK 73169 [2]Cherokee Federal Oklahoma City, OK 73169 [3]Federal Aviation Administration Flight Research and Analysis Group Oklahoma City, OK 73169	**8. Contract or Grant Number** N/A
9. Sponsoring Agency Name & Address Federal Aviation Administration Flight Standards Service, AFS-410 800 Independence Ave., S.W. Washington, DC 20591	**10. Type of Report & Period Covered** N/A
11. Supplementary Notes N/A	

12. Abstract

When flying a Special Authorization Category I (SA CAT I) instrument landing system (ILS) approach, pilots may use a Head-Up Display (HUD), which presents flight symbology on a transparent screen so that the pilot can view primary flight information while looking out the window, along the flightpath. Pilots can also use an Enhanced Flight Vision System (EFVS) on a HUD during this operation, which provides a real-time sensor image of the forward view to enhance runway awareness when transitioning to visual flight references. The Head-Worn Display (HWD) is an emerging technology in civil aviation that is designed to provide the benefits of a HUD; however, the unique optical and physical characteristics of the HWD may change the existing levels of pilot performance and workload during SA CAT I operations flown with a HUD. When flying with a monocular HWD, binocular rivalry occurs, which may impact pilot performance and workload. This raises questions about whether pilot performance and workload are significantly impacted during manual SA CAT I flight operations, in which the pilot flying (PF) uses a monocular HWD with and without an EFVS. To address this concern, a study was carried out in which 11 pilot crews, made up of 22 Airline Transport Pilot (ATP) Captains, flew manual SA CAT I approach, landing, and rollout scenarios in a Boeing 737 Level D-equivalent flight simulator with a HUD and monocular HWD, with and without an EFVS, and in day and night ambient lighting conditions. Simulator motion was disabled to prevent interference with the HWD head tracking system. The PF rated their workload during each scenario using the National Aeronautics and Space Administration Task Load Index (NASA-TLX). The findings of the study suggest that a monocular HWD may not have a significant negative impact on a pilot's ability to manage most aspects of the flightpath during an SA CAT I operation; however, the monocular HWD elevated pilot workload. The monocular HWD also caused increased glideslope deviation during the instrument segment and increased deviation from the runway centerline during rollout. However, these increases were small and may not translate to operational significance. Pilots reported that the EFVS enhanced their awareness of the runway environment when transitioning from instrument to visual flight references; however, because there was not an EFVS kill switch for the PF, it was reported to interfere with natural vision when transitioning to flare, landing, and rollout, regardless of when implemented on a HUD and monocular HWD. Ultimately, this research contributes to the understanding of how SA CAT I operations that are flown with and without an EFVS may be impacted when pilots fly with a monocular HWD in lieu of a HUD.

13. Key Words	14. Distribution Statement
Advanced vision systems, all-weather operations, enhanced vision system, enhanced flight vision system, equivalent visual operations, extended reality, head-mounted display, head-up display, head-worn display, low-visibility operations, pilot performance, pilot workload	N/A

15. Security Classification (of this report)	16. Security Classification (of this page)	17. No. of Pages	18. Price
N/A	N/A	109	N/A

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 their valuable guidance and support throughout the development and execution of this research. 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 support during research plan development, scenario design, and data collection. From Collins Aerospace: Nate Kowash, Alex Reid, 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. The authors also express gratitude to the pilot subject matter experts who supported head-worn display acceptance 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

List of Figures

List of Tables

List of Abbreviations

AGL	Above Ground Level
ALSF-2	High-Intensity Approach Lighting System with Sequenced Flashing Lights
ATP	Airline Transport Pilot
CAE	Canadian Aviation Electronics
CAST	Commercial Aviation Safety Team
CFI	Certified Flight Instructor
CFII	Certified Flight Instructor – Instrument
CLL	Centerline Lighting
DH	Decision Height
DVI2USB	Digital Visual Interface to USB video grabber
DA	Decision Altitude
DV	Dependent Variable
EFVS	Enhanced Flight Vision System
FAA	Federal Aviation Administration
FD	Flight Director
FLIR	Forward-Looking Infrared
FOR	Field of Regard
FOV	Field of View
HGS	Head-Up Guidance System
HIRL	High-Intensity Runway Lighting
HDD	Head-Down Display
HUD	Head-Up Display
HWD	Head-Worn Display
Hz	Hertz (frequency measure)
IAP	Instrument Approach Procedure
IAS	Indicated Airspeed
IMC	Instrument Meteorological Conditions
IFR	Instrument Flight Rules

ILS	Instrument Landing System
IQR	Interquartile Range
IR	Infrared
IV	Independent Variable
LOC-I	Loss-of-Control In-flight
LRU	Line Replaceable Unit
M. HWD	Monocular Head-Worn Display
M_{adj}	Adjusted Mean
MCP	Mode Control Panel
MDA	Minimum Descent Altitude
MEL	Minimum Equipment List
NAS	National Airspace System
NASA-TLX	National Aeronautics and Space Administration Task Load Index
NSP	National Simulation Program
PDX	Portland International Airport
PF	Pilot Flying
PI	Principal Investigator
PM	Pilot Monitoring
RA	Radar Altimeter
RM ANOVA	Repeated Measures Analysis of Variance
RMS	Root Mean Square
RVR	Runway Visual Range
SA CAT I	Special Authorization Category I
SD	Standard Deviation
SME	Subject Matter Expert
SOP	Standard Operating Procedure
TDZ	Touchdown Zone
TDZE	Touchdown Zone Elevation
TOST	Two One-Sided Test

Abstract

When flying a Special Authorization Category I (SA CAT I) instrument landing system (ILS) approach, pilots may use a Head-Up Display (HUD), which presents flight symbology on a transparent screen so that the pilot can view primary flight information while looking out the window, along the flightpath. Pilots can also use an Enhanced Flight Vision System (EFVS) on a HUD during this operation, which provides a real-time sensor image of the forward view to enhance runway awareness when transitioning to visual flight references. The Head-Worn Display (HWD) is an emerging technology in civil aviation that is designed to provide the benefits of a HUD; however, the unique optical and physical characteristics of the HWD may change the existing levels of pilot performance and workload during SA CAT I operations flown with a HUD. When flying with a monocular HWD, binocular rivalry occurs, which may impact pilot performance and workload. This raises questions about whether pilot performance and workload are significantly impacted during manual SA CAT I flight operations, in which the pilot flying (PF) uses a monocular HWD with and without an EFVS. To address this concern, a study was carried out in which 11 pilot crews, made up of 22 Airline Transport Pilot (ATP) Captains, flew manual SA CAT I approach, landing, and rollout scenarios in a Boeing 737 Level D-equivalent flight simulator with a HUD and monocular HWD, with and without an EFVS, and in day and night ambient lighting conditions. Simulator motion was disabled to prevent interference with the HWD head tracking system. The PF rated their workload during each scenario using the National Aeronautics and Space Administration Task Load Index (NASA-TLX). The findings of the study suggest that a monocular HWD may not have a significant negative impact on a pilot's ability to manage most aspects of the flightpath during an SA CAT I operation; however, the monocular HWD elevated pilot workload. The monocular HWD also caused increased glideslope deviation during the instrument segment and increased deviation from the runway centerline during rollout. However, these increases were small and may not translate to operational significance. Pilots reported that the EFVS enhanced their awareness of the runway environment when transitioning from instrument to visual flight references; however, because there was not an EFVS kill switch for the PF, it was reported to interfere with natural vision when transitioning to flare, landing, and rollout, regardless of when implemented on a HUD and monocular HWD. Ultimately, this research contributes to the understanding of how SA CAT I operations that are flown with and without an EFVS may be impacted when pilots fly with a monocular HWD in lieu of a HUD.

Introduction

Low-visibility flight operations are among the most safety-critical in the National Airspace System (NAS). A Commercial Aviation Safety Team (CAST) analysis of worldwide Loss-of-Control in-flight (LOC-I) accidents and incidents found that lack of external visual references, which may be caused by darkness and/or Instrument Meteorological Conditions (IMC), was associated with a lack of aircraft attitude or energy state awareness in 17 of the 18 events analyzed in the study (Mumaw, Billman, & Feary, 2019). These risks are particularly prevalent during the final approach and landing phases of flight. 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). The risks of weather-related visibility restrictions to flight safety, along with their potential impact on NAS operational efficiency, demonstrate an opportunity to investigate the use of flight deck technologies as a safety and efficiency enhancement.

The Head-Up Display (HUD) is a well-established flight deck technology with a demonstrated track record of enhancing pilot performance and safety, particularly when flight visibility is restricted. The HUD is an installed aircraft system that presents information from flight instruments on a transparent screen fixed between the pilot and their view outside the aircraft, enabling them to maintain awareness of primary flight information—such as flightpath guidance, airspeed, and flight mode annunciations—while maintaining a forward line of sight along the flightpath. During an approach and landing, the use of a HUD has been shown to support pilot performance by improving flightpath tracking and airspeed management accuracy, enabling faster visual detection of the runway, and improving landing accuracy compared to when flying with traditional head-down displays (Fischer & Haines, 1980; Goteman et al., 2007; Weintraub et al., 1984). Pilots also 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 (Beringer, Domino, & Kamienski, 2018).

Because of the demonstrated safety enhancements of the HUD, the FAA authorizes pilots to fly manual approach and landing operations with reduced flight visibility using HUD guidance. For example, pilots are authorized to fly a Special Authorization Category I (SA CAT I) approach, landing, and rollout operation with as low as 1400 feet runway visual range (RVR) and a Decision Height (DH) of 150 feet when using flight guidance on a HUD, compared to a minimum of 1800 feet RVR and a Decision Altitude (DA) or DH of 200 feet for manual approach, landing, and rollout operations flown without a HUD.

The Enhanced Flight Vision System

Along with the HUD, an Enhanced Flight Vision System (EFVS) is another flight deck technology that may provide a flight crew with enhanced awareness of aircraft location and

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

trajectory relative to the runway touchdown zone (TDZ), potentially mitigating certain risks during approach and landing operations in low-visibility conditions. An EFVS is a certified, installed aircraft system that uses an electronic means to generate and display the forward external scene to the pilot using real-time imaging sensors, such as forward-looking infrared (FLIR), millimeter-wave radiometry, millimeter-wave radar, and low-light-level image intensification. The EFVS must present real-time sensor imagery, aircraft flight information, and flight symbology on a HUD or an equivalent display, so that the imagery, information, and symbology are clearly visible to the pilot flying in their normal position while they look forward along the flightpath (FAA, 2016).

One of the key operational benefits of an EFVS is that the real-time image of the external scene can extend the pilot's view of the forward scene beyond what natural vision would provide under a given atmospheric condition. This additional distance that the pilot can see using an EFVS, compared to what they would see without the EFVS, is operationally defined as visual advantage (FAA, 2022b; see Figure 1). Because of the visual advantage provided by an EFVS, a pilot may—under appropriate authorization and subject to the specific requirements of 14 CFR § 91.176—descend below the DH of an SA CAT I approach procedure, down to 100 feet above the TDZ, in cases where the required visual references at the DH cannot be seen with natural vision alone (see Figure 2).

In effect, the EFVS allows a pilot to identify the required visual reference identifications when descending below published minima, where a portion of the flight visibility prescribed by the Instrument Approach Procedure (IAP) being flown is satisfied by the visual advantage provided by the EFVS. In addition to the authorization to descend below the DA without natural vision of the required references, lower RVR minima may be authorized, depending on operator-specific approvals, EFVS performance, airfield lighting, and the characteristics of the instrument approach being flown. If the EFVS provides a sufficient visual advantage and all prerequisite requirements are met, pilots may be authorized to fly an SA CAT I operation with a TDZ RVR as low as 1000 feet, compared to 1400 feet for an SA CAT I operation flown without an EFVS (Straight-in landing operations below DA/DH or minimum descent altitude [MDA] using an EFVS under IFR, 2017).

Figure 1

Diagram of Enhanced Flight Vision System Visual Advantage

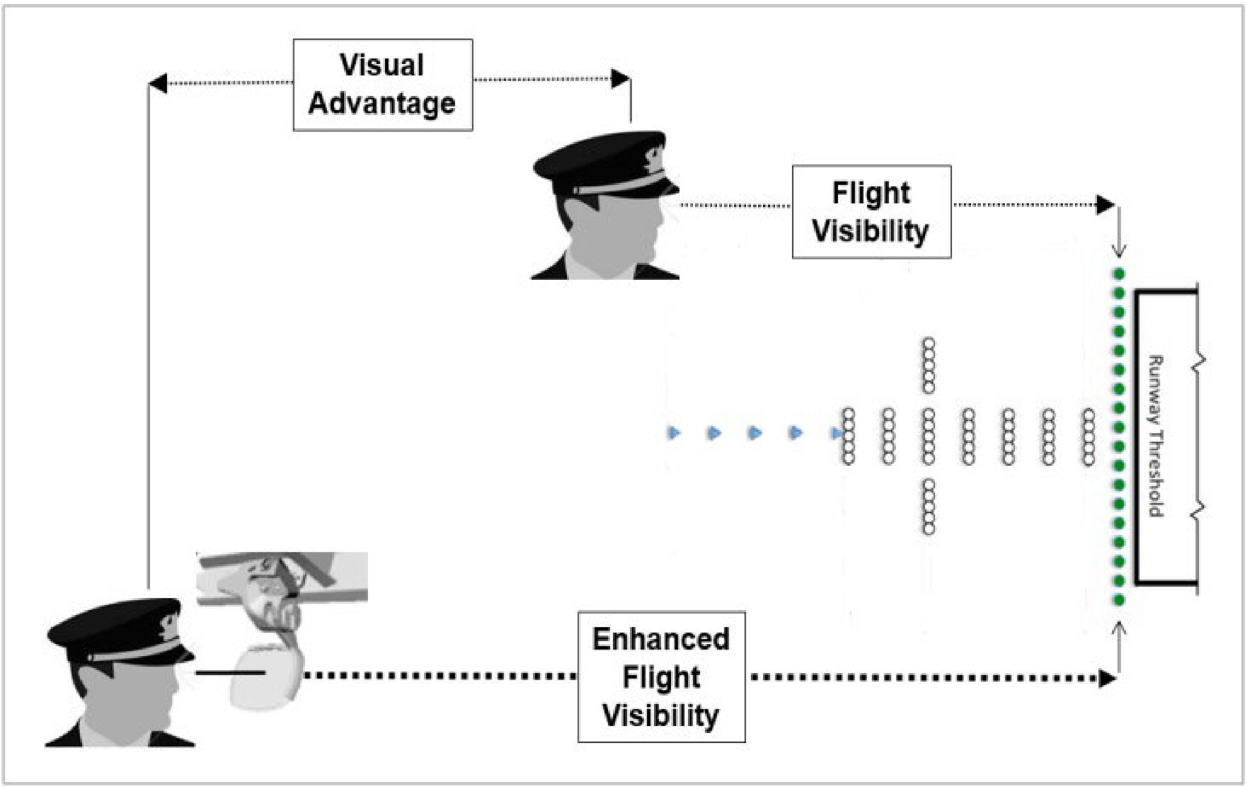


Figure is from "Enhanced Flight Vision System Operations" by the Federal Aviation Administration, 2022 (Advisory Circular No. 90-106B). In the public domain.

Figure 2

Diagram of Special Authorization Category I Enhanced Flight Vision System Operation

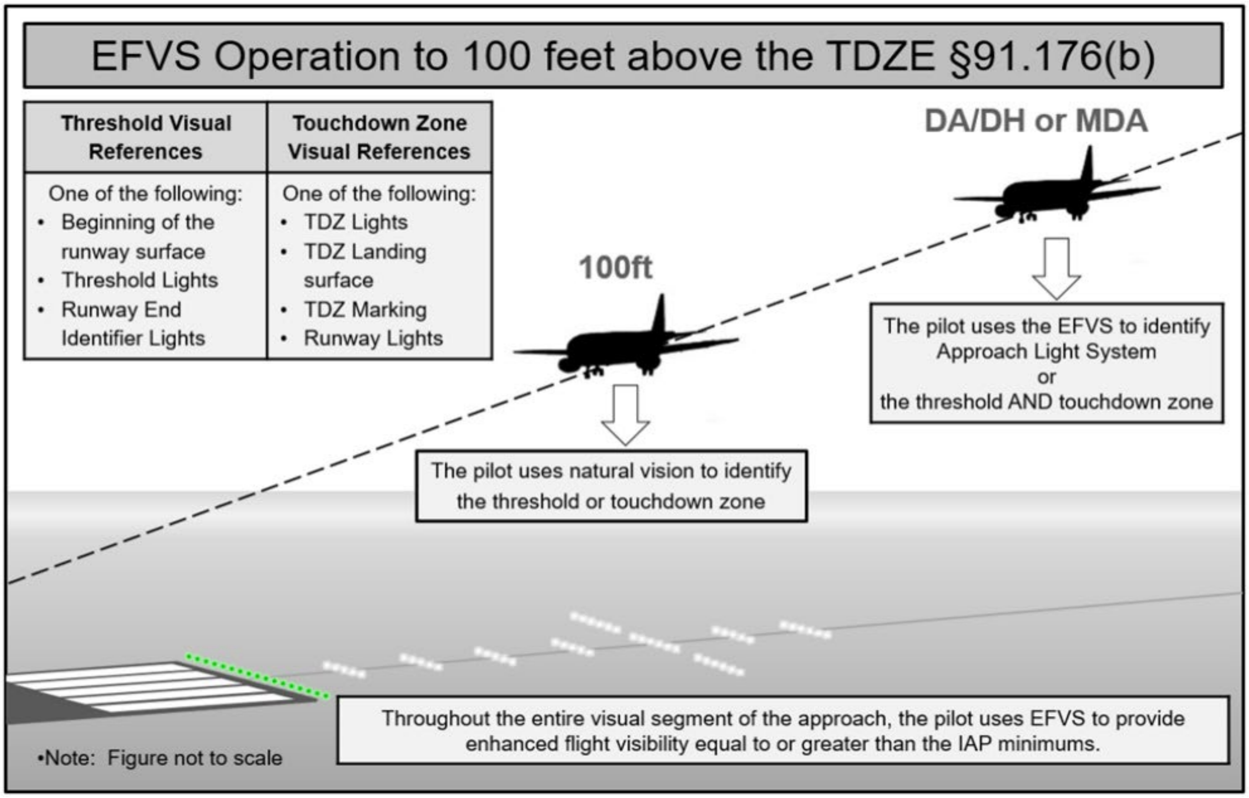


Figure is from "Enhanced Flight Vision System Operations" by the Federal Aviation Administration, 2022 (Advisory Circular No. 90-106B). In the public domain.

Implementation of a Head-Worn Display in Lieu of a Head-Up Display

One barrier that may prevent aircraft operators from being authorized to conduct manual low-visibility flight operations, such as the SA CAT I approach with or without an EFVS, is that such operations often require a HUD to be installed in the aircraft. Installing a HUD may not be a feasible option in all aircraft types, such as those with space restrictions on the flight deck. To address this challenge, there have been recent efforts to implement a Head-Worn Display (HWD) in lieu of a HUD. 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. This provides the pilot with greater freedom of motion compared to when using a HUD, enabling them to view symbology regardless of head position. In certain applications, the HWD may also offer an expansive field of regard, where additional visual information appears when the pilot turns their head to the side or tilts their head up or down (Velger, 1998). A primary motivation behind the development and implementation of HWDs is to expand the availability of existing HUD and EFVS operational benefits to a larger population of end users, including operators of space-restricted aircraft (FAA, 2022a).

The HWD is considered an evolution of the HUD because, like the HUD, it superimposes flight information onto out-the-window visual information, yet it improves upon the HUD by offering expanded field of view (FOV)[2] and field of regard (FOR),[3] reduced form factor, and lower cost (2022b). These design differences between the HUD and HWD challenge the notion that the HWD is equivalent to the HUD from an operational standpoint. The nearer distance of the HWD combiner from the pilot's eyes may impact the pilot's ability to view other information on the flight deck (Newton et al., 2026). Some wide field-of-view HWDs may present certain information at a focal distance nearer than optical infinity to minimize interference with flight deck controls and instruments (Thomas, 2009). HWD systems present flight symbology and sensor imagery to both eyes (i.e., binocular) or to one eye (i.e., monocular). Previous research demonstrates that there are perceptual problems caused by binocular rivalry when the HWD is monocular, which may impact the pilot's ability to use the symbology or sensor imagery (Patterson et al., 2007). Monocular HWDs have also been found to increase pilot workload during manual flight operations compared to binocular HWDs, which could negatively impact performance when pilots fly with EFVS imagery on an HWD (Newton et al., 2026).

Current Study and Research Questions

Because of the distinct design characteristics of the HWD relative to the HUD, human factors research is needed to inform operational evaluation criteria for the use of HWD technologies during low visibility concepts of operation, such as an SA CAT I operation with and without an EFVS. FAA operational authorizations for SA CAT I approach, landing, and rollout operations currently allow approaches with as low as 1400 feet RVR when using primary flight symbology on a HUD and as low as 1000 feet RVR when using an EFVS on a HUD. It is important to determine if an HWD provides the same benefits as a HUD in low-visibility conditions and, therefore, can be granted operational approval for use during SA CAT I operations with and without an EFVS. With HWD integration onto the flight deck, this operational approval could allow for increased NAS throughput without compromising aviation safety. Because of the distinct characteristics of the HWD, human factors research is needed to support the expanded use and approval of emerging HWD technologies during low visibility concepts of operation, including the conduct of SA CAT I approach, landing, and rollout operations with and without an EFVS. This research could also inform decisions that increase the number of viable airports/runways for low-visibility flight operations.

The present research was conducted to evaluate pilot performance and workload when the Pilot Flying uses a monocular HWD during a manual[4] SA CAT I approach, landing, and rollout operation with and without an EFVS. Pilot Flying performance and workload under these conditions will be compared between the use of a monocular HWD and the use of a HUD, as well as between the use of an EFVS and the use of Primary Flight Symbology. This research has the potential to inform operational credit changes that would allow new HWD technologies to be used for reduced visibility operations. As such, the purpose of this study is also to provide

[2] 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.).
[3] FOR refers to the range of virtual environment presented by an XR headset that can be viewed with physical head and body rotation (Ragan et al., 2015).
[4] In the context of this study, "manual" refers to a manual flight operation where the pilot is managing the flightpath while physically controlling pitch, roll, yaw, and thrust with the autopilot and autothrottle disengaged (FAA, 2022c). Other combinations of automation (e.g., Flight Director [FD]) may be enabled.

the FAA Flight Standards Service with data that can be used to create a standardized methodology for evaluating HWD systems with or without an EFVS for human factors considerations that are relevant for operational credit.

Based on these needs, the present research was conducted based on the following research questions. Compared to the approved manual SA CAT I approach, landing, and rollout operations in day and night ambient lighting conditions using a HUD with and without an EFVS:

1. What degree of flightpath and energy management accuracy is present during a manual SA CAT I operation using a monocular HWD with and without an EFVS?
2. What degree of TDZ dispersion at landing is present during a manual SA CAT I operation using a monocular HWD with and without an EFVS?
3. What degree of runway centerline deviation during rollout is present during a manual SA CAT I operation using a monocular HWD with and without an EFVS?
4. What level of workload does the Pilot Flying experience during a manual SA CAT I operation using a monocular HWD with and without an EFVS?

Method

In this research, current Airline Transport Pilot (ATP) Captains flew a series of simulated SA CAT I approach and landing scenarios in day and night ambient lighting conditions with a HUD and a monocular HWD that displayed flight symbology only or an EFVS. All scenarios involved normal, manual flight operations, where no non-normal system failures or environmental hazards occurred, and the PF managed the flightpath using Flight Director (FD) guidance while manually controlling pitch, roll, yaw, and thrust. Objective pilot performance measures were implemented to evaluate the ability of the PF to manage the flightpath and target airspeed of the approach without an autopilot or autothrottle, as well as touchdown performance and directional control during rollout. Subjective mental workload ratings were taken after each flight scenario to facilitate evaluation of the relative impact of ambient lighting, EFVS mode, and display types on task demands during the scenarios.

The HUD in this study was a production-quality, collimated display that presented symbology at a focal distance of optical infinity, whereas the monocular HWD was a non-collimated Microsoft HoloLens 2 with an image focal distance of approximately six feet. 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 flight deck HUDs and HWDs, the use of a non-collimated HWD may limit the ability to generalize findings from this research to all flight deck HWD systems.

Throughout the development and execution of this research, the authors worked with multiple Boeing 737 Type-Rated Pilot Subject Matter Experts (SMEs) 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. This ensured that the simulated flight scenarios were appropriate for the study and that the Dependent Variables (DVs) accurately represented real-world pilot performance and workload outcomes. A pilot study was conducted prior to executing the main study. The purpose of this pilot study was to ensure that the study procedure was appropriately designed, and all data were reliably collected and processed. A combination of quantitative data and qualitative feedback from the pilot study participants was used to make these determinations.

Participants

Eleven Part 121 flight crews participated in this study, with each flight crew consisting of two ATP Captains. All 22 participants were male with an average age of 56.59 years (SD = 8.41 years). Participants were paired into two-person flight crews prior to the study session based on scheduling availability, with each crew consisting of Captains from the same operator (i.e., airline). Each participant (a) held a Boeing 737 Type Rating; (b) was qualified and current according to their operator and FAA requirements as a pilot-in-command of a Boeing 737; (c) was current in using a HUD in all phases of flight; (d) had experience using a HUD within 45 days of participating in the study; and (d) had at least 100 flight hours of HUD use. All participants received monetary compensation for participating in the study.

None of the participants in this study were formally trained in EFVS operations. There were several reasons for not including EFVS-trained flight crews in this research. First, EFVS-trained 737 flight crews are rare, so sampling from that population for this research may have led to an inadequate sample size. Second, prior research that directly compared flight crews who had EFVS training to those who had no EFVS training found no operationally significant differences in performance between the two groups (Beringer et al., 2019).

All participants completed the hole-in-the-card eye dominance test during the pre-experiment briefing. Sixteen of the 22 participants (72.7%) were right-eye dominant, while the remaining six participants (27.3%) were left-eye dominant. Thirteen (59.09%) participants reported wearing corrective lenses while flying.[5] Participants reported an average of 27,875.32 total flight hours (*SD* = 47,921.36), with an average of 42.64 of those hours (*SD* = 28.39) occurring within one month of participating in the study.

Nine (40.91%) of the participants reported having previous experience using an HWD system, primarily through military experience or participation in a prior FAA HWD study.[6] Nine (40.91%) participants indicated they had prior experience flying with a real-time imaging system[7] (e.g., enhanced vision system [EVS], FLIR). The reported total number of approaches participants had flown while using a HUD ranged from 10 to 6500 (*M* = 1245.23; *SD* = 1983.03), with an average of 17.09 approaches (*SD* = 22.87) occurring within the last 30 days prior to participating in the study. Participants reported on average that 44.59 (*SD* = 79.98) of their total number of HUD approaches occurred below standard CAT I minimums. When asked to report the lowest RVR the participants remembered flying, the RVR ranged from 100 feet to 1200 feet. The most frequently reported lowest RVR was 600 feet (*n* = 12; 54.5%).

---

[5] Participants reported wearing corrective lenses when flying included monofocal lenses (n = 5, 22.73%); progressive lenes (n = 5, 22.73%); Contact Lenses (n = 2, 9.09%); and Bi/trifocal lenses (n = 1; 4.55%).
[6] The participants reported having previously used an HWD or helmet-mound display in previous FAA studies (n = 6, 27.27%); system in a F16 (n = 1, 4.55%); system in an attack helicopter (n = 1, 4.55%); and binocular night vision goggles (n = 1, 4.55%).
[7] Participants reported experience with night vision googles (n = 4, 18.18%); FLIR (n = 2, 9.09%); previous FAA studies (n = 2, 9.09%); Other responses included military operations with a helicopter (n = 1, 4.55) and F-16/F-15E (n = 1, 4.55). Some responses included previous experience with multiple system types and is the reason the number of responses sums to greater than the number of '*other*' responses.

Thirteen (59.09%) of the participants indicated their operator requires the use of a HUD beyond the minimum FAA requirements. When the Pilots were asked on a 10-point Likert scale to describe their HUD usage from 1 being "*use as little as possible*" to 10 "*always deployed,*" the average response was 8.82 (*SD* = 2.02). When asked on a 10-point Likert scale how they would feel if the HUD was added to the Minimum Equipment List (MEL) on their next flight from 1 ("*I would feel anxious*") to 10 ("*I would feel just fine*")*,* 8.00 (*SD* = 2.29) was the average response.

Research Design and Independent Variables

This study was designed to determine whether Pilot Flying (PF) performance and workload differ depending on three Independent Variables (IVs): (a) display type (HUD or monocular HWD), (b) EVFS mode (on or off), and (c) ambient lighting (day or night). Toward this end, each crew flew 16 Instrument Landing System (ILS) approach and landing scenarios, with each participant completing eight scenarios as PF from the left seat of the simulator. All scenarios involved normal operations, in which the crews completed a routine approach and landing without the occurrence of aircraft system failures or environmental hazards that may have prevented a landing. An overview of the research design, including all IVs and DVs, is provided in Table 1.

Table 1

Research Design with Independent Variables and Dependent Variables

<table>
<tr><th></th><th colspan="2">IV_2: EFVS Mode</th><th></th></tr>
<tr><th>IV_1: Display Type</th><th>Off</th><th>On</th><th>IV_3: Ambient Lighting</th></tr>
<tr><td rowspan="2">HUD</td><td colspan="2" rowspan="4">Dependent Variables
Instrument Segment Performance
• DV_1: Flightpath Deviation during the Instrument Segment
• DV_2: Airspeed Deviation during the Instrument Segment
• DV_3: Flightpath Deviation at the Decision Height
• DV_4: Airspeed Deviation at the Decision Height
Visual Segment Performance
• DV_5: Flightpath Deviation at 100 feet Above the TDZE
• DV_6: Airspeed Deviation at 100 feet Above the TDZE
• DV_7: Flightpath Deviation at Threshold Crossing
• DV_8: Airspeed Deviation at Threshold Crossing
Landing and Rollout Performance
• DV_9: Centerline Deviation at Touchdown
• DV_{10}: Distance from Runway Threshold at Touchdown
• DV_{11}: Sink Rate at Touchdown
• DV_{12}: Centerline Deviation during Rollout
Pilot Flying Workload
• DV_{13}: NASA-TLX Total Weighted Score
• DV_{14}: NASA-TLX Subscale Scores
Supplemental Measures
• Frequency of Missed Approaches Per Condition
• Demographics Questionnaire
• Usability Questionnaire
• Verbal Feedback</td><td>Day</td></tr>
<tr><td>Night</td></tr>
<tr><td rowspan="2">Monocular HWD</td><td>Day</td></tr>
<tr><td>Night</td></tr>
</table>

IV1: Display Type

Pilot performance and workload were evaluated as a function of whether they flew with a HUD or with a monocular HWD. In the HUD condition, flight symbology was presented to the PF using the HUD that was installed in the aircraft's left seat. In the monocular HWD (M. HWD) condition, the PF wore the M. HWD and viewed symbology that was presented to their dominant sighting eye while sitting in the left seat. In both cases, the PM sat in the right seat without a HUD or M. HWD. As stated elsewhere in this report, the HUD in this study was a production-quality, collimated display with an image focal distance of optical infinity, whereas the monocular HWD was a Microsoft HoloLens 2 with a non-collimated display with an image focal distance of approximately six feet. 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, the use of a non-collimated HWD in this research could limit the ability to generalize data in this report to all HWD systems developed for use on the flight deck.

IV_2: EFVS Mode

Pilot performance and workload were evaluated as a function of whether they flew with flight symbology only or with an EFVS displayed on the HUD or monocular HWD. In the EFVS off condition, the HUD and monocular HWD presented symbology from the Collins HGS-6000 in Primary mode throughout the scenario. In the EFVS on condition, the HUD and monocular HWD presented an EFVS (i.e., EVS imagery and Primary mode HGS-6000 symbology) throughout the scenario. The simulator was not equipped with an EFVS kill switch on the yoke, so the pilots did not have the ability to disable the EFVS sensor image during the scenarios.

IV_3: Ambient Lighting

Pilot performance and workload as a function of display type and EFVS mode were evaluated in day and night ambient lighting conditions. The time of day was set to 15:00 local time in the day condition and 24:00 local time in the night condition.

Dependent Variables

The DVs in this study included a combination of objective and subjective measures that evaluated the impacts of display type, EFVS mode, and ambient lighting on PF performance and workload during SA CAT I approach, landing, and rollout. The objective measures included (a) performance during the instrument segment of the approach; (b) performance at critical points during an EFVS SA CAT I operation (i.e., DH and 100 feet above the TDZE); and (c) landing and rollout performance. The NASA-Task-Load Index (NASA TLX; Hart & Staveland, 1988) was used to collect PF workload ratings during each scenario. The frequency of missed approaches per condition, demographic information, usability questionnaire feedback, and verbal think-aloud feedback captured from inside the simulator supplemented these performance and workload measures. Each of the DVs is described in detail in the following sections.

DV_1: Flightpath Deviation

In this study, flightpath deviation was defined as the linear distance, in feet, between the actual flightpath and the intended flightpath, and was evaluated separately for lateral deviation (i.e., ILS localizer deviation) and vertical deviation (i.e., ILS glideslope deviation). Flightpath deviation was measured during multiple phases of the approach. Deviation during the instrument segment was measured by calculating Root-Mean-Square (RMS) deviation from the ILS localizer and glideslope from simulator release to 300 feet above the TDZE (see Equation 1). Additionally, snapshots of instantaneous lateral and vertical flightpath deviation were captured at the DH (i.e., 157 feet above the TDZE), at 100 feet above the TDZE, and at runway threshold crossing.

Equation 1
Calculation of Root Mean Square Flightpath Deviation

$$Flightpath_{RMS} = \sqrt{\frac{\sum_{i=1}^{n}(d)^2}{n}}$$

Where:

d = instantaneous deviation in feet from center of flightpath (i.e., localizer or glideslope)
n = number of data points capturing instantaneous flightpath deviation

DV_2: Airspeed Deviation

Airspeed deviation was evaluated in terms of actual Indicated Airspeed (IAS) relative to target IAS programmed into the Mode Control Panel (MCP) (i.e., 150 knots, including a 10-knot gust correction). Airspeed deviation was evaluated at multiple phases of the approach, including RMS deviation during the instrument segment (simulator release to 300 feet above the TDZE), as well as snapshots of airspeed deviation captured at the DH (i.e., 157 feet above the TDZE), at 100 feet above the TDZE, and at runway threshold crossing. Airspeed deviation in the instrument segment was calculated as RMS deviation, which is shown in Equation 2.

Equation 2
Calculation of Root Mean Square Airspeed Deviation

$$Airspeed_{RMS} = \sqrt{\frac{\sum_{i=1}^{n}(d)^2}{n}}$$

Where:

d = instantaneous difference between observed IAS and target IAS
n = number of data points capturing IAS

DV_3: Distance From Runway Threshold at Touchdown

The distance down the runway at which the aircraft landed was evaluated by measuring the longitudinal distance, in feet, between the aircraft center of gravity and the runway threshold at touchdown.

DV_4: Sink Rate before Touchdown

The average vertical velocity of the aircraft, measured in feet per second, was measured during the two seconds leading up to touchdown of the main landing gear.

DV_5: Centerline Deviation at Touchdown and During Rollout

Centerline deviation was evaluated at touchdown and during rollout. The ability of the PF to land on the runway centerline at touchdown was evaluated by measuring the distance, in feet, between the aircraft center of gravity and the runway centerline at touchdown. The ability of the PF to track the runway centerline during rollout was evaluated by measuring the distance, in feet, between the aircraft center of gravity and the runway centerline from touchdown to 25 knots ground speed. Before conducting inferential analyses, raw deviation data were transformed into RMS per Equation 1.

DV_6: Pilot Flying NASA-TLX Workload Rating

PF workload was assessed using the NASA-TLX (Hart & Staveland, 1988). The NASA-TLX is a multi-dimensional workload rating scale that integrates the weighted subjective responses driven by perceptions of task demand. The NASA-TLX rating scale was administered at the end of each scenario to the PF (the NASA-TLX rating scale was not administered to the Pilot Monitoring [PM]). After each participant's final scenario as PF, they completed pairwise comparisons of NASA-TLX subscales to evaluate the dimensions of workload that contributed the most to their workload when they were the PF. Responses to the pairwise comparisons were used to weight the total workload rating based on the procedure described in Hart and Staveland (1988).

Supplemental Measures

Frequency of Missed Approaches

Each scenario documented whether the flight crew conducted a missed approach or continued to a landing. The relative number of missed approaches versus landings per condition was evaluated to determine the frequency of missed approaches conducted as a function of display type, EFVS mode, and ambient lighting.

Usability Questionnaire

The authors of this report developed a usability questionnaire for this study based on usability feedback from pilot SMEs during technical working sessions in the simulator, as well as operational suitability criteria outlined in the EFVS Evaluation Aid developed for the FAA Aircraft Evaluation Division (AED; Fercho & Watson, 2024). Each participant rated the usability of the monocular HWD relative to the HUD, as well as the usability of the EFVS versus primary flight symbology only, in the four-part questionnaire, which included 28 5-level Likert items and four text boxes for entering optional open-ended responses. The response options for the Likert scale items were 1 ("Much Worse"), 2 ("Somewhat Worse"), 3 ("About the Same"), 4 ("Somewhat Better"), and 5 ("Much Better"). Usability was evaluated based on each participant's own perception of their ability to:

1. Follow flightpath guidance
2. Maintain target airspeed
3. Visually acquire approach/runway lighting
4. Visually acquire runway paint markings

5. Land the aircraft
6. Track the centerline during rollout
7. Evaluate the safety of the runway environment

The usability questionnaire was administered during the debriefing session using an iPad with Qualtrics. See Appendix A for the complete usability questionnaire.

Simulator and Display Types

Simulator

Simulator scenarios were conducted in a CAE Boeing 737-800 Level D-equivalent Full Flight Simulator, operated by the FAA Flight Research and Analysis Group (AFS-430) at the Mike Monroney Aeronautical Center in Oklahoma City, OK. 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 day/night out-the-window visual model, a comprehensive weather and wind modeling system, and dynamic loading of the flight controls. The out-the-window visual system in the simulator used collimated projectors, so visual information out the windscreen appeared at optical infinity. 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. The simulator was equipped with large-format displays that emulate the Head-Down Displays (HDDs) in the Boeing 737 MAX (see Figure 3). During the study, an iPad was mounted below the left outboard window, which participants used to complete the NASA-TLX rating scale and pairwise comparison items during the study session.

Figure 3
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 4) and a Microsoft HoloLens 2 non-collimated, commercial off-the-shelf augmented reality headset was used as the HWD (see Figure 5). The symbology in the HUD and HWD was set to Primary mode in all scenarios, which included a standard HUD symbology set, a flight director and flightpath vector, and a flare prompt.

Figure 4
Collins Aerospace HGS-6000 Head-Up Display in Simulator

The HoloLens 2 featured a 78% binocular overlap, and the flight symbology fit fully within the binocular region of the total FOV. 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. The HoloLens 2 included a Moving Platform Mode, which allowed for the device to be used in moving environments. 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 between 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 5
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 display types. 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 symbology on a HUD or equivalent display 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 likely not produce confounding effects on pilot performance and workload in this study (Bailey et al., 1989, 2004, 2005; Johns & Funk, 1991).

Enhanced Flight Vision System

The EFVS in the simulator models a Kollsman II EVS camera with an infrared (IR) spectrum of 1 to 5 microns. The simulator EFVS was programmed to a factor of 1.4, which resulted in a visual advantage of 400 feet when TDZ RVR was 1000 feet. The EFVS was presented in identical fashion between the HUD and HWD, and the brightness/gain of the EFVS video remained at the same level across all scenarios. In scenarios where the EFVS was enabled, it remained enabled for the duration of the scenario, and the simulator was not equipped with a kill switch on the yoke for the PF to disable the EFVS sensor image. It is important to note that the HWD range of brightness was limited compared to that of the HUD. Therefore, in order to maintain equivalent perceived brightness across all scenarios, the HUD brightness was set at a higher brightness than pilots might normally use, particularly when ambient lighting levels are low. Figure 6 - Figure 9 present images taken through the HUD with the EFVS on and off in the day and night ambient lighting conditions.

Figure 6

Image of HUD with EFVS off in the Day Ambient Lighting Condition with TDZ RVR of 1400 ft

Figure 7

Image of HUD with EFVS on in the Day Ambient Lighting Condition with TDZ RVR of 1000 ft

Figure 8

Image of HUD with EFVS off in the Night Ambient Lighting Condition with TDZ RVR of 1400 ft

Figure 9

Image of HUD with EFVS on in the Night Ambient Lighting Condition with TDZ RVR of 1000 ft

Approach, Landing, and Rollout Scenarios

Each flight crew flew 16 ILS approach and landing scenarios under Instrument Flight Rules (IFR) beginning six miles from the runway threshold and ending when the aircraft had decelerated to 25 knots ground speed on the runway surface during rollout. At the beginning of each scenario, the aircraft was aligned with the localizer and glideslope, traveling at the target airspeed (150 knots, including a 10-knot gust correction) and configured for landing, and control was transferred to the PF. All scenarios were conducted without automated flight control systems (e.g., autopilot, autothrottle, and autoland) and with FD guidance on the HUD and monocular HWD.

Runway Environment and Approaches

All scenarios involved flying the published SA CAT I ILS approach procedure into Portland International Airport (PDX) RWY 10R. Runway lighting infrastructure and runway markings remained consistent across all scenarios, with the airfield equipped with high-intensity runway lighting (HIRL), TDZ lighting, centerline lighting (CLL), and an Approach Lighting System with Sequenced Flashing Lights (ALSF-2).

Weather and Winds

For all scenarios, the cloud ceiling was set at 200 feet above the TDZE and cloud tops at 3,000 feet above the TDZE. Within the clouds, visibility was 0 feet. The reported TDZ RVR in each scenario was dependent upon whether the EFVS was ON or OFF during that scenario. For EFVS off scenarios, reported TDZ RVR was set at 1400 feet. For EFVS on scenarios, reported TDZ RVR was set at 1000 feet. The EFVS was tuned so that it provided a visual advantage of 400 feet. This approach yielded constant visibility from an operational standpoint for the PF across all scenarios, with the source of that visibility dependent on whether the EFVS was ON or OFF.

Throughout each approach, gusting winds with variable direction and velocity were present from the start of each scenario until the aircraft reached approximately 375 feet AGL, the purpose of which was to increase the difficulty of maintaining a stabilized approach to determine if flying with a monocular HWD in lieu of a HUD with or without an EFVS 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 5 knots from 90° offset from the runway heading with gusts up to 8 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 15 knots with gusts up to 23 knots. The wind direction then shifted back so that when the aircraft descended below 375 feet AGL, the gusting winds with variable wind heading dissipated, resulting in sustained 14 knot winds offset 60° left or right of the runway heading (12.12 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 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.

Study Procedures

The FAA Civil Aerospace Medical Institute Institutional Review Board approved all study procedures prior to data collection. Each study session lasted approximately 2.5 hours, including breaks. During this time, participants completed an initial briefing, a simulator familiarization session, an alignment eye dominance test, an experimental session with 16 ILS approach and landing scenarios, and a post-experiment debriefing (Table 2).

Table 2
Approximate Schedule of Study Session

Activity	Pilot Flying	Start Time	End Time	Duration (mins)
Pre-Experiment Briefing		0800	0830	30
Familiarization Session		0830	0900	30
Break		0900	0910	10
Session 1 (4 Scenarios)	Participant 1	0910	0925	15
Session 2 (4 Scenarios)	Participant 2	0925	0940	15
Break		0940	0950	10
Session 3 (4 Scenarios)	Participant 1	0950	1005	15
Session 4 (4 Scenarios)	Participant 2	1005	1020	15
Debriefing		1020	1030	10
			Total Time	2.5 hours

Research Briefing

Upon arrival, the participant was provided with an Informed Consent form that outlined their rights, responsibilities, and the purpose of the study. Following their written consent to participate, participants completed a demographics questionnaire. This questionnaire asked each participant to report basic demographic information, aviation experience (e.g., total flight hours, total flight hours with a HUD, experience with low-visibility approach and landing operations), and information about their vision (i.e., type of corrective lenses used). A detailed study briefing followed, during which the Principal Investigator (PI) and a Boeing 737 Type-Rated FAA Pilot described the purpose of the study, the approach and landing operations they would be conducting, the EFVS operational concept, and the procedures to be followed during the simulator sessions. Part of this briefing included an explanation of the performance expectations during the approach and landing operations, including (a) maintaining the vertical and lateral profiles of the approach as best as possible, (b) maintaining the target airspeed of the approach as best as possible, and (c) landing as close to the runway touchdown point as possible with minimal sink rate. The PI also described the NASA-TLX, including its subscales and when the pilots were expected to complete their responses. They were given an opportunity to handle a demonstration unit of the HoloLens 2 headset and become familiar with the adjustment points. Cumulatively, the research briefing took approximately 30 minutes.

Eye Dominance Testing

At the end of the briefing session, each participant was tested for sighting eye dominance using the distance hole-in-the-card test (e.g., see Johannson et al., 2015). For this test, the participant was given an 8.5” x 11” piece of black cardstock with a 3-centimeter diameter circular hole cut in the center of the card. They were instructed to hold the card at arm’s length while viewing a single letter with both eyes open, positioned 10 feet in front of the participant. The PI then asked the participant to bring the card close to their face while maintaining visibility of the letter. The eye that was aligned with the hole in the card was recorded as the dominant eye. For the experimental session, the monocular HWD was configured to present symbology to that participant’s dominant eye.

Simulator Familiarization

After the initial briefing and sighting eye dominance test, the participants entered the simulator. A Boeing 737 Type-Rated FAA Pilot gave each participant a flight deck walk-through and familiarization session to acquaint them with the controls, displays, and iPad mounted below the left outboard window. The participants then completed a practice session designed to familiarize them with the specific approach and landing scenarios used in this study, as well as the combinations of display type and EFVS mode implemented in the study. In this session, each participant completed two approach and landing practice scenarios as PF from the left seat, and a second set of two as PM from the right seat. Practice scenarios began with the aircraft positioned 6 miles from the runway threshold. Before each HWD scenario, the Microsoft HoloLens calibration process was performed to ensure that the symbology was aligned. Table 3 lists the Familiarization Session scenarios.

Table 3
List of Familiarization Session Scenarios

Scenario No.	Display Type	EFVS Mode	RVR/Ceiling (feet)	Ambient Lighting	Pilot Flying	Wind Dir.	Gust Model
1	HUD	Off	1400/200	Day	Participant 1	Left	1
4	M. HWD	On	1000/200	Day	Participant 1	Left	2
			Participants Switch Seats				
1	HUD	Off	1400/200	Day	Participant 2	Left	3
4	M. HWD	On	1000/200	Day	Participant 2	Left	4

All participants completed the familiarization session scenarios in sequential order.

Experimental Session

Following the familiarization session, each participant completed eight ILS approach and landing scenarios as the PF from the left seat of the simulator and another eight scenarios as PM from the right seat, for a total of 16 scenarios per crew. The PI was seated behind the crew in the simulator jump seat. The participant who was acting as PF operated the flight controls from the left seat and used the HWD or HUD. The participant who was acting as PM during the scenarios sat in the right seat and performed monitoring duties without an HWD or HUD throughout each approach as per their company's Standard Operating Procedures (SOPs), FAA regulations, and personal techniques.

Before each scenario, the PI briefed the participants on the display type in use and whether the EFVS would be used for the upcoming scenario. The participant who acted as PM then completed the approach checklist. The approach procedure chart for each scenario was available on the back side of the laminated pre-approach briefing checklist. When the scenario required the use of the HUD, the PF was asked to remove the HWD, hand it to the PI, and deploy the HUD combiner. When the scenario required the use of the HWD, the PF was asked to put on the HWD and stow the HUD combiner. Each time the HWD was used for a scenario, the PF was asked to verify the alignment of the symbology by ensuring that the HWD symbology was fully overlapped with the HUD symbology; adjustments were made as necessary by the simulator engineering team from the remote operating station. When the checklist was completed and the HWD alignment had been verified, the PI cleared the participants for landing and started the simulation.

After the approach and landing was completed, the PF brought the aircraft to a stop on the runway. The simulator automatically paused the current scenario and began loading the next scenario after the aircraft had decelerated to 25 knots on the runway surface. At this point, the PF hit the “NEXT” button on the iPad Qualtrics screen to advance to the NASA-TLX rating scale. Each scenario, which includes the pre-flight approach briefing, approach and landing, and NASA-TLX rating scale, took approximately four minutes to complete. After each participant completed their final scenario as PF and had completed the NASA-TLX rating scale for that scenario, the iPad Qualtrics screen advanced to the pairwise comparisons portion of the NASA-TLX.

Scenario Sequence and Session Structure

Each of the IVs was fully permutated within each participant so that all participants experienced every level of each IV an equal number of times. The sequence of display type and EFVS mode was counterbalanced using a Latin Square design across all scenarios to control for order and carryover effects in the data. Ambient lighting was counterbalanced in blocks across all crews so that the first set of eight scenarios for 50% of crews was day, and the second set of eight was night; for the remaining 50% of crews, the first set of eight was night and the second set of eight was day. The 16 scenarios were structured into four sessions of four scenarios each. The participant who acted as a PM in the preceding session became the PF and vice versa.

Post-Experiment Procedures

After the final block of the experimental session, the participants and PI exited the simulator and returned to the briefing room, where they completed the Usability Questionnaire. The PI and FAA Research Pilot then debriefed with the participants and closed out the study session.

Results

PF performance and workload were assessed during approach and landing across two display type conditions (HUD/monocular HWD), two EFVS modes (on/off), and two ambient lighting conditions (day/night). Before conducting the analyses, the data were examined for outliers and checked to ensure they met the assumptions required for the Repeated Measures Analysis of Variance (RM ANOVA). The assumption of sphericity was considered met for this design as all IVs have only two levels. If factors have only two levels, there is a single set of difference scores and one variance, meaning that unequal variances are not possible. All statistical analyses were conducted using both the Statistical Package for Social Science (SPSS version 28.0) and RStudio (version 4.2). The Shapiro-Wilk test was employed to assess the normality assumption of the residuals. When the assumption of normality was violated, a Box-Cox transformation was applied to the raw data (Malik et al., 2018).

If the normality assumption remained unmet after applying the Box-Cox transformation, the Box-and-Whisker plots of the transformed data were examined for potential outliers. If a potential outlier seemed to be causing the deviation from normality, further investigation was conducted to identify potential outliers, and normality was reassessed after their removal. For each model, information regarding the application of the Box-Cox transformation and the exclusion of outliers to address non-normality issues is provided. However, caution is advised when interpreting the results of models that violate the normality assumption, even after applying the Box-Cox transformation. For all comparisons, a significance level (α) of less than .05 was considered statistically significant. Adjusted Means (M_{adj}) and corresponding line plots represent transformed data with 95% confidence intervals unless stated otherwise.

This study also sought to determine the degree to which the performance levels associated with flying SA CAT I operations with monocular HWD were statistically equivalent to those when flying with a HUD. In this research, a non-significant result of the traditional parametric statistics would provide evidence that there is no significant difference in pilot performance between flying with a HUD and flying with a monocular HWD. To investigate whether performance when flying with a HUD was significantly equivalent to performance when flying with a monocular HWD, the two one-sided test (TOST) procedure (Lakens, 2017) was conducted as a follow-up test to any non-significant main effect for display type for pilot performance DVs. The following section describes the TOST procedure in detail, along with guidelines for interpreting the test outcomes.

Two One-Sided *t*-Test

Non-significant *p* values are traditionally interpreted as having a “null” effect (i.e., no significant effect of the IV was found). Traditional hypothesis testing typically sets the alternative hypothesis as identifying a significant difference between groups (between-subjects designs) or conditions (within-subject designs), and the null hypothesis as failing to identify a significant difference. This method does not evaluate the hypothesis that conditions are equal or “significantly equivalent” (Hauck & Anderson, 1984).

The TOST method is a statistical approach used to test for equivalence between two groups. Rather than testing for a difference, it checks whether the observed effect size falls within a pre-defined equivalence range, showing that any observed difference is small enough to be considered practically irrelevant. Significant equivalence is frequently discussed in the scientific literature and in aviation safety research. For example, Schuirmann (1987) employed the TOST method to determine equivalent efficacy between a new medical drug and one that was already approved and on the market. Beringer and Fercho (2020) and Williams et al. (2021) implemented the TOST procedure to provide statistical evidence that a new flight deck technology or operational concept provides an equivalent level of safety to current practices.

There are three possible outcomes for the TOST test: undetermined, statistically different, or statistically equivalent. Outcomes are *undetermined* if one confidence interval bound falls within the equivalence bounds while the other does not, indicating inconclusive evidence of equivalence. *Statistically different* outcomes occur when the traditional two-sided hypothesis test is significant, suggesting groups differ beyond a trivial range without evidence of equivalence. Lastly, outcomes are considered *statistically equivalent* when both bounds of the confidence interval fall within the equivalence range, with no significant difference in the traditional hypothesis test (Lakens, 2017; Seaman & Serlin, 1998).

Frequency of Missed Approaches Versus Landings

Prior to analyses, each scenario was evaluated to determine whether the flight crew continued to a landing or conducted a missed approach to evaluate the likelihood of a flight crew conducting a missed approach as a function of display type, EFVS mode, or ambient lighting. Across all 352 scenarios flown by participants in the study, no missed approaches occurred, demonstrating that a monocular HWD, with or without an EFVS, should support the ability of the PF to maintain a stabilized approach and continue to a landing during SA CAT I operations. Table 4 presents the number of landings that occurred in each experimental condition of the study.

Table 4
Frequency of Landings in Each Experimental Condition

		Display Type		
Ambient Lighting	**EFVS Mode**	**HUD**	**M. HWD**	**Total**
Day	EFVS Off	*22*	*22*	44
	EFVS On	*22*	*22*	44
Night	EFVS Off	*22*	*22*	44
	EFVS On	*22*	*22*	44
	Total	88	88	**352**

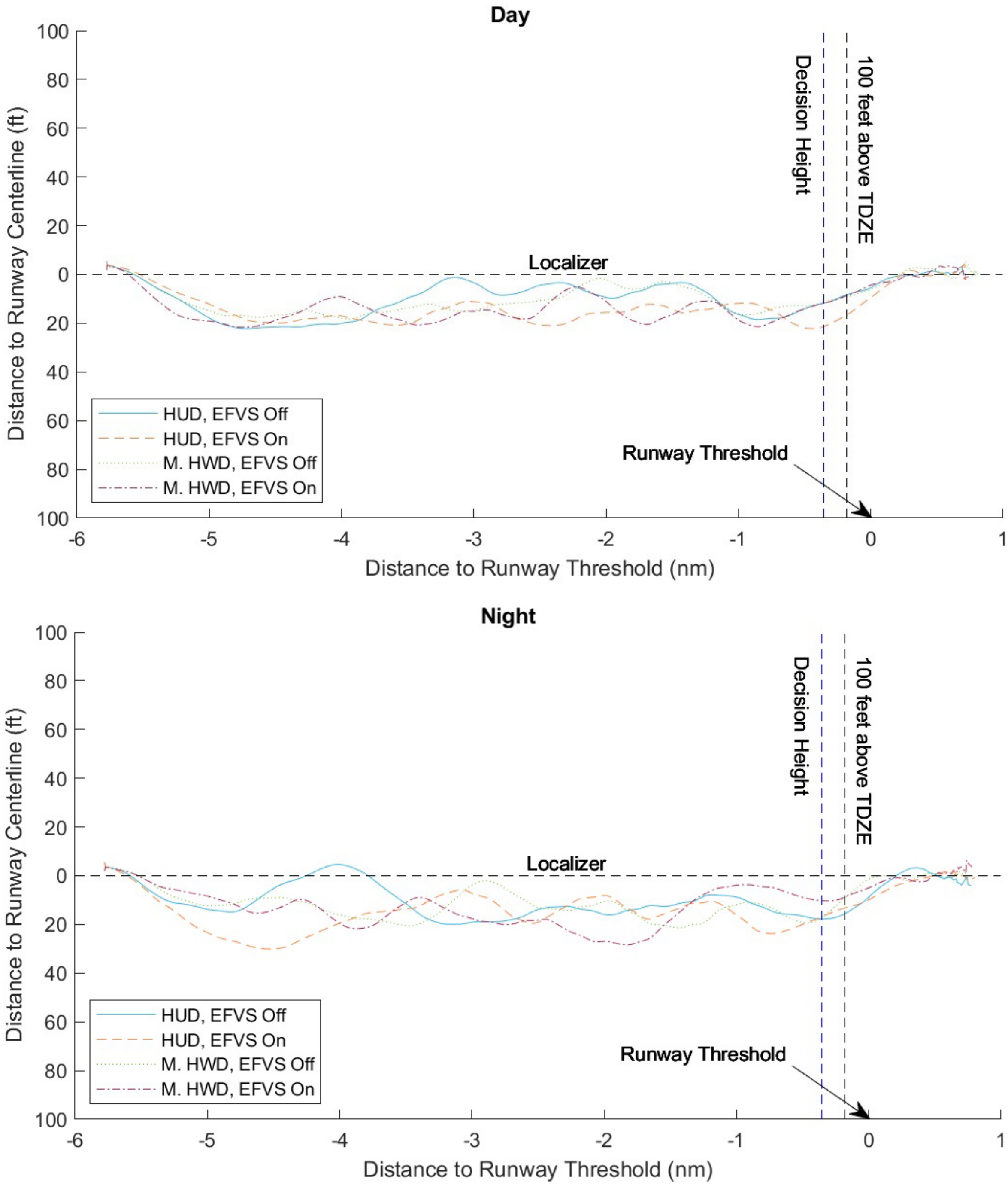
Localizer Deviation

Descriptive Overview of Lateral Flightpath Tracking

Figure 10 presents moving average plots of the lateral flightpath profile in each experimental condition from the start of each scenario until the aircraft had decelerated to 25 knots ground speed on the runway surface. Observations from these descriptive plots suggest that pilots' ability to track the lateral flightpath of an SA CAT I approach, landing, and rollout operation in day and night ambient lighting was largely consistent regardless of whether they flew with a monocular HWD or a HUD, or whether they flew with or without an EFVS. The greatest degree of deviation from the lateral flightpath appeared to occur throughout the instrument segment of the approach and when transitioning below the DH, likely due to the gusting winds that were present throughout the instrument segment in each scenario. As the winds began to dissipate and the aircraft descended below the DH, pilots were able to line up with the runway centerline prior to landing and execute the touchdown on a suitable point on the runway. This ability appeared to be consistent across all combinations of display type, EFVS mode, and ambient lighting condition.

Figure 10

Moving Average Plots of Lateral Flightpath Profile in Each Experimental Condition



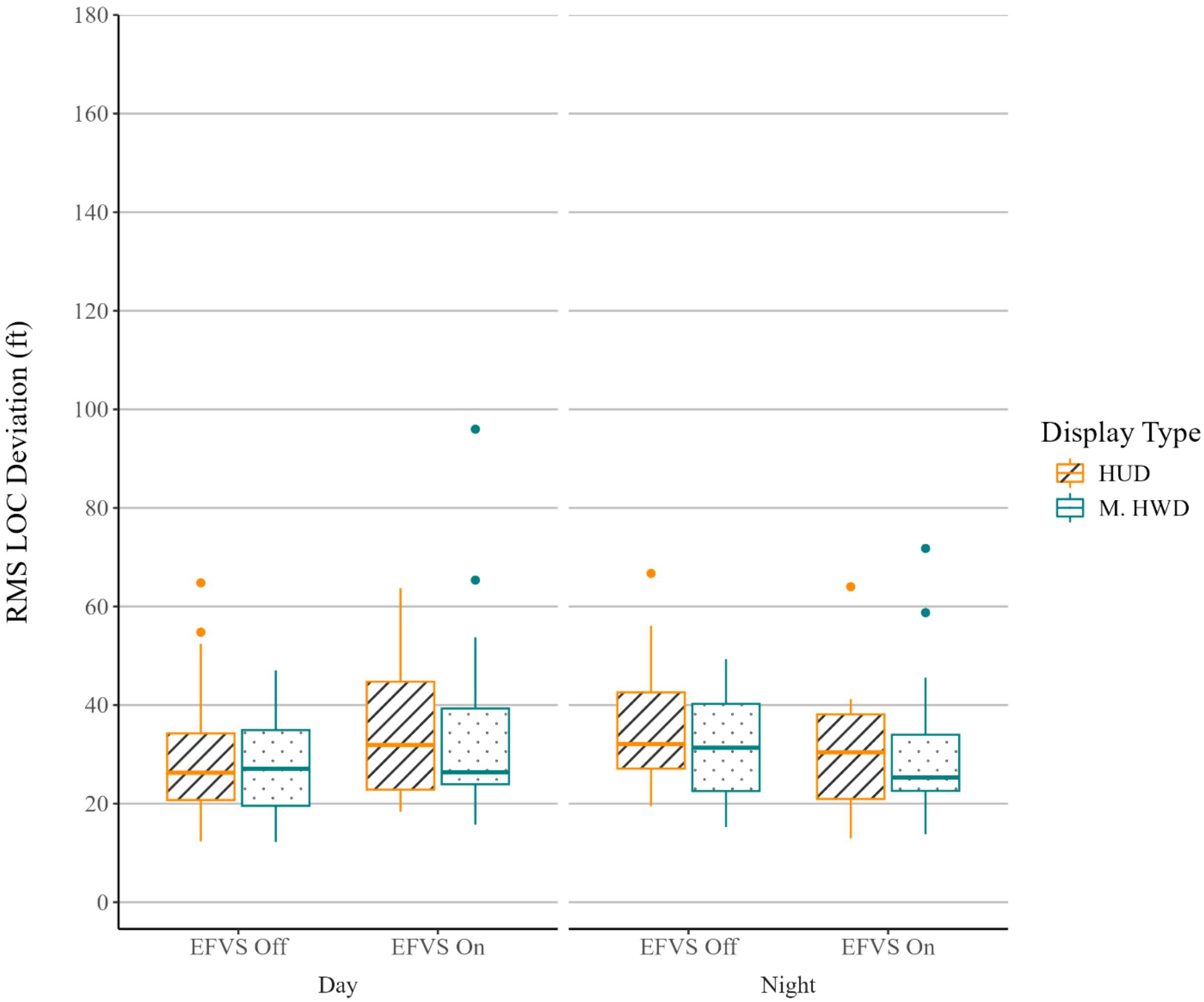
The lateral position of the reference lines for the DH and 100 feet above the TDZE is based on where these points occur along the 3° glideslope defined in the published approach procedure.

Root Mean Square Localizer Deviation During the Instrument Segment

Lateral flightpath deviation throughout the instrument segment of the approach was measured relative to the ILS localizer by calculating the RMS of the localizer deviation from the start of the scenario to 300 feet above the TDZE. Figure 11 presents a Box-and-Whisker plot of RMS localizer deviation during the instrument segment for each of the experimental conditions. As shown in the figure, several outliers were present in the RMS localizer deviation data across the experimental conditions, indicating that the gusting winds with variable direction present in the instrument segment of each scenario made it challenging to maintain the localizer. The consistent distribution of these outliers across conditions indicates that this pattern occurred regardless of display type, EFVS mode, and ambient lighting.

Figure 11

RMS Localizer Deviation During the Instrument Segment by Display Type, EFVS Mode, and Ambient Lighting



In this and all subsequent Box-and-Whisker plots, the median is indicated by a horizontal line within each box. The box represents the Interquartile Range (IQR), covering the middle 50% of the data. The whiskers extend to 1.5 × IQR, illustrating the variability beyond the upper and lower quantiles. Data points that fall outside the whiskers are considered outliers.

The Shapiro-Wilk test indicated that the RMS localizer deviation data violated the normality assumption. After applying a Box-Cox transformation (λ = -0.191), the normality assumption was met. The 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA results with transformed data revealed no significant main effects involving display type, EFVS mode, or ambient lighting on RMS localizer deviation during the instrument segment ($p > .05$ for all models). However, there was a significant interaction effect between EFVS mode and ambient lighting, F (1, 21) = 11.66, $p = .003$, $\eta_p^2 = .36$.

To investigate the directionality of this interaction, post-hoc pairwise comparisons were conducted for each factor combination (see Table 5). Post-hoc pairwise comparisons revealed that RMS localizer deviation in the instrument segment was lower when the EFVS was off compared to when it was on with daytime ambient lighting ($p = .018$, $d = -0.08$). However, with night ambient lighting, RMS localizer deviation was higher when the EFVS was off compared to when it was on ($p = .024$, $d = 0.07$). Pairwise comparison results also revealed a smaller RMS localizer deviation with day ambient lighting than with night ambient lighting (p = .025, d = -0.09), but only when the EFVS was off (see Table 6 and Table 7). However, as demonstrated by the differences in mean values presented in Table 5 - Table 7, these effects were small in magnitude and likely do not translate to operationally significant differences in pilot performance. The follow-up TOST test indicated that both the upper and lower 90% confidence interval bounds (-0.091, 0.034) fell outside the equivalence bounds of (-0.003, 0.003), t (21) = -0.69, $p = .751$ (see Appendix B, Figure B1). Therefore, pilots' absolute localizer deviation in the instrument segment was not found to be significantly equivalent between the HUD and Monocular HWD conditions.

Table 5

M_{adj} of RMS Localizer Deviation During the Instrument Segment Between Ambient Lighting and EFVS Mode

Ambient Lighting	EFVS	M_{BT}	M_T	SE	95% Confidence Interval	
					Lower Bound	Upper Bound
Day	Off	26.21	2.43	0.04	2.36	2.51
	On	30.50	2.51	0.03	2.45	2.58
Night	Off	31.09	2.52	0.03	2.47	2.58
	On	27.21	2.45	0.03	2.39	2.52

In this and all subsequent tables, M_T refers to the adjusted mean of the transformed data, and M_{BT} refers to the back-transformed M_T.

Table 6

Pairwise Comparisons of RMS Localizer Deviation During the Instrument Segment within Ambient Lighting by EFVS Mode

Ambient Lighting	**EFVS Comparison**	**Deviation (ΔM_T)**	***SE***	***p***	**η_p^2**	**95% Confidence Interval**
Day	Off - On	-0.08	0.03	.018	.24	[-0.15, -0.02]
Night	Off - On	0.07	0.03	.024	.22	[0.01, 0.13]

In this and all subsequent tables, ΔM_T refers to the difference between the adjusted means of the transformed data at two levels.

Table 7

Pairwise Comparisons of RMS Localizer Deviation During the Instrument Segment within EFVS Mode by Ambient Lighting

EFVS	**Ambient Lighting Comparison**	**Deviation (ΔM_T)**	***SE***	***p***	**η_p^2**	**95% Confidence Interval**
Off	Day - Night	-0.09	0.04	.025	.22	[-0.17, -0.01]
On	Day - Night	0.06	0.03	.064	.16	[-0.00,[8] 0.13]

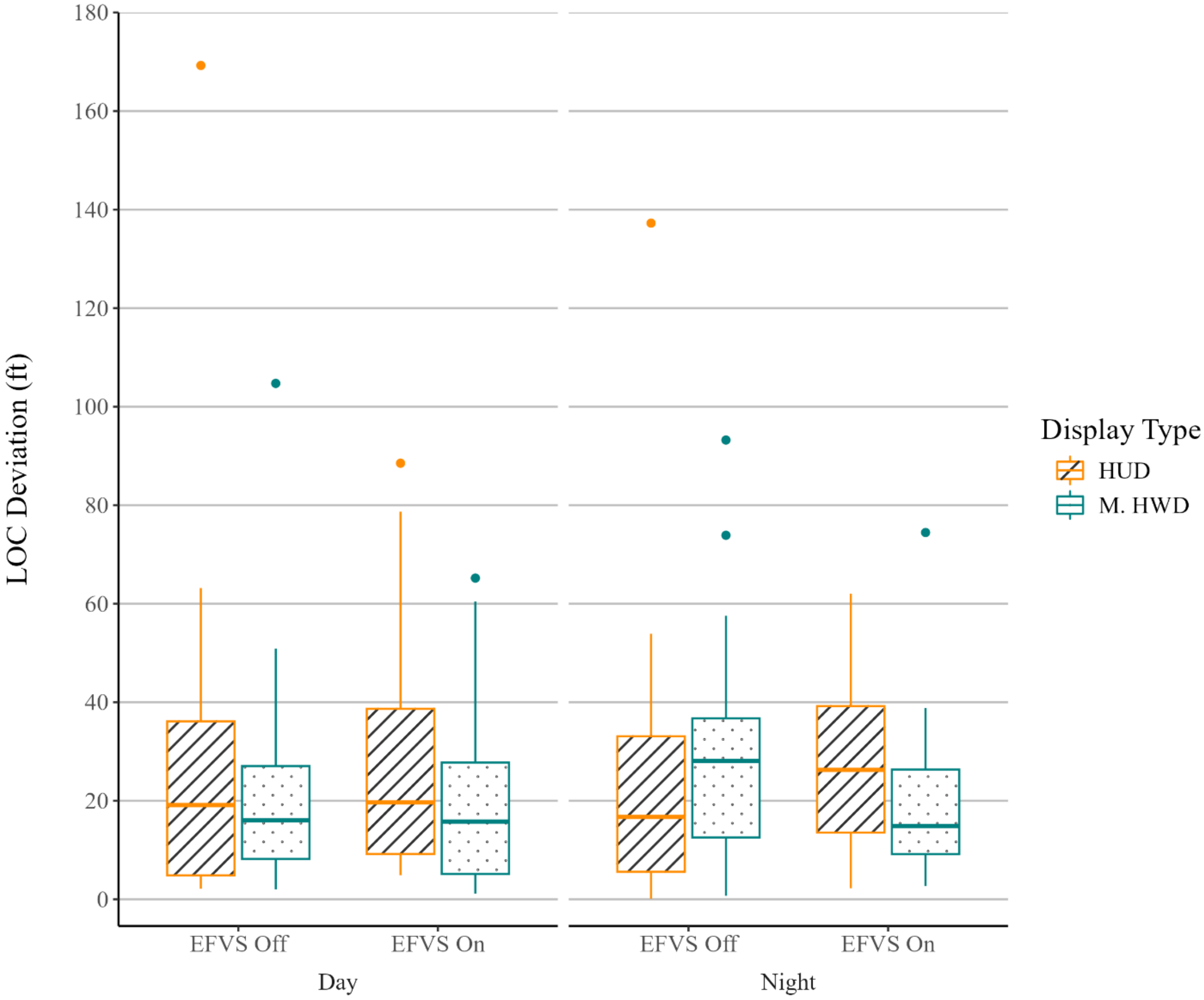
Absolute Localizer Deviation at the Decision Height

The lateral flightpath deviation at the DH of the approach was assessed by measuring the absolute deviation from the ILS localizer guidance, in feet, when the Radar Altimeter (RA) altitude of the aircraft reached 157 feet. Figure 12 presents a Box-and-Whisker plot of localizer deviation at the DH in each experimental condition. As shown in the figure, several outliers were present in the localizer deviation data across the experimental conditions, indicating that the gusting winds with variable direction present in the instrument segment of each scenario made it challenging to maintain the localizer when crossing the DH. The consistent distribution of these outliers across conditions indicates that this pattern occurred regardless of display type, EFVS mode, and ambient lighting.

[8] The original value is -0.004, which is rounded to two decimal places in this table.

Figure 12

Absolute Localizer Deviation at the Decision Height by Display Type, EFVS Mode, and Ambient Lighting



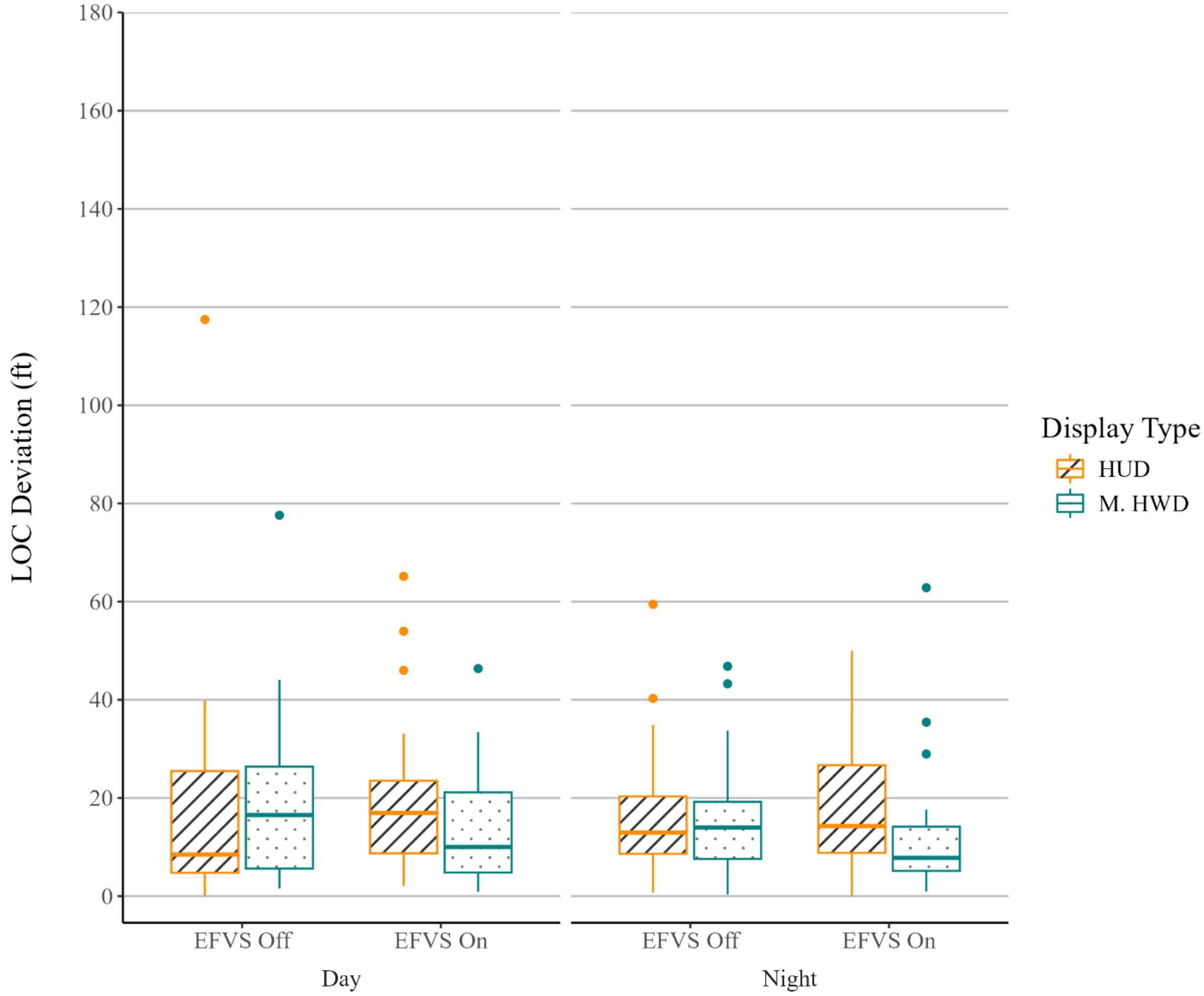
The ability of the PF to track the runway centerline at the DH was evaluated as a function of display type, EFVS mode, and ambient lighting using a 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA. Shapiro-Wilk test results revealed that the model residuals violated the normality assumption, so a Box-Cox transformation with $\lambda = 0.259$ was applied. After the transformation, Shapiro-Wilk test results showed that the model residuals met the normality assumption. The analysis with the transformed data did not reveal any significant main effects of display type, EFVS mode, or ambient lighting on localizer deviation at the DH, nor any significant interaction effects ($p > .05$ for all models). The follow-up TOST test indicated that both the upper and lower 90% confidence interval bounds (-0.621, 0.170) fell outside the equivalence raw bounds of (-0.023, 0.023), $t(21) = -0.88$, $p = .805$ (see Appendix B, Figure B2). Therefore, pilots' absolute localizer deviation at the DH was not found to be significantly equivalent between the HUD and Monocular HWD conditions.

Absolute Localizer Deviation at 100 feet Above the Touchdown Zone Elevation

The lateral flightpath deviation when the aircraft reached 100 feet above the TDZE was evaluated by measuring the absolute deviation from the ILS localizer guidance, in feet. Figure 13 presents a Box-and-Whisker plot of absolute localizer deviation at 100 feet above the TDZE in each experimental condition. As shown in the figure, several outliers were present in the localizer deviation data across the experimental conditions, indicating that the winds present in each scenario made it challenging to maintain the localizer. The somewhat consistent distribution of these outliers across conditions indicates that this pattern occurred regardless of display type, EFVS mode, and ambient lighting.

Figure 13

Absolute Localizer Deviation at 100 feet Above the TDZE by Display Type, EFVS Mode, and Ambient Lighting



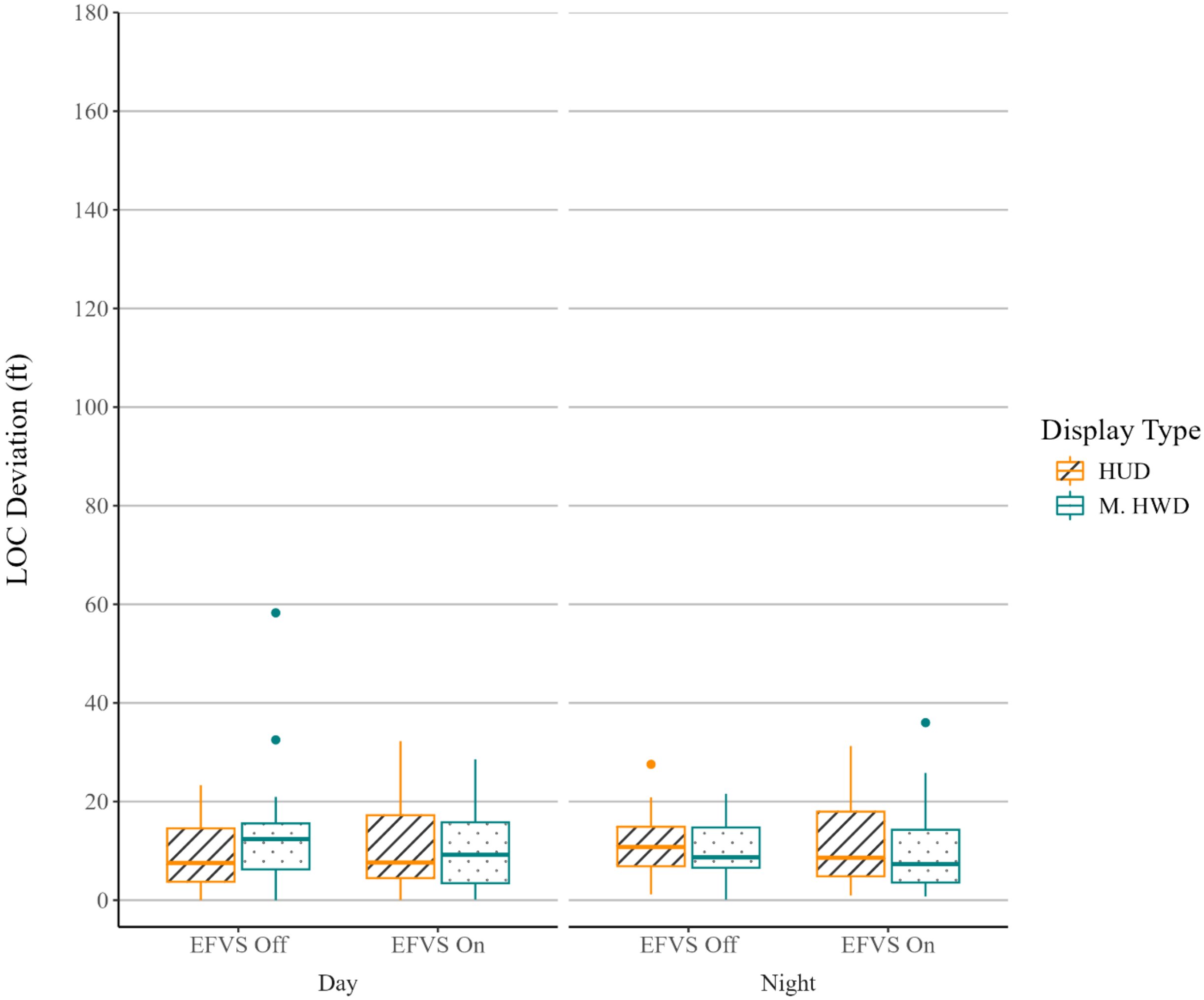
The effects of display type, EFVS mode, and ambient lighting conditions on localizer deviation at 100 feet above the TDZE were analyzed using a 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA. The model residuals violated the normality assumption based on the results of the Shapiro-Wilk test. Therefore, a Box-Cox transformation with $\lambda = 0.307$ was applied to the data, which resolved the non-normality. The analysis of transformed data showed no significant main effect of display type, EFVS mode, or ambient lighting on pilots' absolute airspeed deviation at 100 feet above the TDZE, and no significant interaction effects between the factors ($p > .05$ for all models). The follow-up TOST test indicated that both the upper and lower 90% confidence interval bounds (-0.734,0.283) fell outside the equivalence bounds of (-0.112, 0.112), $t(21) = -0.38$, $p = .647$ (see Appendix B, Figure B3). Therefore, pilots' absolute localizer deviation at 100 ft. above the touchdown zone was not found to be significantly equivalent between the HUD and Monocular HWD conditions.

Absolute Localizer Deviation at Threshold Crossing

The lateral flightpath deviation when the aircraft center of gravity crossed the runway threshold was assessed by measuring the absolute deviation from the ILS localizer guidance, in feet. Figure 14 presents a Box-and-Whisker plot of absolute localizer deviation at threshold crossing in each experimental condition. As shown in the figure, several outliers are present in the data, however there are fewer outliers and an overall smaller amount of deviation in the data at threshold crossing compared to the localizer deviation data at earlier points in the approach (e.g., at the DH), indicating that participants were able to correct flightpath deviations in the visual segment of the approach, and they were successful in doing so across all conditions of display type, EFVS mode, and ambient lighting.

Figure 14

Absolute Localizer Deviation at Threshold Crossing by Display Type, EFVS Mode, and Ambient Lighting



The Shapiro-Wilk test showed that the model residuals for absolute localizer deviation at threshold crossing did not meet the normality assumption. A Box-Cox transformation with $\lambda = 0.450$ was applied. Following the transformation, the model residuals met the normality assumption. The 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA with the transformed data did not show a significant main effect of display type, EFVS mode, or ambient lighting on pilots' absolute localizer deviation at threshold crossing, and no significant interaction effects between the factors ($p > .05$ for all models).

The follow-up TOST test indicated that both the upper and lower 90% confidence interval bounds (-0.591, 0.478) fell outside the equivalence bounds of (-0.273, 0.273), $t(21) = 0.70$, $p = .247$ (see Appendix B, Figure B4). Therefore, pilots' absolute localizer deviation at threshold crossing was not found to be significantly equivalent between the HUD and Monocular HWD conditions.

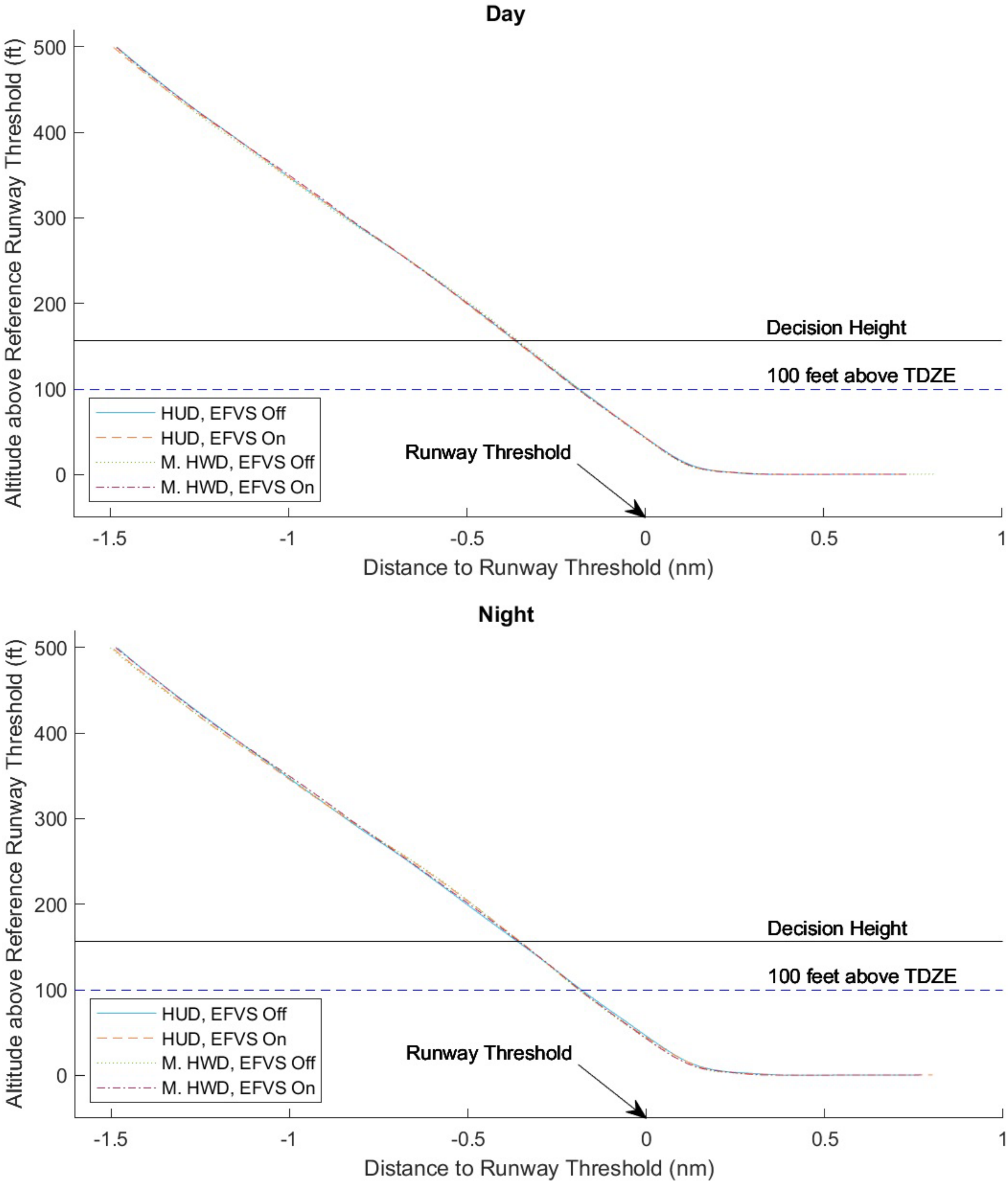
Glideslope Deviation

Descriptive Overview of Vertical Flightpath Tracking

Figure 15 presents moving average plots of the vertical flightpath profile in each experimental condition from 500 feet above the runway threshold until the aircraft had decelerated to 25 knots ground speed on the runway surface. Observations from these descriptive plots suggest that pilots' ability to follow a defined vertical flightpath during the final stages of the instrument segment and during the visual segment of a manual SA CAT I approach with and without an EFVS was largely consistent regardless of whether they flew with a monocular HWD or a HUD.

Figure 15

Moving Average Plots of Vertical Flightpath Profile in Each Experimental Condition



The vertical position of the reference lines for the DH and 100 feet above the TDZE is based on where these points occur along the 3° glideslope defined in the published approach procedure.

Root Mean Square Glideslope Deviation During the Instrument Segment

Vertical flightpath deviation throughout the instrument segment of the approach was measured relative to the ILS glideslope by calculating the RMS of the glideslope deviation from the start of the scenario to 300 feet above the TDZE. Figure 16 presents a Box-and-Whisker plot of RMS glideslope deviation during the instrument segment in each experimental condition. The Shapiro-Wilk test showed that the residuals of the RMS glideslope deviation model met the assumption of normality; therefore, no transformation was applied. The 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA revealed a significant effect of display type, $F(1, 21) = 8.05$, $p = .010$, $\eta_p^2 = .28$, with a greater RMS glideslope deviation observed with monocular HWD compared to HUD, as shown by post-hoc pairwise comparisons results ($p = .01$, $d = 1.39$). Trends among the mean deviation values in each condition indicate that while this difference was statistically significant, the practical significance of the effect is small, as RMS glideslope deviation was only 1.39 feet greater with the monocular HWD than it was with the HUD (see Table 8). There were no significant effects of EFVS mode or ambient lighting on pilots' RMS glideslope deviation, with no significant interaction effects observed ($p > .05$ for all models).

Figure 16

RMS Glideslope Deviation during the Instrument Segment by Display Type, EFVS Mode, and Ambient Lighting

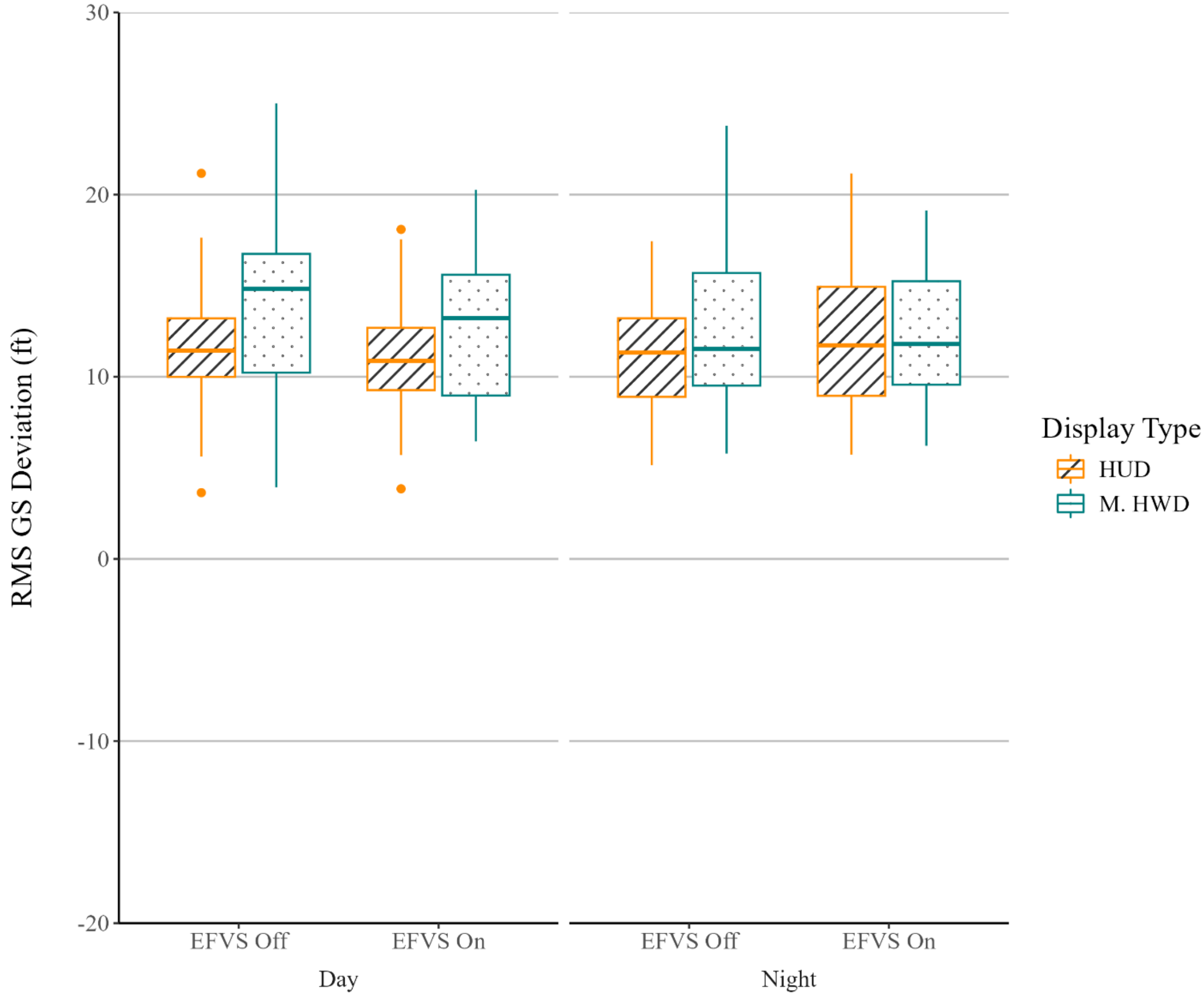


Table 8

M_{adj} of RMS Glideslope Deviation in Instrument Segment Between Display Type Conditions

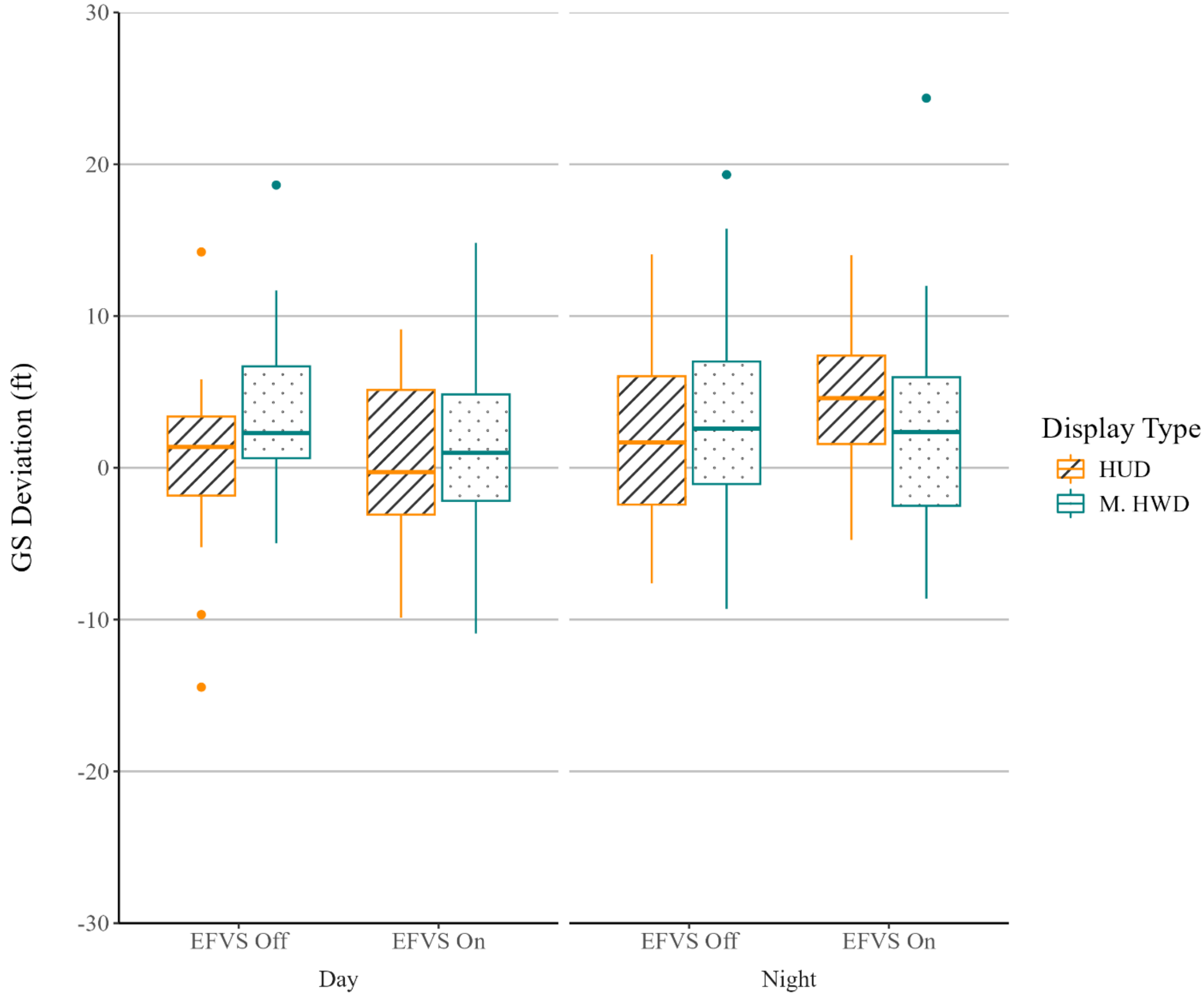
Display Type	*M*	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
HUD	11.55	0.56	10.39	12.72
M. HWD	12.94	0.74	11.40	14.48

Glideslope Deviation at Decision Height

Vertical flightpath deviation at the DH was assessed by measuring the deviation from the ILS glideslope, in feet, when the RA altitude of the aircraft reached 157 feet. Figure 17 presents a Box-and-Whisker plot of glideslope deviation at the DH in each experimental condition.

Figure 17

Glideslope Deviation at the DH by Display Type, EFVS Mode, and Ambient Lighting



In the above figure, positive values represent deviation above the glideslope and negative values represent deviation below the glideslope.

A 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA was conducted to evaluate the effects of display type, EFVS mode, and ambient lighting on pilots' glideslope deviation at the DH. Shapiro-Wilk test results indicated that the model residuals followed a normal distribution. The RM ANOVA revealed a significant main effect of ambient lighting on glideslope deviation at the DH, F (1, 21) = 5.34, p = .031, η_p^2 = .20. Post-hoc t-tests showed that participants crossed the DH with greater glideslope deviation during the 'night' scenarios compared to the 'day' scenarios (p = .031, d = 1.36). While this difference was statistically significant, the magnitude of the difference was small from an operational standpoint; the glideslope deviation at the DH was 1.37 feet greater with night ambient lighting than it was with day ambient lighting (see Table 9). The analysis did not show a significant effect of display type or EFVS mode on pilots' glideslope deviation at the DH, with no significant interaction effects observed (p > .05 for all models). The follow-up TOST test indicated that the lower 90% confidence interval bounds (-0.091, 2.023) fell outside the equivalence bounds (-0.441, 0.441), t (21) = 0.85, p = .799 (see Appendix B, Figure B5). Therefore, pilots' glideslope deviation at the DH was not found to be significantly equivalent between the HUD and Monocular HWD conditions.

Table 9

M_{adj} of Glideslope Deviation at the DH Between Ambient Lighting Conditions

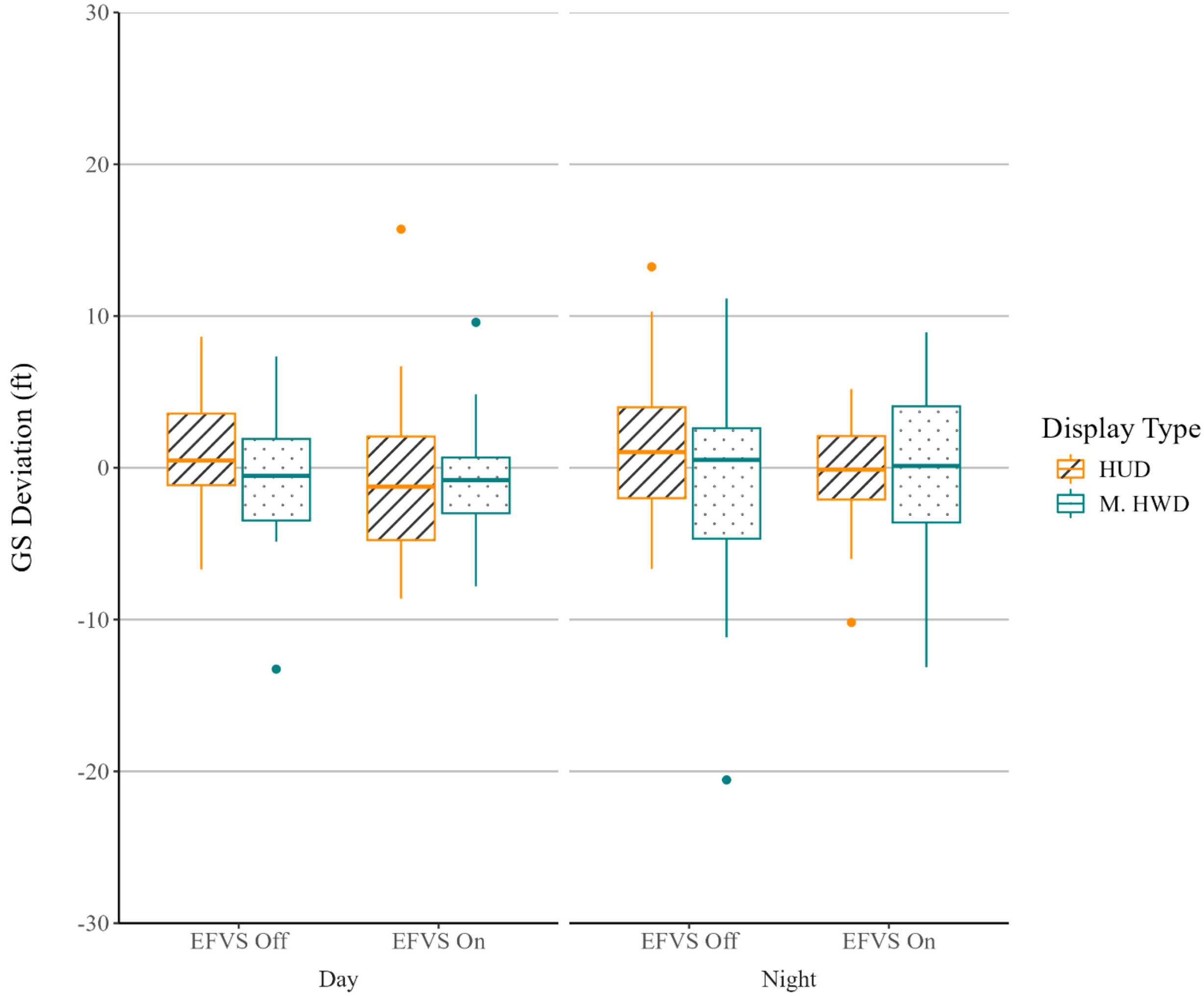
Ambient Lighting	*M*	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
Day	1.48	0.89	-0.36	3.33
Night	2.85	0.91	0.96	4.73

Glideslope Deviation at 100 feet Above the Touchdown Zone Elevation

The vertical flightpath deviation when the aircraft was 100 feet above the TDZE was evaluated by measuring the deviation from the ILS glideslope, in feet. Figure 18 presents a Box-and-Whisker plot of glideslope deviation at 100 feet above the TDZE in each experimental condition.

Figure 18

Glideslope Deviation at 100 feet Above the TDZE by Display Type, EFVS Mode, and Ambient Lighting



In the above figure, positive values represent deviation above the glideslope and negative values represent deviation below the glideslope.

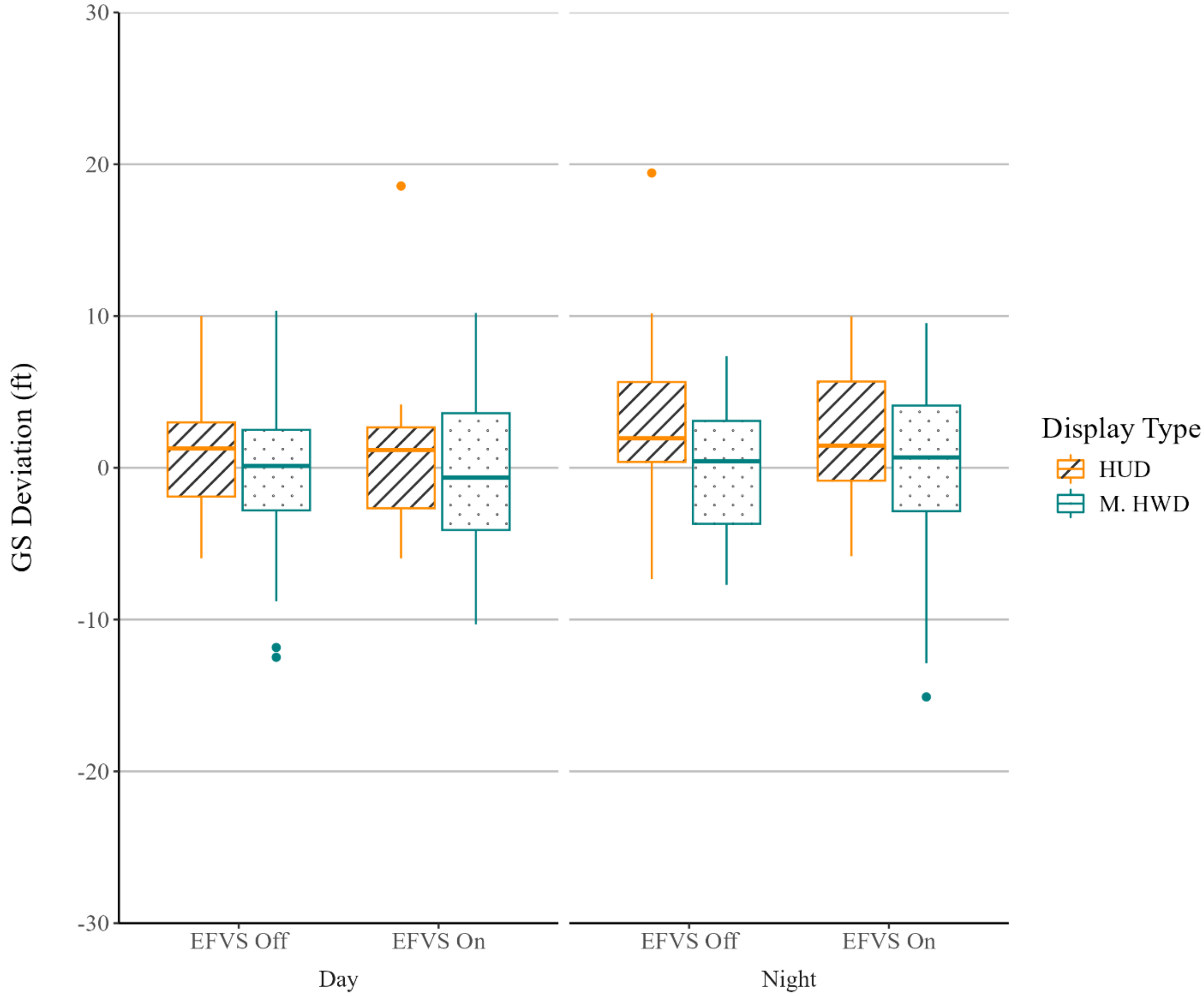
The Shapiro-Wilk test confirmed that the model's residuals met the assumption of normality. The effect of display type, EFVS mode, and ambient lighting on pilots' glideslope deviation at 100 feet above the TDZE was examined through a 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA, which revealed that there was not a significant main effect of display type, EFVS mode, or ambient lighting, nor any interaction effect of the factors on pilots' glideslope deviation at 100 feet above the TDZE ($p > .05$ for all models). The follow-up TOST test indicated that both the upper and lower 90% confidence interval bounds (-1.969, 0.097) fell outside the equivalence bounds of (-0.913, 0.913), $t\,(21) = -0.04$, $p = .515$ (see Appendix B, Figure B6). Therefore, pilots' glideslope deviation at threshold crossing was not found to be significantly equivalent between the HUD and Monocular HWD conditions.

Glideslope Deviation at Threshold Crossing

The vertical flightpath deviation when the aircraft center of gravity crossed the runway threshold was evaluated by calculating the deviation from the ILS glideslope, in feet. Figure 19 presents a Box-and-Whisker plot for glideslope deviation at threshold crossing in each experimental condition.

Figure 19

Glideslope Deviation at Threshold Crossing by Display Type, EFVS Mode, and Ambient Lighting



In the above figure, positive values represent deviation above the glideslope and negative values represent deviation below the glideslope.

The Shapiro-Wilk test results suggested that the model residuals did not meet the normality assumption. Therefore, a Box-Cox transformation with $\lambda = 0.958$ was applied to the data. However, even after the Box-Cox transformation, the Shapiro-Wilk test was significant, indicating a deviation from normality ($p = .024$). To further investigate the source of non-normality, a Shapiro-Wilk test was conducted on the model residuals after removing an outlier identified under the HUD, EFVS on, and day ambient lighting conditions. After removing this outlier, the Shapiro-Wilk test results showed an improvement in normality; however, the test result was bordering on statistically significant (p = .053). Because only a single data point appeared to cause the non-normality issue, an RM ANOVA was conducted on the Box-Cox-transformed data, including the outlier.

The 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA revealed a significant effect of display type on pilots' glideslope deviation at threshold crossing; F (1,21) = 6.48, p = .019, η_p^2 = .24. Post-hoc t-test results revealed that pilots exhibited a larger glideslope deviation at threshold crossing when flying with the HUD compared to when flying with the monocular HWD (p = 0.019, d = 1.68). However, this difference was small from an operational standpoint, as glideslope deviation at threshold crossing was only 1.91 feet greater with the HUD than with the monocular HWD (see Table 10). There was not a significant effect of EFVS mode or ambient lighting on pilots' glideslope deviation at threshold crossing, nor were there any significant interaction effects (p > .05 for all models).

Table 10

M_{adj} of Glideslope Deviation at Threshold Crossing Between Display Type Conditions

Display Type	M_{BT}	M_T	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	17.68	15.31	0.60	14.06	16.55
M. HWD	15.77	13.62	0.67	12.23	15.01

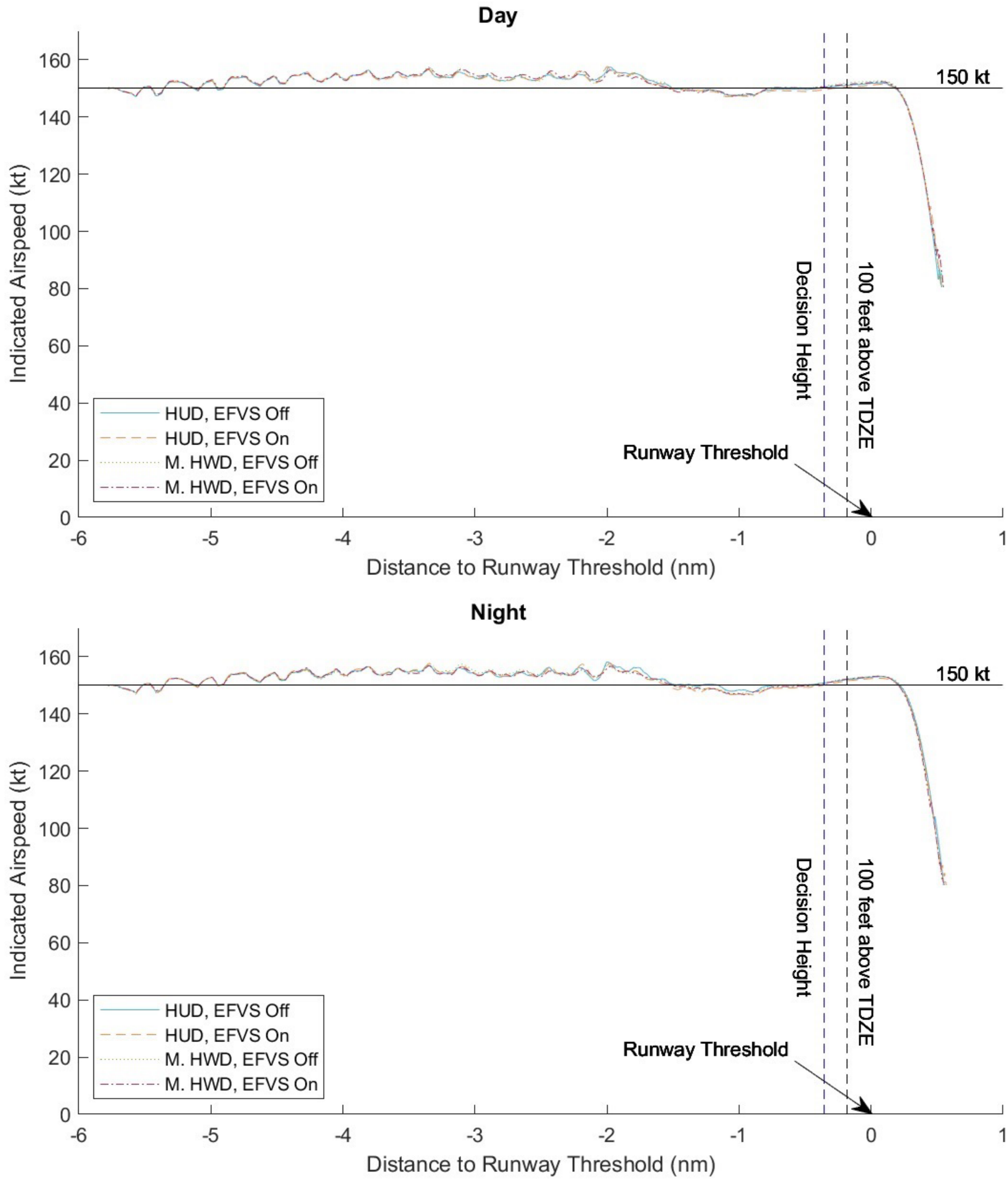
Airspeed Deviation

Descriptive Overview of Energy Management Performance

Figure 20 presents moving average plots of the IAS in each experimental condition from the start of each scenario until the aircraft had decelerated to 80 knots ground speed on the runway surface. Observations from this descriptive overview of energy management performance suggest that pilots' ability to manage the aircraft's energy state during a manual SA CAT I operation is largely consistent regardless of whether they flew with a monocular HWD or a HUD, or whether they flew with or without an EFVS. As indicated in the plots, deviations from the IAS target of 150 knots were apparent throughout most of the instrument segment of the approach. This is likely due to the gusting winds aloft that were present from the start of each scenario until dissipating at 375 feet AGL, which corresponds with approximately 1.34 nautical miles from the runway threshold. As shown in the plots, pilots were consistent in their ability to bring the IAS of the airspeed back to the target IAS at roughly this point of the approach, regardless of the display type, EFVS mode, and ambient lighting. Energy management behavior beyond this point of the approach was consistent across the experimental factors as well, indicating that pilots were consistent in their ability to manage airspeed during the visual segment, retard the throttles at touchdown, and decelerate the aircraft during the high-speed portion of the rollout segment, regardless of the display type, EFVS mode, or ambient lighting.

Figure 20

Moving Average Plots of Airspeed in Each Experimental Condition



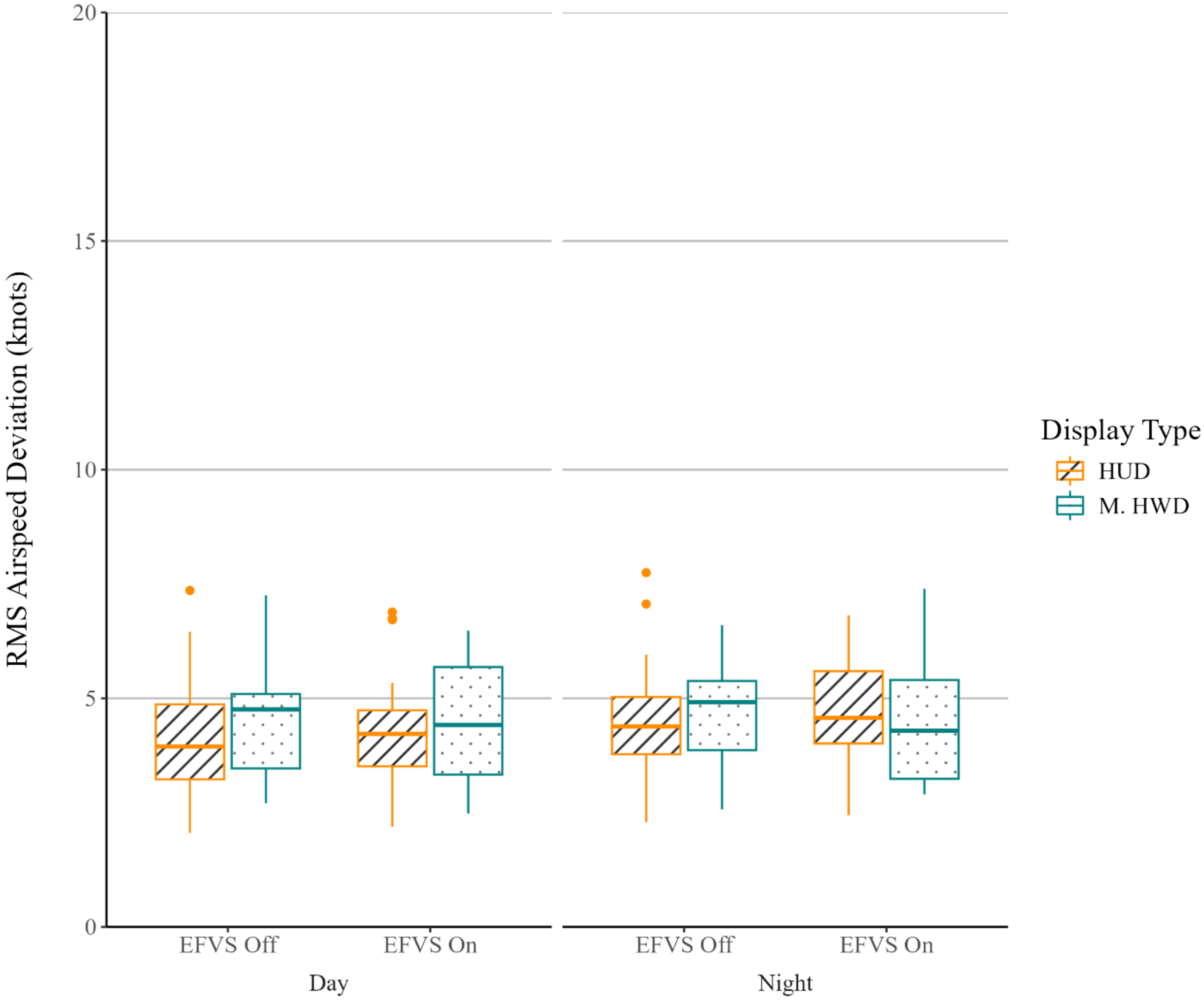
The lateral position of the reference lines for the DH and 100 feet above the TDZE is based on where these points occur along the 3° glideslope defined in the published approach procedure.

Root Mean Square Airspeed Deviation in the Instrument Segment

RMS airspeed deviation was assessed by comparing the actual IAS to the target IAS set in the MCP (i.e., 150 knots) from the start of the approach until the aircraft had descended to 300 feet AGL. Figure 21 presents a Box-and-Whisker plot of RMS airspeed deviation in the instrument segment for each experimental condition.

Figure 21

RMS Airspeed Deviation in Instrument Segment by Display Type, EFVS Mode, and Ambient Lighting



The pilot's ability to maintain the target airspeed, measured in knots IAS, during the instrument segment was evaluated across experimental conditions using a 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA. Shapiro-Wilk test results indicated that the model residuals were normally distributed ($p > .05$). The RM ANOVA revealed that there were no significant main effects of display type, EFVS mode, or ambient lighting on the RMS airspeed deviation, nor were there any significant interaction effects between the experimental factors ($p > .05$ for all models).

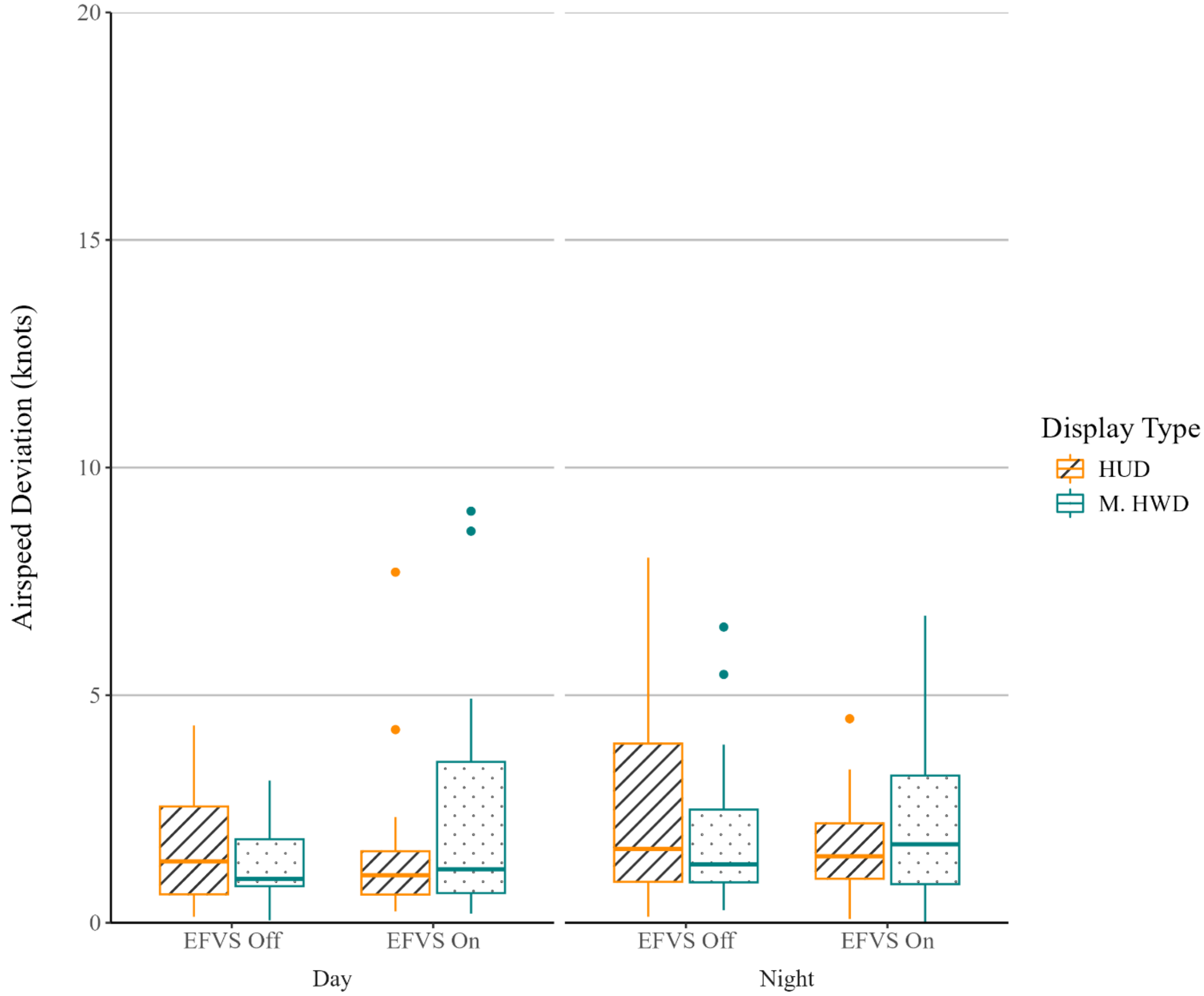
The follow-up TOST test indicated that both the upper and lower 90% confidence interval bounds (-0.031, 0.271) fell outside the equivalence bounds of (-0.131, 0.131), t (21) = -0.13, p = .448 (see Appendix B, Figure B7). Therefore, pilots' absolute deviation from airspeed at the DH was not found to be significantly equivalent between the HUD and Monocular HWD conditions.

Absolute Airspeed Deviation at the Decision Height

Deviation from the target airspeed of the approach when the aircraft crossed the DH was measured by calculating the difference, in knots, between the observed IAS and the target IAS programmed in the MCP (i.e., 150 knots), when the RA altitude of the aircraft reached 157 feet RA altitude. Figure 22 presents a Box-and-Whisker plot of absolute airspeed deviation at the DH in each experimental condition.

Figure 22

Absolute Airspeed Deviation at the DH by Display Type, EFVS Mode, and Ambient Lighting



Pilots' ability to fly at the target indicated airspeed of the approach when crossing the DH was evaluated using a 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA. The

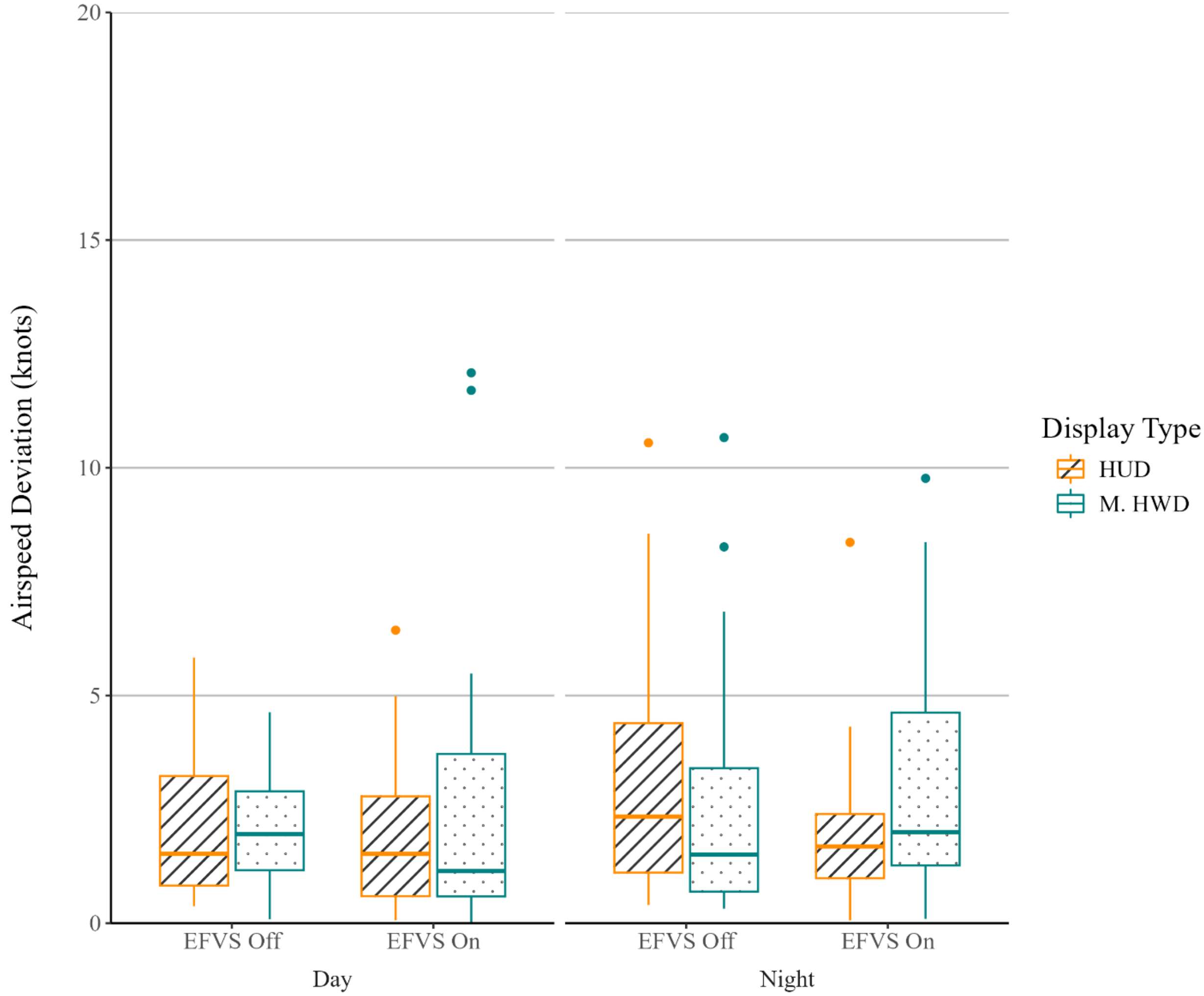
Shapiro-Wilk test revealed that the residuals violated the normality assumption. After applying a Box-Cox transformation ($\lambda = 0.213$), the normality assumption was met. The RM ANOVA results with transformed data showed that there was not a significant main effect of display type, EFVS mode, or ambient lighting on pilots' absolute airspeed deviation at the DH, nor were there any significant interaction effects between the factors ($p > .05$ for all models). The follow-up TOST test indicated that both the upper and lower 90% confidence interval bounds (-0.248, 0.369) fell outside the equivalence bounds of (-0.158, 0.158), $t\,(21) = -0.55$, $p = .295$ (see Appendix B, Figure B8). Therefore, pilots' absolute localizer deviation at threshold crossing was not found to be significantly equivalent between the HUD and Monocular HWD conditions.

Absolute Airspeed Deviation at 100 feet Above the Touchdown Zone Elevation

The absolute airspeed deviation at 100 feet above the TDZE was assessed by calculating the absolute difference, in knots, between the observed IAS and the target IAS set in the MCP (i.e., 150 knots). Figure 23 presents the Box-and-Whisker plot of absolute airspeed deviation at 100 feet above the TDZE.

Figure 23

Absolute Airspeed Deviation at 100 feet Above the TDZE by Display Type, EFVS Mode, and Ambient Lighting



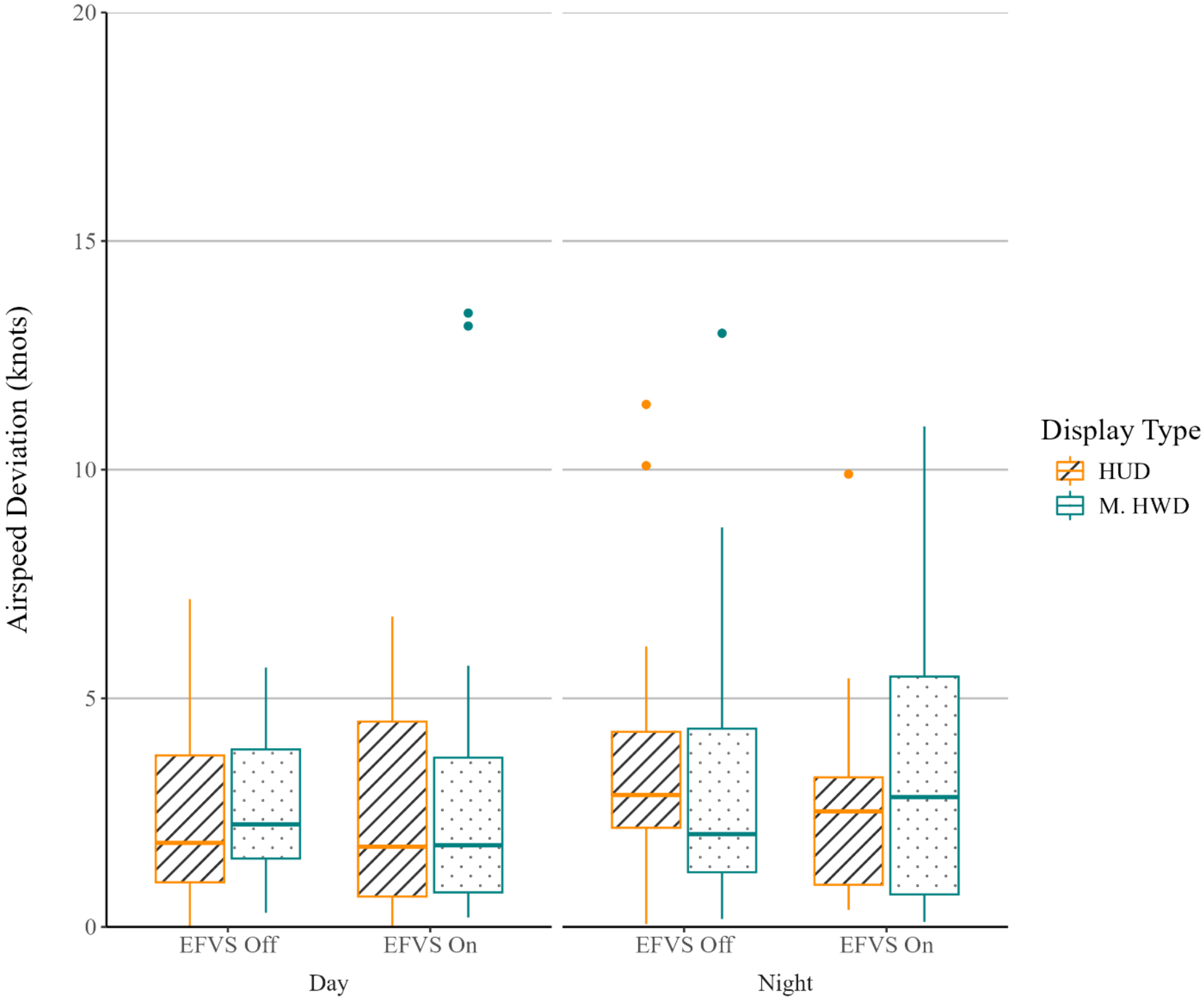
The Shapiro-Wilk test indicated that the model residuals for absolute airspeed deviation at 100 feet above the TDZE violated the normality assumption. Consequently, a Box-Cox transformation with $\lambda = 0.250$ was applied. After the transformation, the model residuals met the normality assumption. The 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA with the transformed data showed that there was not a significant main effect of display type, EFVS mode, or ambient lighting on pilots' absolute airspeed deviation at 100 feet above the TDZE, nor were there any significant interaction effects between the factors ($p > .05$ for all models). The follow-up TOST test indicated that both the upper and lower 90% confidence interval bounds (-0.198, 0.358) fell outside the equivalence bounds of (-0.187, 0.187), $t(21) = -0.66$, $p = .258$ (see Appendix B, Figure B9). Therefore, pilots' airspeed deviation at 100 feet above the TDZE was not found to be significantly equivalent between the HUD and Monocular HWD conditions.

Absolute Airspeed Deviation at Threshold Crossing

The airspeed deviation when the aircraft center of gravity crossed the runway threshold was assessed by calculating the absolute difference, in knots, between the actual IAS and the target IAS set in the MCP (i.e., 150 knots). Figure 24 presents a Box-and-Whisker plot for absolute airspeed deviation at threshold crossing in each experimental condition.

Figure 24

Absolute Airspeed Deviation at Threshold Crossing by Display Type, EFVS Mode, and Ambient Lighting



The Shapiro-Wilk test indicated that the model residuals violated the normality assumption, requiring a transformation. Applying a Box-Cox transformation with $\lambda = 0.323$ resulted in a normal distribution of the residuals. The 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA with the transformed data revealed that there was not a significant main effects of display type, EFVS mode, or ambient lighting condition, nor were there any significant interaction effects between the experimental factors ($p > .05$ for all models).

The follow-up TOST test indicated that both the upper and lower 90% confidence interval bounds (-0.131, 0.472) fell outside the equivalence bounds of (-0.149, 0.149), $t\ (21) = 0.12$, p =

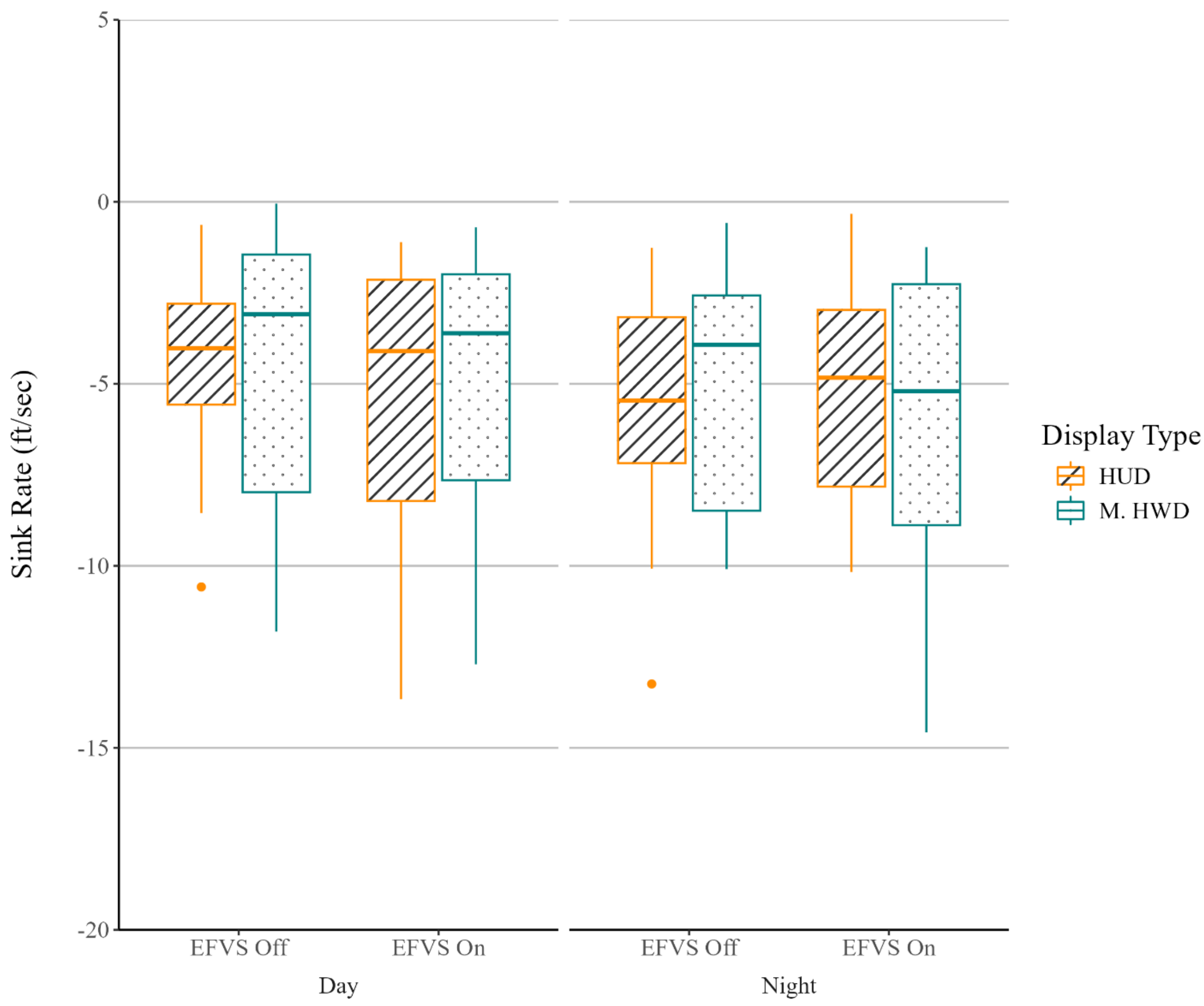
.548 (see Appendix B, Figure B10). Therefore, pilots' airspeed deviation at threshold crossing was not found to be significantly equivalent between the HUD and Monocular HWD conditions.

Sink Rate at Touchdown

The aircraft's average vertical velocity, measured in feet per second, was recorded during the two seconds leading up to touchdown of the main landing gear to evaluate pilots' ability to manage the sink rate of the aircraft leading up to landing as a function of display type, EFVS mode, and ambient lighting. Figure 25 presents a Box-and-Whisker plot for sink rate at touchdown in each experimental condition. As shown in the figure, some outliers were present in the data, with several landings in most experimental conditions exceeding a sink rate of 10 feet/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 25

Sink Rate at Touchdown by Display Type, EFVS Mode, and Ambient Lighting



The Shapiro-Wilk test indicated that the sink rate model residuals did not meet the normality assumption. A Box-Cox transformation with $\lambda = 2.072$ was applied. After the Box-Cox transformation, the model residuals still slightly violated normality. However, given the Shapiro-Wilk *p*-value's proximity to .05 and the absence of detected outliers, the analysis was carried out with the transformed data. The results, however, should be interpreted with caution.

The 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA with transformed data revealed no significant main effects of display type, EFVS mode, or ambient lighting on pilots' sink rate before the touchdown, and no significant interaction effects were observed ($p > .05$ for all models).

The follow-up TOST test indicated that the lower 90% confidence interval bounds (-6.827, 9.104) fell outside the equivalence bounds of (-0.821, 0.821), t (21) = 0.07, p = .527 (see Appendix B, Figure B11). Therefore, when considered alongside the non-significant main effect of display type in the RM ANOVA, sink rate before touchdown was determined to not be significantly equivalent between the HUD and Monocular HWD conditions.

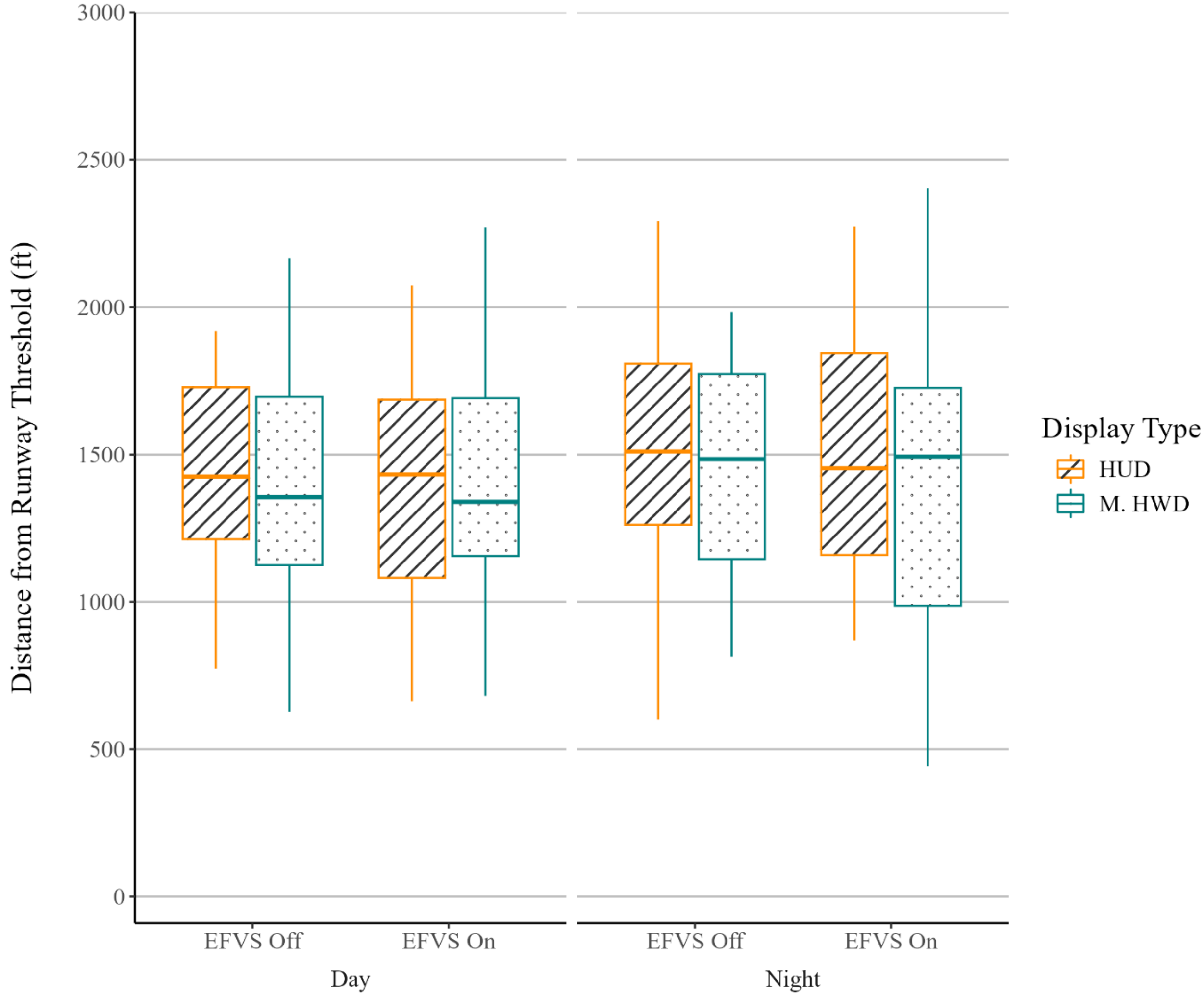
Distance from the Runway Threshold at Touchdown

The distance along the runway at which the aircraft landed was evaluated by measuring the longitudinal distance in feet between the aircraft center of gravity[9] and the runway threshold when the main gear contacted the runway surface. Figure 26 presents a Box-and-Whisker plot for the distance from the runway threshold at touchdown in each experimental condition.

[9] The aircraft center of gravity (CG) was located 3.92 feet ahead of and 10.7 feet above the ground contact point of the main landing gear.

Figure 26

Distance from the Runway Threshold at Touchdown by Display Type, EFVS Mode, and Ambient Lighting



The Shapiro-Wilk test was non-significant ($p > .05$), indicating that the model residuals were normally distributed; thus, no Box-Cox transformation was required. The 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA revealed that there were no significant main effects of display type, EFVS mode, or ambient lighting observed in the RM analysis, and no significant interaction effects were found ($p > .05$ for all models).

The follow-up TOST test indicated that both the upper and lower 90% confidence interval bounds (-147.160, 56.371) fell outside the equivalence bounds of (-68.827, 68.827), $t(21) = 0.4$, $p = .348$ (see Appendix B, Figure B12). Therefore, when considered alongside the non-significant main effect of display type in the RM ANOVA, pilots' distance from runway threshold at touchdown was determined to not be significantly equivalent between the HUD and Monocular HWD conditions.

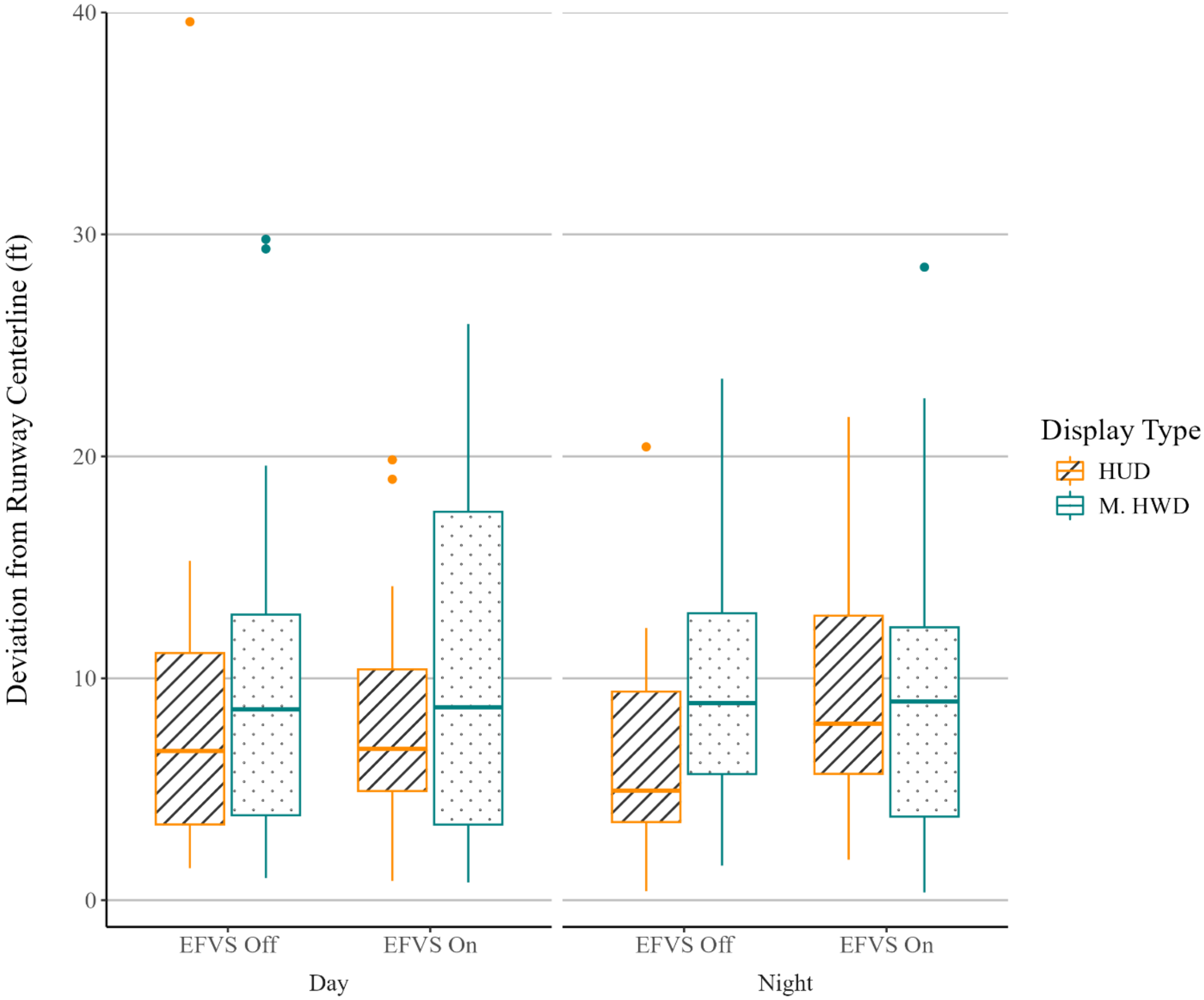
Runway Centerline Deviation

Absolute Centerline Deviation at Touchdown

The aircraft's absolute centerline deviation at touchdown was assessed by measuring the lateral deviation, in feet, between the aircraft center of gravity and the runway centerline at touchdown. Figure 27 presents a Box-and-Whisker plot for absolute centerline deviation at touchdown in each experimental condition.

Figure 27

Absolute Centerline Deviation at Touchdown by Display Type, EFVS Mode, and Ambient Lighting



The Shapiro-Wilk test indicated that the model residuals did not meet the normality assumption. Therefore, a Box-Cox transformation with $\lambda = 0.294$ was applied, resulting in transformed data that met the normality assumption. The RM ANOVA with transformed data revealed no significant main effects of display type, EFVS mode, or ambient lighting condition, and no significant interaction effects were observed ($p > .05$ for all models).

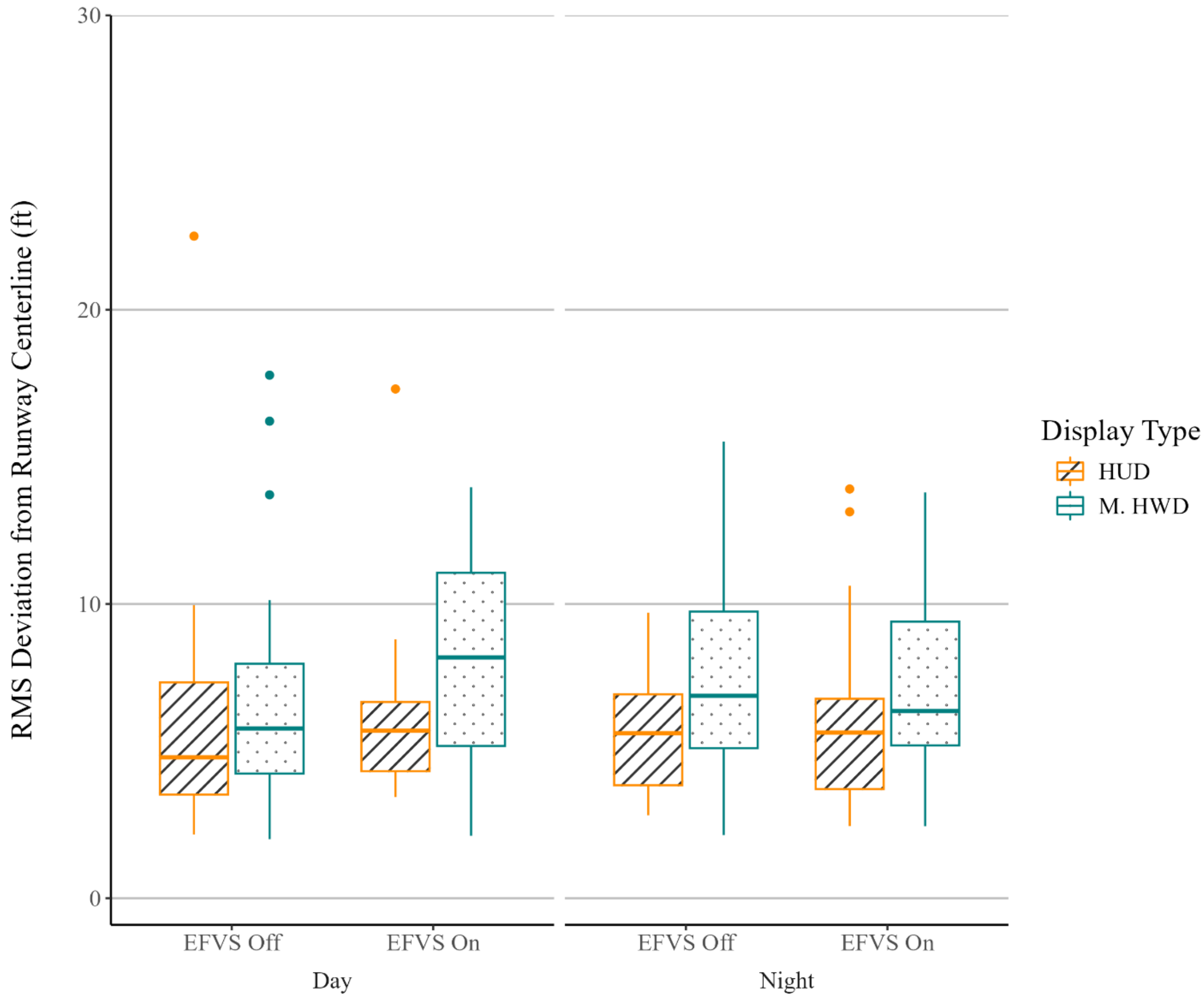
The follow-up TOST test indicated that the lower 90% confidence interval bound (-0.005, 0.718) fell outside the equivalence bounds of (-0.074, 0.074); however, the upper bound was within the boundaries, $t(21) = 1.35$, $p = .904$ (see Appendix B, Figure B13). Therefore, when considered alongside the non-significant main effect of display type in the RM ANOVA, absolute centerline deviation at touchdown was determined to not be significantly equivalent between the HUD and Monocular HWD conditions.

Root Mean Square Centerline Deviation during Rollout

The pilot's ability to maintain alignment with the runway centerline during rollout as a function of display type, EFVS mode, and ambient lighting was assessed by measuring the RMS error in feet between the aircraft center of gravity and the runway centerline from touchdown until the aircraft had decelerated to 25 knots ground speed. Figure 28 presents a Box-and-Whisker plot for RMS centerline deviation during rollout in each experimental condition.

Figure 28

RMS Centerline Deviation during Rollout by Display Type, EFVS Mode, and Ambient Lighting



Shapiro-Wilk results revealed that the residuals violated the normality assumption, so a Box-Cox transformation with $\lambda = -0.122$ was conducted to address the non-normality. The Box-Cox transformation resulted in a normal distribution of the residuals. The 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA revealed a significant effect of display type on pilots' RMS centerline deviation during the rollout; $F\ (1,21) = 9.79$, $p = .005$, $\eta_p^2 = .32$. Post-hoc results showed that pilots exhibited a larger RMS centerline deviation during rollout with the monocular HWD than they did with the HUD ($p = .005$, $d = 0.16$). Trends among the means between these conditions indicate that this difference, while statistically significant, may not be operationally significant, as RMS centerline deviation during rollout was 1.18 feet greater with the monocular HWD than it was with the HUD (see Table 11). There were no significant main effects of EFVS mode or ambient lighting on pilots' RMS centerline deviation during the rollout, nor were there any significant interaction effects between the experimental factors ($p > .05$ for all models).

Table 11

M_{adj} of Deviation from Runway Centerline during Rollout Between Display Type Conditions

Display Type	M_{BT}	M_T	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	5.37	1.51	0.05	1.42	1.61
M. HWD	6.55	1.68	0.05	1.58	1.78

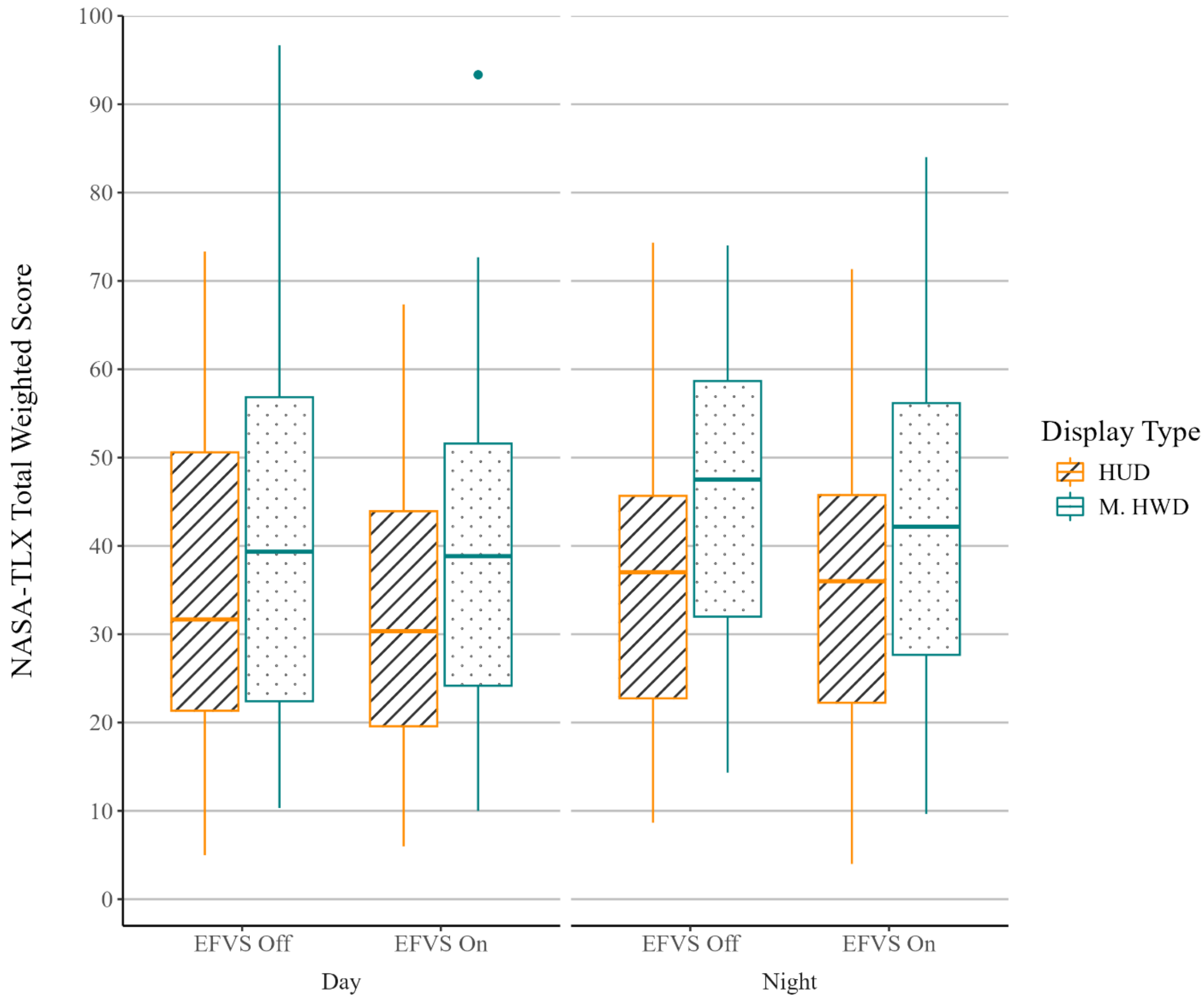
Pilot Flying Workload

NASA-TLX Total Weighted Scores

The total weighted NASA-TLX score was calculated using the procedure described by Hart and Staveland (1988), resulting in a total weighted score ranging from 0 to 100, with higher scores indicating a greater overall workload. Figure 29 presents a Box-and-Whisker plot of the NASA-TLX total weighted scores in each experimental condition.

Figure 29

NASA-TLX Total Weighted Score by Display Type, EFVS Mode, and Ambient Lighting



A 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVA was conducted to assess the effect of display type, EFVS mode, and ambient lighting on PF NASA-TLX total weighted scores. The normality assumption was met based on the results of the Shapiro-Wilk test ($p > .05$). Total weighted scores were significantly impacted by display type, $F\,(1,21) = 23.52$, $p < .001$, $\eta_p^2 = .53$. The post-hoc *t*-test revealed that mean total workload was significantly higher when pilots flew with a monocular HWD compared to when they flew with a HUD ($p < .001$, $d = 7.88$) (see Table 12). In contrast, there was no significant effect of EFVS mode or ambient lighting on the NASA-TLX total weighted score, nor were there any interaction effects between the experimental factors ($p > .05$ for all models).

Table 12

M_{adj} of NASA-TLX Total Weighted Scores Between Display Type Conditions

Display Type	*M*	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
HUD	34.39	3.52	27.07	41.72
M. HWD	42.27	4.03	33.89	50.66

NASA-TLX Subscale Scores

To further break down the impact of display type, EFVS mode, and ambient lighting on pilot workload beyond the NASA-TLX total weighted score, scores on each of the six NASA-TLX subscale scores were analyzed. The NASA-TLX subscales evaluated workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each subscale was rated on a scale from 0 to 20, with higher scores indicating a greater workload.[10] Figure 30 - Figure 35 present Box-and-Whisker plots of the mental demand, physical demand, temporal demand, performance, effort, and frustration scores, respectively, for each experimental condition.

[10] For the performance subscale in the NASA-TLX, higher scores indicate poorer performance. For the purposes of the NASA-TLX, poorer performance is interpreted as being correlated with a higher workload (Hart & Staveland, 1988).

Figure 30

Mental Demand Score by Display Type, EFVS mode, and Ambient Lighting

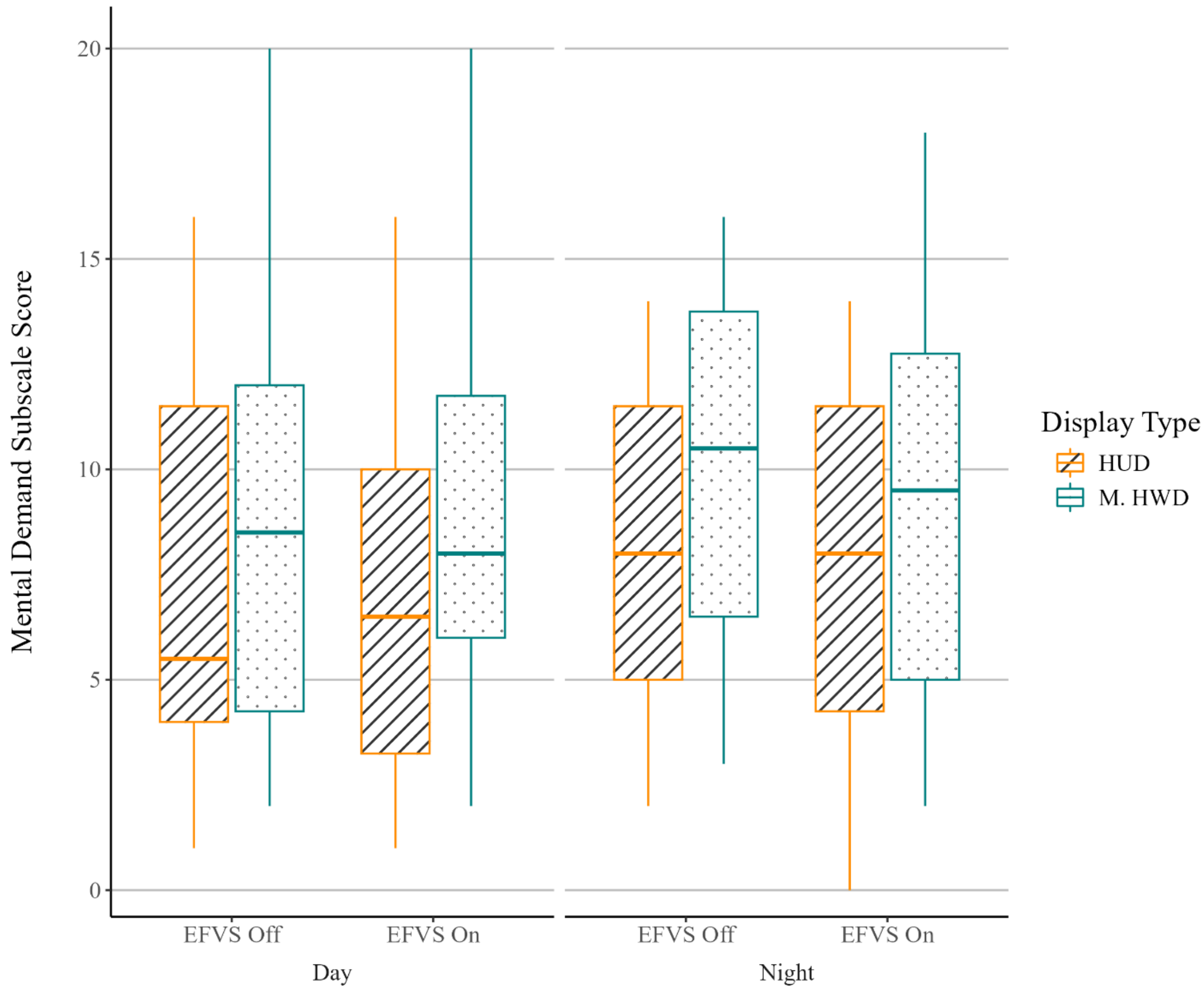


Figure 31
Physical Demand Score by Display Type, EFVS Mode, and Ambient Lighting

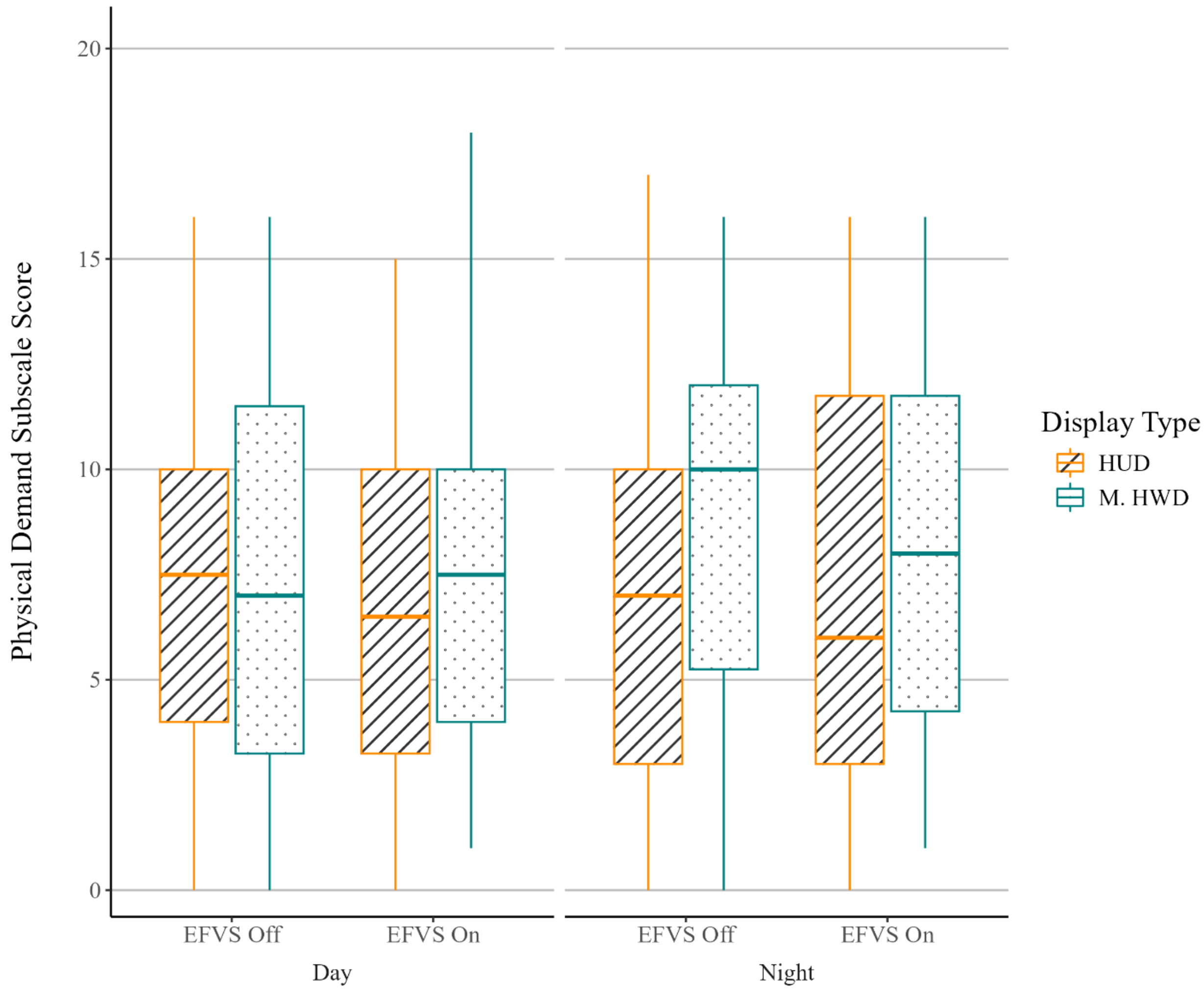


Figure 32

Temporal Demand Score by Display Type, EFVS Mode, and Ambient Lighting

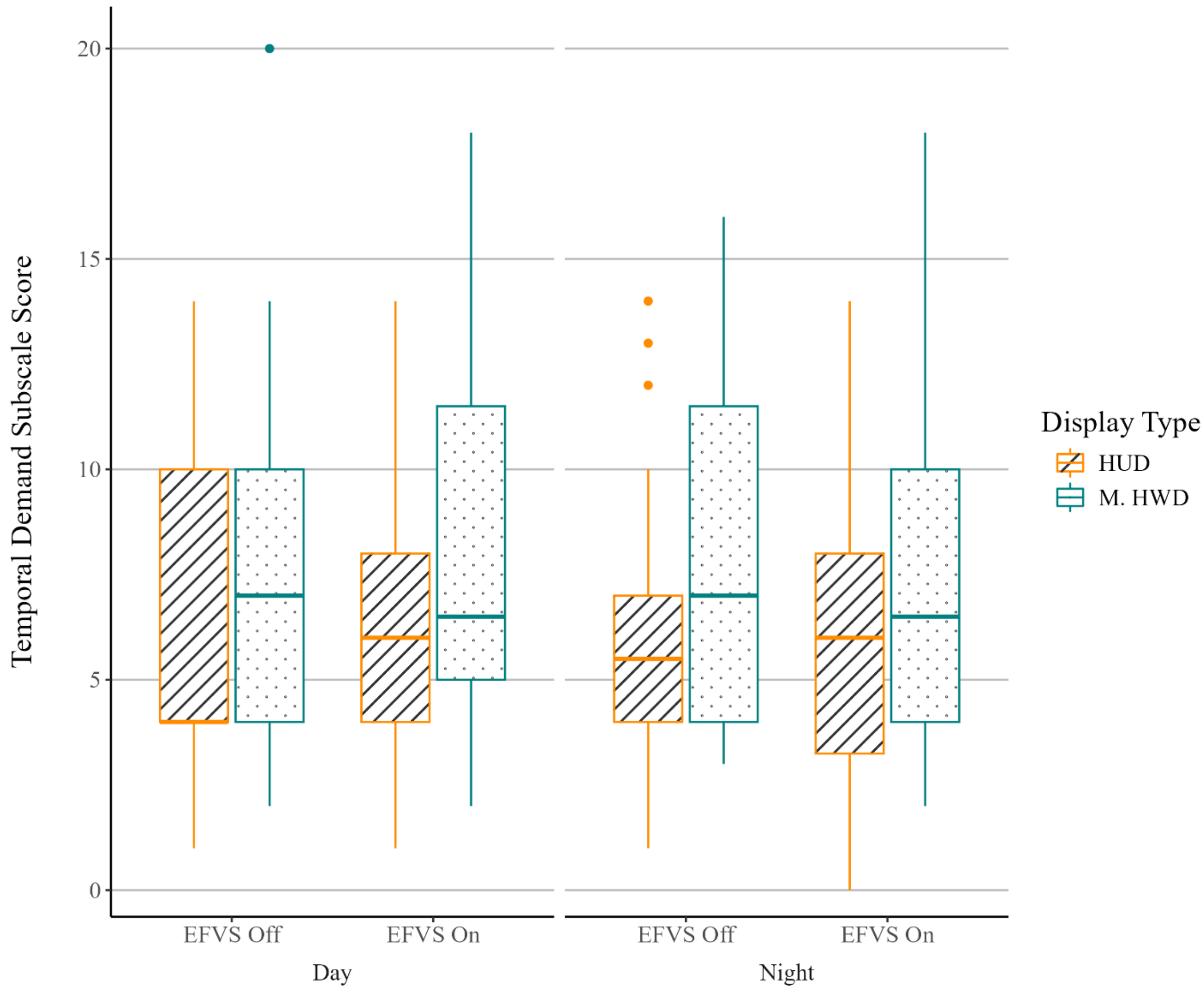
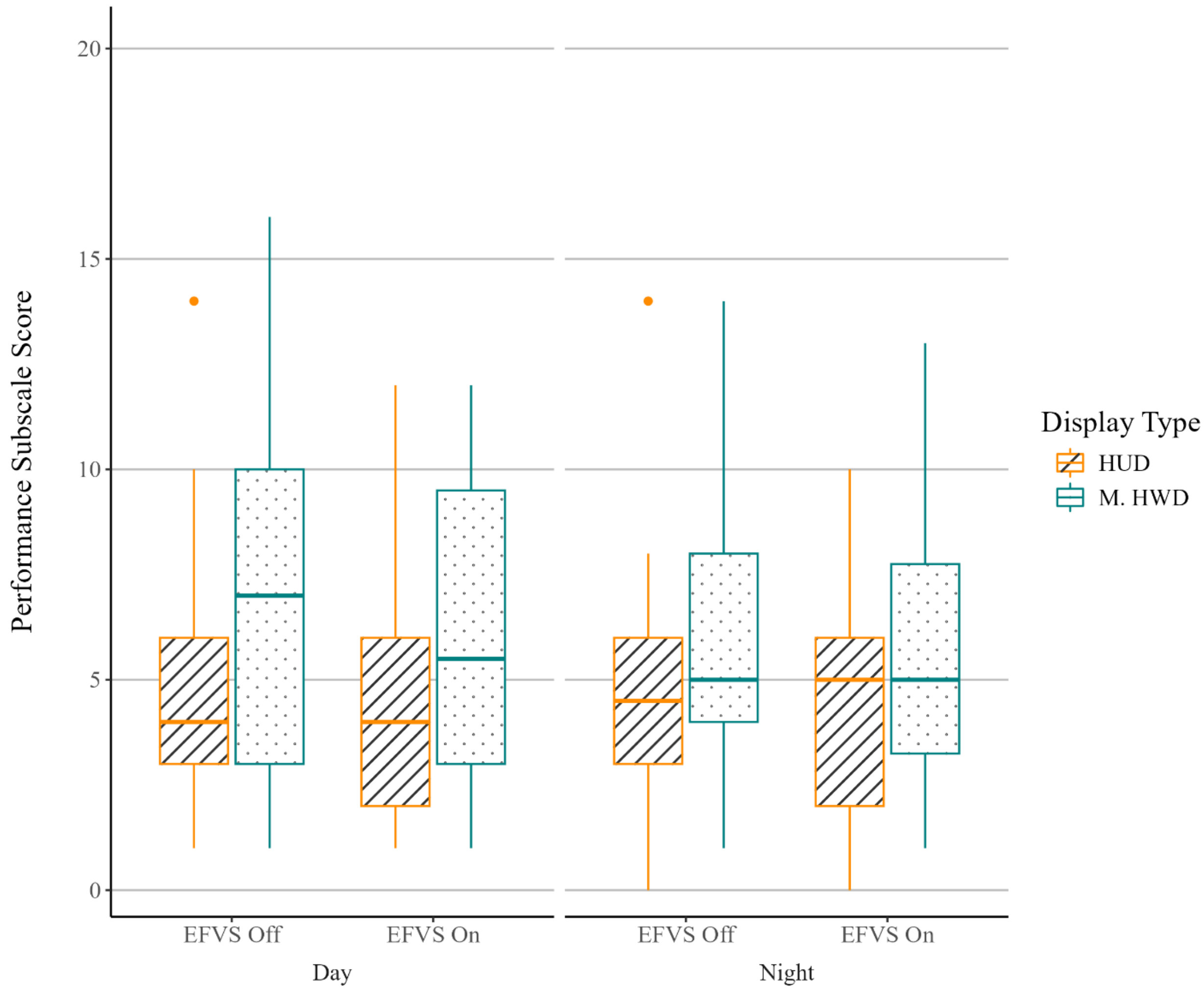


Figure 33

Performance Score by Display Type, EFVS Mode, and Ambient Lighting



In the NASA-TLX Performance subscale, 0 represents “Perfect” and 20 represents “Failure.”

Figure 34

Effort Score by Display Type, EFVS Mode, and Ambient Lighting

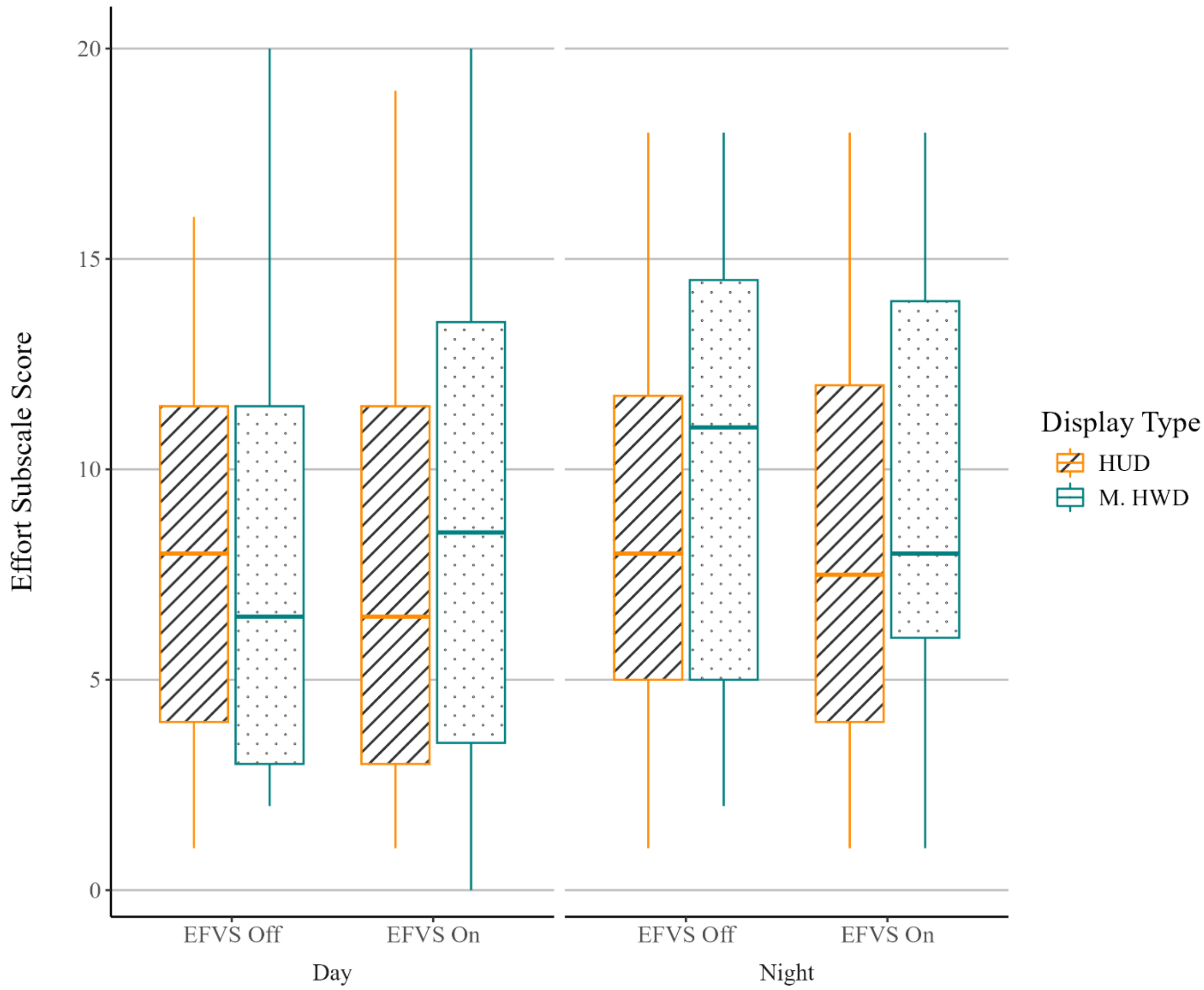
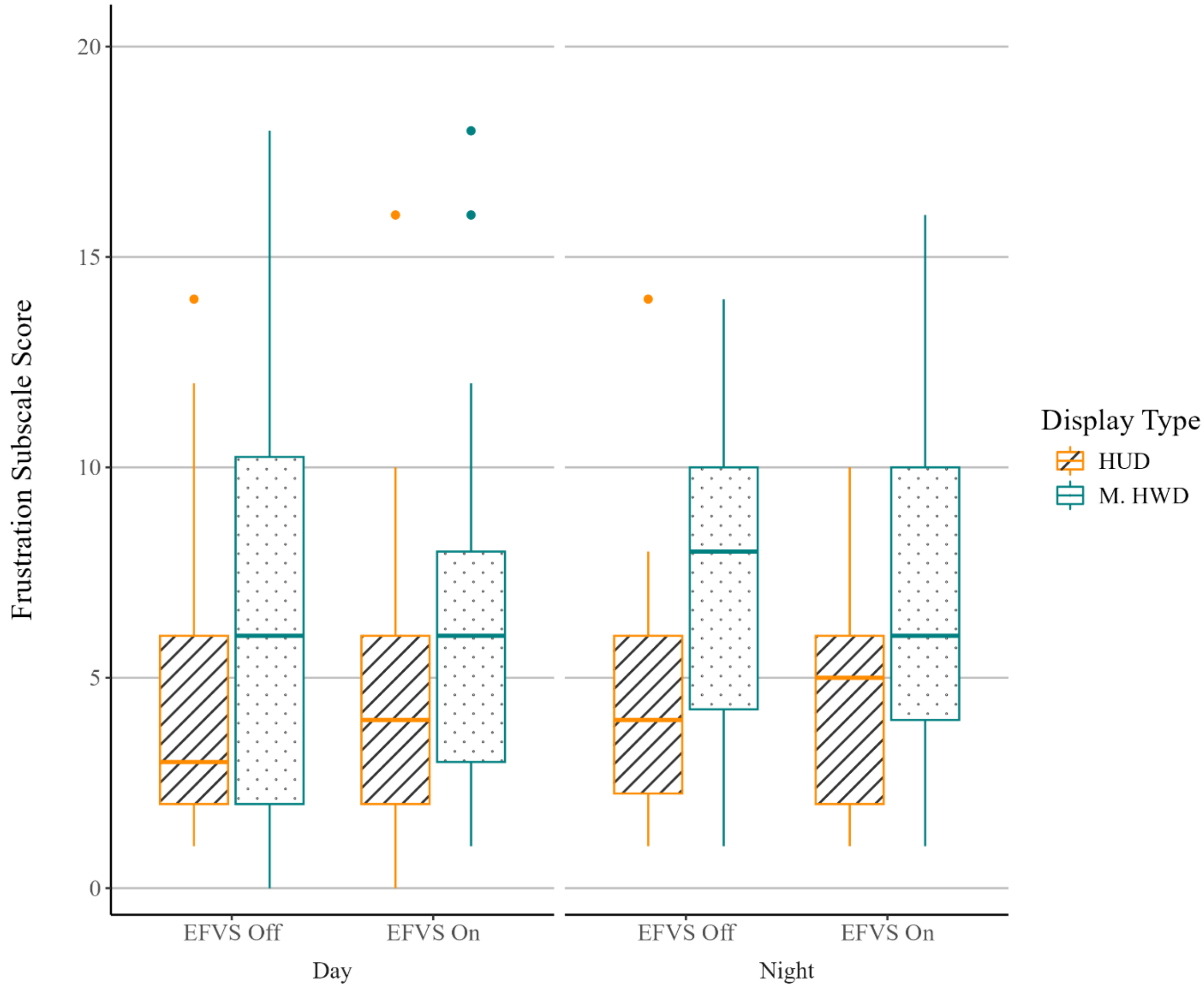


Figure 35
Frustration Score by Display Type, EFVS Mode, and Ambient Lighting



A three-way RM Multivariate Analysis of Variance (MANOVA) was initially performed to examine the effects of display type, EFVS mode, and ambient lighting on NASA-TLX subscale scores. However, significant Shapiro-Wilk test results revealed that the normality assumption was violated for at least one experimental combination in all subscales except for physical demand (see Table 13). Furthermore, significant multicollinearity was detected among some subscales (r >.8). Due to these violations of the assumptions that must be met to conduct an RM MANOVA, separate 2 (display type) × 2 (EFVS mode) × 2 (ambient lighting) RM ANOVAs were conducted for each subscale in lieu of the RM MANOVA.

Table 13

Shapiro-Wilk Test Results of Raw Scores Violating Normality Assumption

Subscale	Condition	SW statistic	df	*p*
Mental Demand	HUD/EFVS Off/Day	0.90	22	.024
Mental Demand	M. HWD/EFVS Off/Night	0.90	22	.029
Temporal Demand	HUD/EFVS Off/Day	0.87	22	.008
Temporal Demand	M. HWD/EFVS Off/Day	0.91	22	.048
Temporal Demand	M. HWD/Off/Night	0.89	22	.016
Performance	HUD/EFVS Off/Day	0.90	22	.028
Performance	HUD/EFVS Off/Night	0.90	22	.034
Effort	M. HWD/EFVS Off/Day	0.91	22	.041
Frustration	HUD/EFVS On/Day	0.86	22	.004
Frustration	HUD/EFVS On/Night	0.90	22	.031
Frustration	HUD/EFVS Off/Day	0.84	22	.002
Frustration	HUD/EFVS Off/Night	0.90	22	.028
Frustration	M. HWD/EFVS On/Day	0.91	22	.049

Scores were transformed using the Box-Cox power transformation method with a unique lambda for each of the six subscales (see Table 14). Lambdas for each subscale were calculated using the individual subscale columns. The mental demand, temporal demand, and frustration subscales were found to violate the normality assumption, even after the Box-Cox transformation was applied (see Table 15). Despite of the violation of normality in above mentioned cases, the RM ANOVA was conducted with the transformed data including outliers for the following reasons: 1) only one out of eight residuals violated normality in two models (mental demand and temporal demand), and two out of eight in the frustration subscale; 2) only minimal skewness was observed in models' residuals; 3) *p* values were close to .05; and 4) a consistent analysis across six subscales was maintained for ease of comparison. Nevertheless, caution should be exercised when interpreting the results.

Table 14
Box-Cox Transformation Lambdas per Subscale

Subscale[11]	λ
Mental Demand	0.621
Temporal Demand	0.279
Performance	0.311
Effort	0.518
Frustration	0.276

Table 15
Normality Test Results with Box-Cox Transformed Scores

Subscale	Condition	SW statistic	df	*p*	skew
Mental Demand	M. HWD/EFVS On/Night	.88	22	.010	-0.65
Temporal Demand	M. HWD/EFVS Off/Night	.89	22	.021	0.15
Frustration	HUD/EFVS On/Night	.91	22	.041	-0.11
Frustration	HUD/EFVS Off/Day	.91	22	.043	0.60

The RM ANOVAs revealed that display type had a significant impact on subscale scores for mental demand, physical demand, temporal demand, performance, and frustration. However, it did not significantly affect the effort subscale ($p > .05$). Scores associated with the following subscales were significantly affected by Display Type: mental demand, F (1,21) = 15.47, $p < .001$, $\eta_p^2 = .42$; physical demand, F (1,21) = 8.52, $p < .01$, $\eta_p^2 = .29$; temporal demand, F (1,21) = 25.68, $p < .001$, $\eta_p^2 = .55$; performance, F (1,21) = 17.01, $p < .001$, $\eta_p^2 = .45$; and frustration, F (1,21) = 67.87, $p < .001$, $\eta_p^2 = .76$.

Post-hoc analysis revealed that for all subscales except for effort, the mean workload represented by each subscale was found to be significantly greater when pilots flew with a monocular HWD compared to when they flew with a HUD. Specifically, when flying with the monocular HWD, participants reported higher mental demand ($p <.001$, $d = 0.69$), higher physical demand ($p = .008$, $d = 0.90$), higher temporal demand ($p <.001$, $d = 0.34$), poorer performance ($p <.001$, $d = 0.35$), and greater frustration ($p <.001$, $d = 0.58$) (see Table 16 – Table 20).

[11] The Physical Demand subscale met the normality assumption, so no lambda was calculated for this subscale.

No significant effects of EFVS mode or ambient lighting on NASA-TLX subscale scores were found, and no two-way or three-way interaction effects were observed in any of the subscale analyses ($p > .05$ for all models).

Table 16

M_{adj} of Mental Demand Between Display Type Conditions

Display Type	M_{BT}	M_T	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	8.25	4.36	0.38	3.56	5.16
M. HWD	9.84	5.05	0.41	4.21	5.89

Table 17

M_{adj} of Physical Demand Between Display Type Conditions

Display Type	*M*	*SE*	95% Confidence Interval	
			Lower Bound	Upper Bound
HUD	7.06	0.92	5.15	8.97
M. HWD	7.95	0.91	6.05	9.85

Table 18

M_{adj} of Temporal Demand Between Display Type Conditions

Display Type	M_{BT}	M_T	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	6.55	2.47	0.17	2.10	2.83
M. HWD	7.96	2.81	0.17	2.45	3.16

Table 19

M_{adj} of Performance Between Display Type Conditions

Display Type	M_{BT}	M_T	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	5.19	2.15	0.18	1.77	2.53
M. HWD	6.36	2.50	0.20	2.08	2.91

Table 20

M_{adj} of Frustration Between Display Type Conditions

Display Type	M_{BT}	M_T	*SE*	95% Confidence Interval	
				Lower Bound	Upper Bound
HUD	4.95	2.01	0.18	1.64	2.38
M. HWD	7.10	2.60	0.20	2.18	3.01

Usability Questionnaire

The distribution of participants' responses to each usability questionnaire item is detailed in Table 21 - Table 24, as well as Figure 36 - Figure 39. The response options for the Likert scale items were 1 ("Much Worse"), 2 ("Somewhat Worse"), 3 ("About the Same"), 4 ("Somewhat Better"), and 5 ("Much Better"). In addition to the Likert-scale items, the usability questionnaire gave participants the option to provide written feedback on the various combinations of display type and EFVS mode evaluated in this study. Thirteen participants (59.1%) elected to provide additional written feedback. The authors reviewed the responses and grouped them into categories based on themes. The most common themes that the authors identified among the responses included (a) broad comparisons between the monocular HWD and the HUD, and (b) broad comparisons between EFVS on and EFVS off. Table 25 presents thematically grouped highlights among the written feedback that participants provided in the usability questionnaire.

Table 21

Descriptive Statistics for Usability Questionnaire Items 1 – 7

Compared to flying without the EFVS on the HUD…						
	Question	*M*	Median	*SD*	Min	Max
Q1	…your ability to follow flightpath guidance symbology while flying with the EFVS on the HWD was:	3.82	4.00	.91	2	5
Q2	…your ability to maintain target airspeed while flying with the EFVS on the HWD was:	3.18	3.00	.73	1	5
Q3	…your ability to visually acquire approach/runway lighting while flying with the EFVS on the HWD was:	4.32	5.00	1.09	2	5
Q4	…your ability to visually acquire runway paint markings while flying with the EFVS on the HWD was:	3.59	4.00	1.37	1	5
Q5	…your ability to land the aircraft while flying with the EFVS on the HWD was:	3.18	3.00	1.01	2	5
Q6	…your ability to track the centerline during rollout while flying with the EFVS on the HWD was:	3.14	3.00	.94	2	5
Q7	…your ability to evaluate the safety of the runway environment while flying with the EFVS on the HWD was:	3.50	4.00	1.06	2	5

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

Figure 36

Stacked Bar Chart of Usability Questionnaire Responses Items 1 – 7

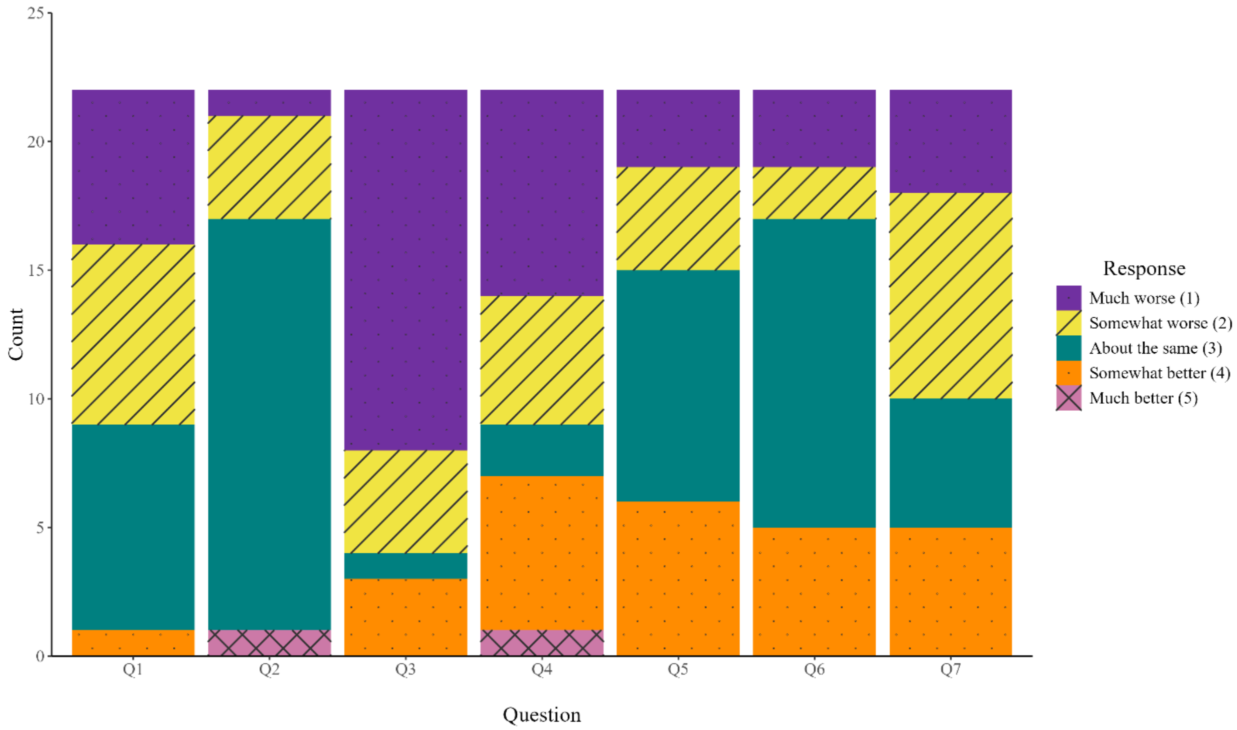


Table 22

Descriptive Statistics for Usability Questionnaire Items 8 – 14

Compared to flying without EFVS on the monocular HWD…						
	Question	***M***	**Median**	***SD***	**Min**	**Max**
Q8	...your ability to follow flightpath guidance symbology while flying with the EFVS on the monocular HWD was:	3.05	3.00	.72	2	5
Q9	...your ability to maintain target airspeed while flying with the EFVS on the monocular HWD was:	3.05	3.00	.65	1	5
Q10	...your ability to visually acquire approach/runway lighting while flying with the EFVS on the monocular HWD was:	3.77	4.00	1.07	2	5
Q11	...your ability to visually acquire runway paint markings while flying with the EFVS on the monocular HWD was:	3.27	3.50	1.16	1	5
Q12	...your ability to land the aircraft while flying with the EFVS on the monocular HWD was:	3.05	3.00	1.00	1	5
Q13	...your ability to track the centerline during rollout while flying with the EFVS on the monocular HWD was:	3.05	3.00	.90	2	5
Q14	…your ability to evaluate the safety of the runway environment while flying with the EFVS on the monocular HWD was:	3.09	3.00	1.07	1	5

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

Figure 37

Stacked Bar Chart of Usability Questionnaire Responses Items 8-14

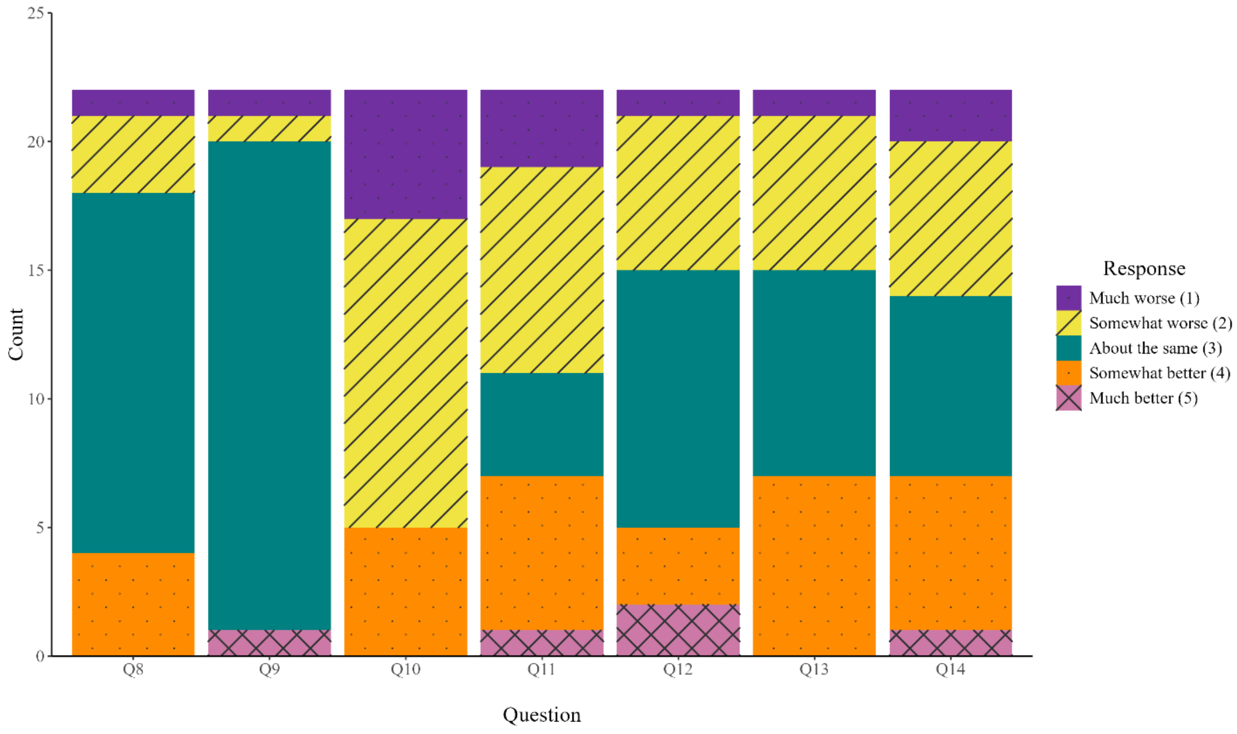


Table 23

Descriptive Statistics for Usability Questionnaire Items 15 – 21

Compared to flying with the HUD without EFVS…						
	Question	***M***	**Median**	***SD***	**Min**	**Max**
Q15	...your ability to follow flightpath guidance symbology while flying with the monocular HWD without the EFVS was:	3.00	3.00	.76	2	4
Q16	...your ability to maintain target airspeed while flying with the monocular HWD without the EFVS was:	3.00	3.00	.54	2	4
Q17	...your ability to visually acquire approach/runway lighting while flying with the monocular HWD without the EFVS was:	3.05	3.00	.72	2	5
Q18	...your ability to visually acquire runway paint markings while flying with the monocular HWD without the EFVS was:	2.91	3.00	.75	1	4
Q19	...your ability to land the aircraft while flying with the monocular HWD without EFVS was:	2.82	3.00	.50	2	4
Q20	...your ability to track the centerline during rollout while flying with the monocular HWD without the EFVS was:	3.14	3.00	.56	2	4
Q21	...your ability to evaluate the safety of the runway environment while flying with the monocular HWD without the EFVS was:	2.91	3.00	.75	1	4

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

Figure 38

Stacked Bar Chart of Usability Questionnaire Responses Items 15-21

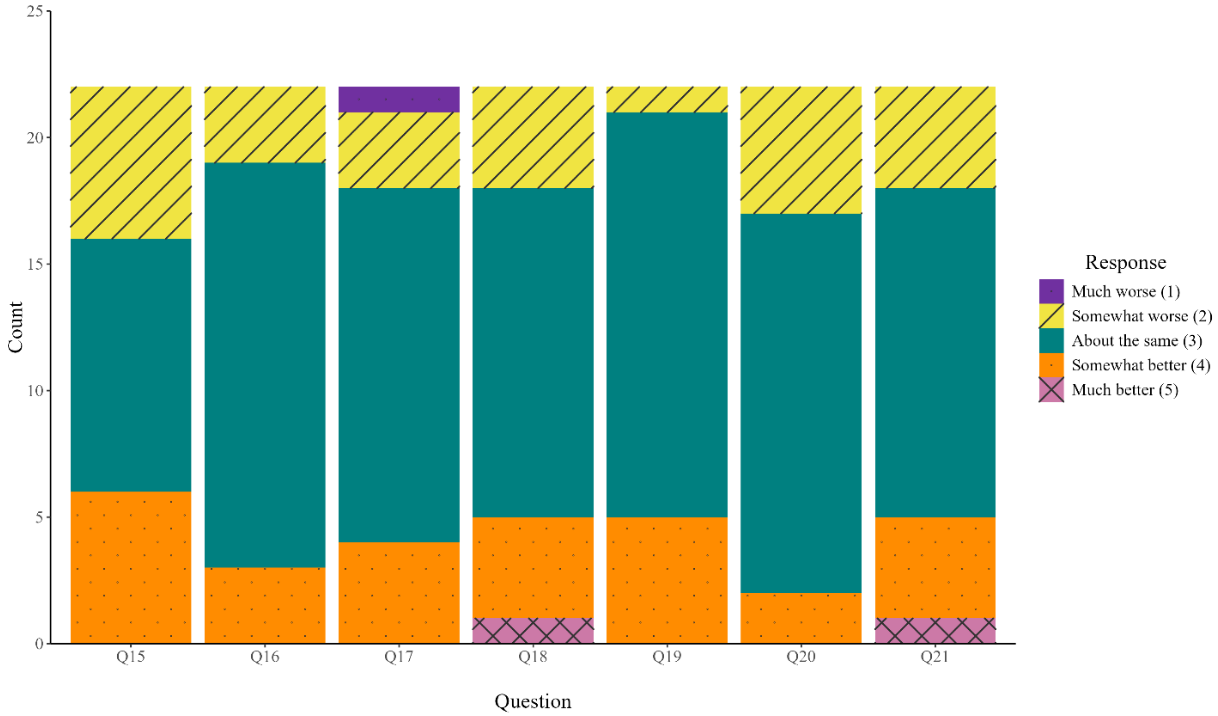


Table 24

Descriptive Statistics for Usability Questionnaire Items 22 – 28

	Compared to flying with the monocular HWD without EFVS…					
	Question	***M***	**Median**	***SD***	**Min**	**Max**
Q22	...your ability to follow flightpath guidance symbology while flying with the monocular HWD with the EFVS was:	3.23	3.00	.75	2	5
Q23	...your ability to maintain target airspeed while flying with the monocular HWD with the EFVS was:	3.09	3.00	.75	1	5
Q24	...your ability to visually acquire approach/runway lighting while flying with the monocular HWD with the EFVS was:	3.32	3.00	.84	2	5
Q25	...your ability to visually acquire runway paint markings while flying with the monocular HWD with the EFVS was:	3.27	3.00	.88	2	5
Q26	...your ability to land the aircraft while flying with the monocular HWD with EFVS was:	2.82	3.00	.80	1	5
Q27	...your ability to track the centerline during rollout while flying with the monocular HWD with the EFVS was:	3.09	3.00	.75	2	5
Q28	...your ability to evaluate the safety of the runway environment while flying with the monocular HWD with the EFVS was:	3.18	3.00	.91	1	5

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

Figure 39

Stacked Bar Chart of Usability Questionnaire Responses Items 15-21

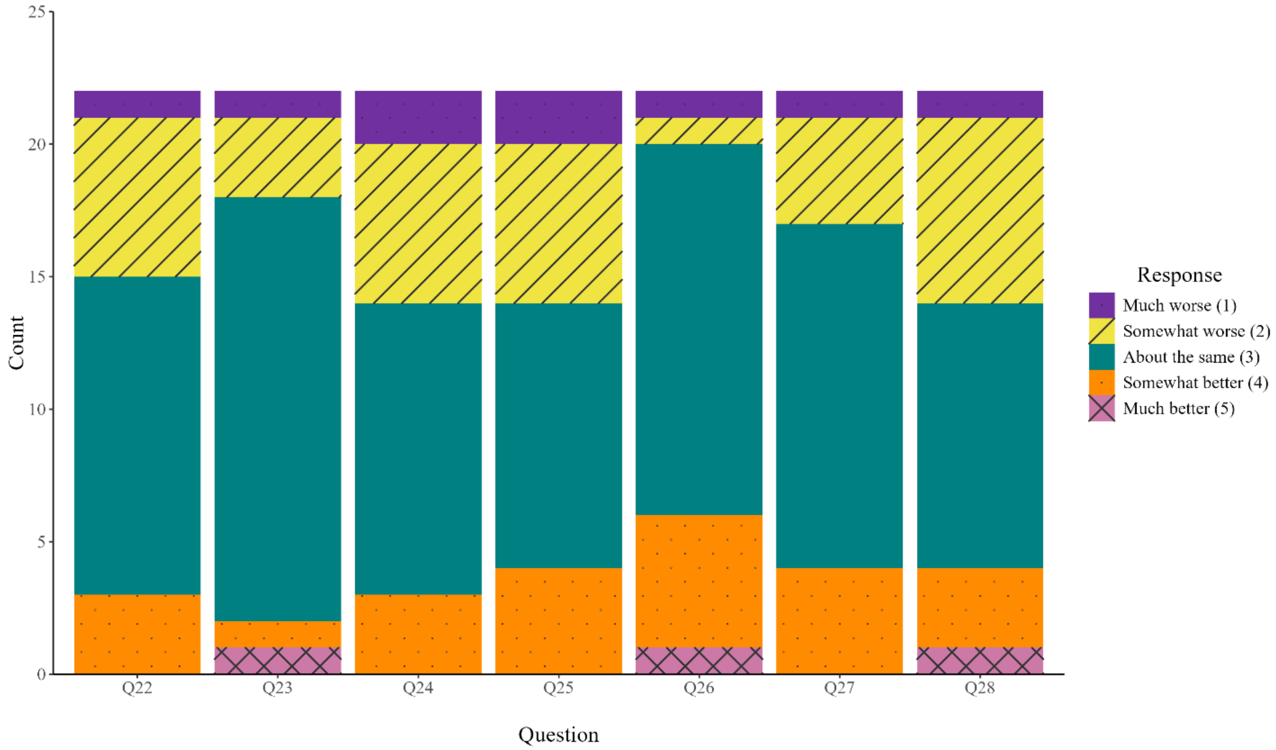


Table 25

Themes and Highlighted Written Feedback from Participants

Monocular HWD Versus HUD
• EFVS on the HUD was much preferred. • Any enhancements to vision day or night is always good but a cost vs use will have to be determined. • Love the EFVS. [It] gives you great situational awareness • The HWD is definitely more difficult than the HUD as far as the approach and especially the transition to landing. Part of this is the ability to better see “through” the HUD plus the wider general field of view in the HUD, and then partially the monocular HWD only giving the info to one eye which is harder for my brain to interpret. • With the [HWD] there was some double vision or parallax of the EFVS my right eye and my left unaided left eye. • [HWD] gave certain advantages however also gave some challenges like monocular vision • The symbology was the same in both [display types] but the HUD just provides better clarity, field of view, and a stereo optical experience that is much better and easier to use than the monocular display. • HWD [h]as a more difficult crosscheck but results were about the same • Much prefer HUD [over HWD] with or without EFVS. • I liked the [HWD]. • Not a fan of the Head-Worn display. The symbology is not as sharp especially under daylight conditions. • Challenge on Head-Worn display was monocular • The head worn display tends to make the entire approach and landing experience more difficult for me as it limits my field of view and depth perception. • HUD … is light years better than HWD. • [I experienced] double vision or parallax with [the HWD].
EFVS On Versus EFVS Off
• EFVS made situation awareness much higher. • [With both the HUD and the HWD] the ability to declutter EFVS data either manually or automatically at a certain height would greatly enhance the transition to flare for the PF. The EFVS data makes it difficult to acquire the visual cues normally used to accomplish this. • While the pilot flying will have more information [with the EFVS], it will be just as important to keep the pilot monitoring in the loop when picking up the approach lights or the landing environment. • Subjectively, in order I liked HUD no EVS, HUD with EVS. Most of the issues with EVS can be addressed with a kill switch for use just prior to the flare. • [EVS] helped me capture the runway environment quicker. However, once the runway was in sight, EVS hampered my efforts to land and maintain centerline. • [With both the HUD and the HWD] EFVS provides better SA in acquiring the runway environment • With EVFS, approach lights and runway are much easier to acquire, as there is no real transition that typically occurs in reduced visibility.

- EVS is a big help in acquiring the runway lights, but a hindrance in the flare and roll out.
- EVS helped me to more quickly acquire the runway environment. However, once the runway was in sight, the distractions caused by the limitations of the head worn display, coupled with the ‘fixation effect’ of the EVS, made landing and rollout much more difficult.
- Some distraction from physical CL lights diverging from the [EFVS] virtual CL lights drove a focus to the near runway environment verses a crosscheck of near to far and airspeed versus runway position and runway remaining.
- An auto blanking system at X ft or Weight on wheels would be good.
- Only thing substantially better [with the EFVS] was the ability to acquire lights when below dashboard level.
- Less depth perception but better runway awareness [with the EFVS]
- Love the EFVS gives you great situational awareness

Discussion

The present study focused on determining pilot performance and workload levels while flying SA CAT I approach, landing, and rollout operations in day and night ambient lighting conditions with a monocular HWD with and without an EFVS, with the goal of identifying whether those factors differed substantially from when pilots flew those same operations with a HUD with or without an EFVS. The following section summarizes the findings from this research as they relate to the initial research questions and presents key operational takeaways of the findings.

Findings

RQ$_1$: What degree of flightpath and energy management accuracy is present during a manual SA CAT I operation using a monocular HWD with and without an EFVS?

Across all analyses of flightpath and energy management performance during the approach, there were statistically significant differences in glideslope deviation among the study conditions; however, none of these differences appeared to reach operational significance. In the instrument segment, RMS glideslope deviation was higher when flying with a monocular HWD than when flying with a HUD; however, the difference was small. RMS glideslope deviation was 1.39 feet greater, on average, with a monocular HWD than with a HUD (see Table 8). This trend in glideslope deviation was not present when crossing the DH and 100 feet above the TDZ; however, pilots crossed the runway threshold with a small (1.91 feet) increase in glideslope deviation when flying with the HUD compared to when flying with the monocular HWD (see Table 10). Pilots were able to successfully correct any differences in glideslope deviation due to display type before landing, as evidenced by a lack of difference in touchdown performance across all combinations of display type, EFVS mode, and ambient lighting. During the visual segment, there was a significant—yet small magnitude—impact of ambient lighting: On average, pilots crossed the DH 2.85 feet above the glideslope during nighttime scenarios compared to 1.48 feet above the glideslope during daytime scenarios (see Table 9); however, this pattern was not present elsewhere in the data and the very small difference is not operationally significant.

Taken together, while some performance measures appeared to be impacted by display type and ambient lighting, there were no consistent trends, and pilots appeared able to correct any temporary flightpath deviations before landing. The takeaway from this set of findings is that there is existing variability in pilot flightpath and energy management performance that occurs due to differences in environmental conditions, such as ambient light levels and wind conditions. Additional variability in performance that occurs as a function of whether pilots fly with a HUD or a monocular HWD may be present as well; however, it may not supersede any existing variability caused by environmental factors. Moreover, the consistency in performance regardless of whether the EFVS was off or on suggests that an EFVS with sufficient visual advantage enables pilots to fly an SA CAT I approach with as little as 1000 feet RVR without causing significant changes in performance from what is present during an SA CAT I approach flown without an EFVS.

RQ_2: What degree of touchdown zone dispersion at landing is present during a manual SA CAT I operation using a monocular HWD with and without an EFVS?

Results from landing performance analyses in this research suggest that pilots may exhibit a similar degree of TDZ dispersion and sink rate at touchdown during an SA CAT I operation with and without an EFVS when flying with a monocular HWD as they exhibit when flying with a HUD. The results of this set of analyses suggest that pilots may be able to land the aircraft in the runway TDZ during an SA CAT I approach with and without an EFVS just as well when flying with a monocular HWD as when flying with a HUD. However, pilot feedback on the interference of EFVS imagery during landing suggests that the ability to disable the EFVS imagery once the aircraft has descended below 100 feet above the TDZE may improve EFVS usability and acceptance. Indeed, several pilots commented that their ability to flare and derotate the aircraft would have been improved if they had been able to switch the EFVS imagery off after descending below 100 feet.

RQ_3: What degree of runway centerline deviation during rollout is present during a manual SA CAT I operation using a monocular HWD with and without an EFVS?

During rollout, pilots exhibited small increases in deviation from the runway centerline when flying with a monocular HWD compared to when flying with a HUD. Post-hoc *t*-tests and trends among the means suggest that this effect is present regardless of whether the EFVS was on or off, and regardless of the ambient lighting during the scenario. This effect might have been driven by differences between the non-collimated HWD and the collimated HUD used in this research. Past research suggests that a mismatch in focal distance between information on a HUD and HWD and the background scenery can hinder the ability to divide attention between the two, which is a process that is involved with maintaining lateral control of the aircraft after landing using HUD and HWD symbology and imagery in tandem with runway centerline lights and markings (Newton et al., 2026; Weintraub & Endsing, 1992).

In addition to the performance analysis, verbal feedback and responses to the usability questionnaire from pilots during the study contained a theme: they found that the EFVS was a hindrance once the aircraft had touched down, making it more difficult for them to flare, land, and track the runway centerline. Pilots indicated that this may have been a result of the EFVS impacting their ability to perceive depth cues using natural vision. However, this feedback did not appear to translate into significant impacts on performance, as there was no significant difference in runway centerline deviation during rollout as a function of EFVS mode. The results from this analysis suggest that authorizations regarding the use of a monocular HWD for SA CAT I operations should be developed in a manner that is sensitive to the potential for differences in runway centerline tracking performance with some HWD configurations, as well as the potential for an EFVS to provide a benefit during the transition to visual flight references but to be a hindrance once visual references are established and the pilot transitions to landing.

RQ$_4$: What level of workload does the Pilot Flying experience during a manual SA CAT I operation using a monocular HWD with and without an EFVS?

Pilots reported their workload was significantly higher when flying with a monocular HWD than when flying with a HUD. This replicates the findings of Newton et al. (2026) and extends those findings to nighttime ambient lighting levels and to operations flown with an EFVS. This finding also aligns with early research on civil aviation HWDs and suggests that increased effort is required to use flight symbology presented on some near-to-eye display concepts 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 five of the six NASA-TLX subscale ratings: Mental Demand, Physical Demand, Temporal Demand, Performance, and Frustration.

Trends in pilot workload as a function of whether they flew with a HUD or an HWD suggest that the physical and optical characteristics of the HWD in this study created additional sources of workload. In past research, increased workload during the use of extended reality headsets has been linked to the additional weight of the headset (Drouot et al., 2022). This may have been the case in this research as well. This effect might have been exacerbated by the absence of a collimated display in the HWD used in this study. Conversely, there were no significant impacts of EFVS mode or ambient lighting on pilot workload ratings. This suggests that the use of an EFVS does not change pilots' perception of task demands during an SA CAT I operation, even when that operation is conducted at the visibility minima for an EFVS SA CAT I operation (i.e., 1000 ft TDZ RVR). It also suggests that the ambient lighting levels during an SA CAT I operation do not contribute to pilots' overall perception of task demand.

Usability Questionnaire and Pilot Feedback

Participants' feedback on the usability and utility of the monocular HWD and the EFVS aligned with the performance and workload findings, with some key exceptions. Participants verbally reported that greater effort was required to fly the approach when using the monocular HWD, citing that the monocular configuration of the HWD made it more difficult to use the symbology and EFVS image, particularly during the rollout phase of the operation. Some pilots reported that this experience subsided as the study session progressed, suggesting a possible learning and practice effect. Participants reported that the EFVS sensor image made it easier to transition from instrument to visual flight references prior to the DH; however, a primary point of feedback was that the EFVS hindered the ability to see runway visual features with natural vision when landing the aircraft and when tracking the runway centerline during rollout, especially when it was presented on the monocular HWD. A common point of feedback among the participants was that it would have been beneficial if they were able to clear the EFVS sensor image from the HUD or HWD after approach and runway lighting could be seen with natural vision.

Limitations

There are several limitations to this research that should be considered alongside the findings. A primary limitation of this research was that the monocular HWD, a Microsoft HoloLens 2, was not originally designed for use on a flight deck and may not best represent HWD technology that will be used in an operational setting by ATPs. The use of the HoloLens 2 bolstered the internal validity of the study by presenting identical flight symbology regardless of whether pilots flew with a HUD or a monocular HWD, thus avoiding a major confound in the research design that would be present if the symbology differed between the display types. Nevertheless, some aspects of its design may limit the generalizability of these research findings to all certified, production-quality aviation HWDs.

In particular, some study participants commented that the HWD imagery appeared to be lower resolution than the HUD imagery, which may have been a source of distraction during the scenarios. Moreover, the HoloLens 2 did not offer the same range of brightness adjustability as the HUD, so some participants commented that brightness levels were higher than they would prefer, particularly during nighttime scenarios flown with the EFVS. Participants also commented on the prominent combiner of the HWD, which restricted their view of other displays on the flight deck. Additionally, although participants did not comment on the disparity in image focal distance between the HUD and HWD, the use of a non-collimated HWD in comparison to a collimated HUD is an important factor when interpreting the findings of this study. While display collimation could not be fully controlled in this study, it is a foundational aspect of flight deck HUD design and, therefore, may have contributed to patterns in performance and workload data in this research. The specific characteristics of the HWD used in this study, including the non-collimated display, may differ from those of HWDs designed to be certified for use on a flight deck. As such, future research should continue investigating pilot performance and workload while flying with a production aviation HWD that features a collimated 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.

Conclusions and Future Directions

This research contributes to the understanding of how pilot performance and workload during SA CAT I operations with and without an EFVS may be impacted when pilots fly with a monocular HWD in lieu of a HUD. The findings from this research suggest that pilots are able to manually fly SA CAT I approach, landing, and rollout operations with approximately the same degree of flightpath and energy management accuracy during most phases of the operation when using a monocular HWD compared to when using a HUD. This is the case when the weather is at SA CAT I minima for operations conducted with HUD symbology (i.e., 1400 feet RVR) as well as when the weather is at minima for operations with an EFVS (i.e., 1000 feet RVR). It appears that a monocular HWD may not have an impact on the pilot's ability to manage the lateral and vertical flightpath and maintain the target IAS during most phases of an SA CAT I approach; however, pilots experienced elevated workload levels. This trend in workload data is consistent with findings from previous research and may primarily be driven by the monocular configuration of the HWD, as well as the non-collimated image on the HWD (Newton et al., 2026). Pilots also exhibited a greater degree of glideslope deviation during the instrument segment, glideslope deviation at the DH, and centerline deviation during rollout when flying with the monocular HWD; however, these differences were small in magnitude and may not carry substantial practical significance.

The EFVS had no measurable impact on pilot performance and workload in this study, suggesting that the reduced flight visibility authorizations for EFVS use may not negatively impact the safety of a routine SA CAT I approach and landing, however pilot feedback on usability reinforces the value of being able to clear the EFVS image below the DH, so the pilot can benefit from improved detection of runway visual features before the DH from the EFVS, after which point the sensor image could be cleared by the pilot to facilitate natural visibility of the runway during landing and rollout. This function is currently required for HUD EFVS installations. Ultimately, this research suggests that pilots should be able to fly SA CAT I approach, landing, and rollout operations with and without an EFVS using a monocular HWD and exhibit performance levels similar to those when they fly those same operations with a HUD. Crucially, however, pilots experience elevated workload levels when flying with a monocular HWD. As such, it would be important to develop future operational authorizations for HWD use during SA CAT I operations in a manner that is sensitive to the potential for increased workload and runway centerline deviation during rollout compared to currently authorized SA CAT I operations using a HUD with and without an EFVS.

References

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*.

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., & Fercho, K. A. (2020). The use of Enhanced Flight Vision Systems (EFVS) for low-visibility takeoffs in Part 121 operations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 64*(1), 198–202. https://doi.org/10.1177/1071181320641048

Beringer, D. B., Sparko, A. L., & Jaworski, J. M. (2019). Using Enhanced Flight Vision Systems (EFVS) for low-visibility taxi in transport category aircraft. In *Proceedings of the 2019 International Symposium on Aviation Psychology*.

Boeing Aircraft Co. (2025, April). *Statistical summary of commercial jet airplane accidents* (1959-2024). Arlington, VA: Boeing Aircraft Co.

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

Drouot, M., Le Bigot, N., Bricard, E., De Bougrenet, J. L., & Nourrit, V. (2022). Augmented reality on industrial assembly line: Impact on effectiveness and mental workload. *Applied Ergonomics*, *103*, 103793.

Federal Aviation Administration. (2016). *Airworthiness approval of Enhanced Vision System, Synthetic Vision System, Combined Vision System, and Enhanced Flight Vision System Equipment* (Advisory Circular No. 20-167A). Washington, D.C.: Federal Aviation Administration.

Federal Aviation Administration. (2022a). *Report to Congress: Advanced cockpit displays*. Washington, D.C.: Federal Aviation Administration.

Federal Aviation Administration. (2022b). *Enhanced Flight Vision System operations* (Advisory Circular No. 90-106B). Washington, D.C.: Federal Aviation Administration.

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

Fercho, K. A., & Watson, B. L. (2024). *Federal Aviation Administration Aircraft Evaluation Division operations and human factors evaluation aid for Enhanced Flight Vision*

Systems. (Report No. DOT/FAA/AM-24/21). Washington, D.C.: Federal Aviation Administration.

Federal Aviation Administration National Simulator Program. (2012). *Head-up display (HUD) simulator qualification* (FSTD Guidance Bulletin 03-02). Washington, D.C.

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.

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

Hauck, W. W., & Anderson, S. (1984). A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *Journal of Pharmacokinetics and Biopharmaceutics*, *12*(1), 83–91.

IBM Corp. Released 2011. *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY: IBM Corp.

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-

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

Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355-362. https://doi.org/10.1177/1948550617697177

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).

Mumaw, R. J., Billman, D., & Feary, M. S. (2019). *Factors that influenced airplane state awareness accidents and incidents: CAST SE-210 output 2: Report 2 of 6* (Report No. NASA/TM-20205010985). Moffett Field, CA: National Aeronautics and Space Administration.

Newton, D. C., Greenlee, E. T., Mofle, T. C., Shahini, F., DiDomenica, R., Choi, I., & Weaver, R. (2026, January). *Use of a monocular and binocular head-worn display in lieu of a head-up display for approach, landing, and rollout: Human factors evaluation of pilot performance and workload* (Report No. DOT/FAA/AM-26/02). Washington, D.C.: Federal Aviation Administration.

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

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

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.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics, 15*(6), 657–680. https://doi.org/10.1007/BF01068419

Seaman, M. A., & Serlin, R. C. (1998). Equivalence confidence intervals for two-group comparisons of means*. Psychological Methods, 3*(4), 403–411. https://doi.org/10.1037/1082-989X.3.4.403

Straight-in landing operations below DA/DH or MDA using an Enhanced Flight Vision System (EFVS) under IFR, 14 C.F.R. § 91.176 (2017). https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-B/subject-group-ECFRef6e8c57f580cfd/section-91.176

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.

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

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.

Williams, K. W., Mofle, T. C., & Choi, I. (2021). *Evaluating the preventive alert function for UAS detect and avoid systems* (Report No. DOT/FAA/AM-21/17). Washington, D.C.: Federal Aviation Administration.

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.

Appendix A

Usability Questionnaire

Based on your experience as Pilot Flying today, please rate your experience on the following scale:

1 = “much worse” 2 = “somewhat worse” 3 = “about the same” 4 = “somewhat better” 5 = “much better”

Flight Symbology Only vs. Enhanced Flight Vision System (EFVS) on the HUD

Compared to flying without EFVS on the HUD, your ability to **follow flightpath guidance** symbology while flying with EFVS on the HUD was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the HUD, your ability to **maintain target airspeed** while flying with EFVS on the HUD was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the HUD, your ability to **visually acquire approach/runway lighting** while flying with EFVS on the HUD was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the HUD, your ability to **visually acquire runway paint markings** while flying with EFVS on the HUD was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the HUD, your ability to **land the aircraft** while flying with EFVS on the HUD was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the HUD, your ability to **track the centerline during rollout** while flying with EFVS on the HUD was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the HUD, your ability to **evaluate the safety of the runway environment** while flying with EFVS on the HUD was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

In the Qualtrics survey, a text box appeared after the preceding set of items with the yes/no prompt, "would you like to provide additional comments on flying with EFVS versus without EFVS on the HUD?" If response was "yes," a text box for open-ended written feedback appeared.

Flight Symbology Only vs. Enhanced Flight Vision System (EFVS) on the Head-Worn Display

Compared to flying without EFVS on the Head-Worn Display, your ability to **follow flightpath guidance** symbology while flying with EFVS on the Head-Worn Display was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the Head-Worn Display, your ability to **maintain target airspeed** while flying with EFVS on the Head-Worn Display was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the Head-Worn Display, your ability to **visually acquire approach/runway lighting** while flying with EFVS on the Head-Worn Display was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the Head-Worn Display, your ability to **visually acquire runway paint markings** while flying with EFVS on the Head-Worn Display was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the Head-Worn Display, your ability to **land the aircraft** while flying with EFVS on the Head-Worn Display was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the Head-Worn Display, your ability to **track the centerline during rollout** while flying with EFVS on the Head-Worn Display was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying without EFVS on the Head-Worn Display, your ability to **evaluate the safety of the runway environment** while flying with EFVS on the Head-Worn Display was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

In the Qualtrics survey, a text box appeared after the preceding set of items with the yes/no prompt, "would you like to provide additional comments on flying with EFVS versus without EFVS on the Head-Worn Display?" If response was "yes," a text box for open-ended written feedback appeared.

HUD vs. Head-Worn Display with Flight Symbology Only (No EFVS)

Compared to flying with the HUD without EFVS, your ability to **follow flightpath guidance symbology** while flying with the Head-Worn Display without EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD without EFVS, your ability to **maintain target airspeed** while flying with the Head-Worn Display without EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD without EFVS, your ability to **visually acquire approach/runway lighting** while flying with the Head-Worn Display without EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD without EFVS, your ability **visually acquire runway paint markings** while flying with the Head-Worn Display without EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD without EFVS, your ability to **land the aircraft** while flying with the Head-Worn Display without EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD without EFVS, your ability to **track the centerline during rollout** while flying with the Head-Worn Display without EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD without EFVS, your ability to **evaluate the safety of the runway environment** while flying with the Head-Worn Display without EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better

In the Qualtrics survey, a text box appeared after the preceding set of items with the yes/no prompt, "would you like to provide additional comments on HUD versus Head-Worn Display without EFVS?" If response was "yes," a text box for open-ended written feedback appeared.

HUD vs. Head-Worn Display with Enhanced Flight Vision System (EFVS)

Compared to flying with the HUD with EFVS, your ability to **follow flightpath guidance symbology** while flying with the Head-Worn Display with EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD with EFVS, your ability to **maintain target airspeed** while flying with the Head-Worn Display with EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD with EFVS, your ability to **visually acquire approach/runway lighting** while flying with the Head-Worn Display with EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD with EFVS, your ability **visually acquire runway paint markings** while flying with the Head-Worn Display with EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD with EFVS, your ability to **land the aircraft** while flying with the Head-Worn Display with EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD with EFVS, your ability to **track the centerline during rollout** while flying with the Head-Worn Display with EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

Compared to flying with the HUD with EFVS, your ability to **evaluate the safety of the runway environment** while flying with the Head-Worn Display with EFVS was:

Much worse	Somewhat worse	About the same	Somewhat better	Much better
○	○	○	○	○

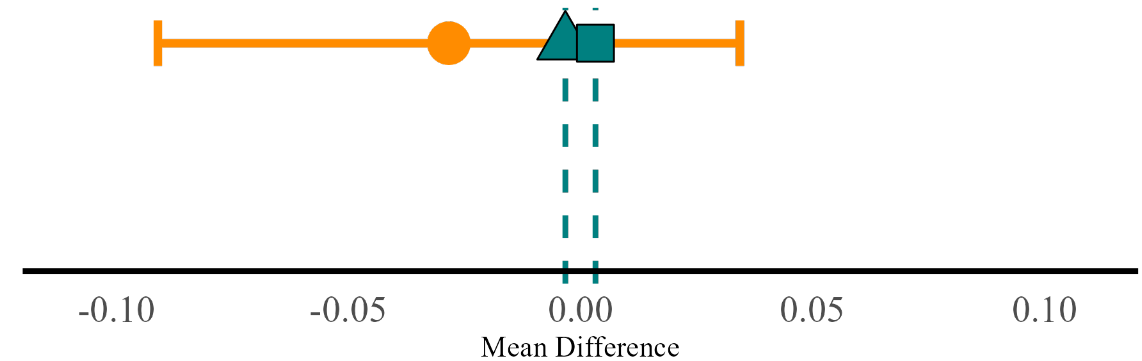
In the Qualtrics survey, a text box appeared after the preceding set of items with the yes/no prompt, "would you like to provide additional comments on HUD versus Head-Worn Display with EFVS?" If response was "yes," a text box for open-ended written feedback appeared.

Appendix B

Two One-Sided *t*-Test Figures

Figure B1

Equivalency Graph for Absolute Localizer Deviation During the Instrument Segment



In this and all subsequent equivalency graphs, ▲ = Lower Equivalence Bound; ■ = Upper Equivalence Bound; ● = Mean Difference.

Figure B2

Equivalency Graph for Absolute Localizer Deviation at the Decision Height

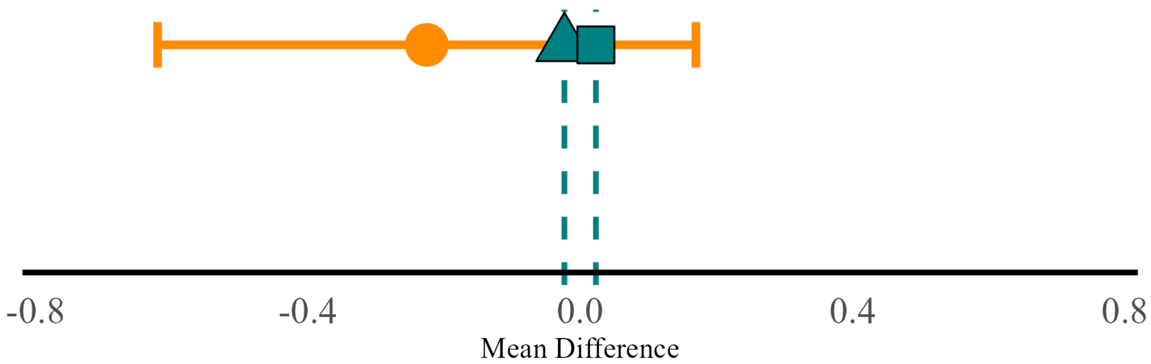


Figure B3

Equivalency Graph for Absolute Localizer Deviation at 100 ft Above the Touchdown Zone

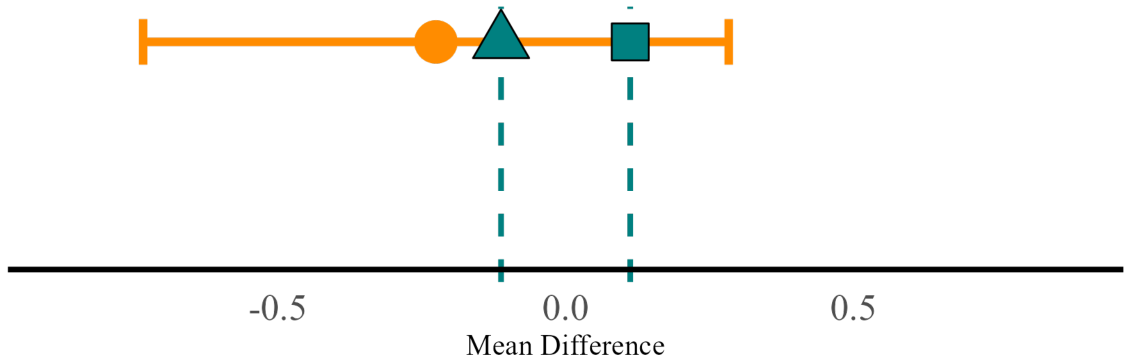


Figure B4

Equivalency Graph for Absolute Localizer Deviation at Threshold Crossing

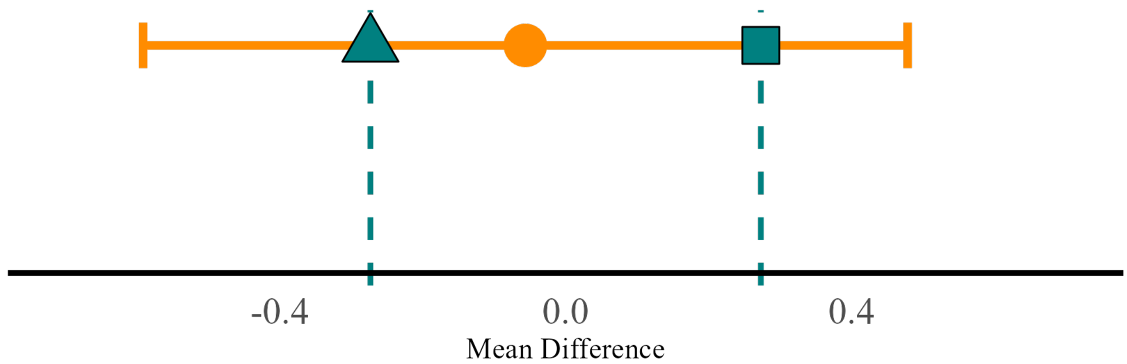


Figure B5

Equivalency Graph for Glideslope Deviation at the Decision Height

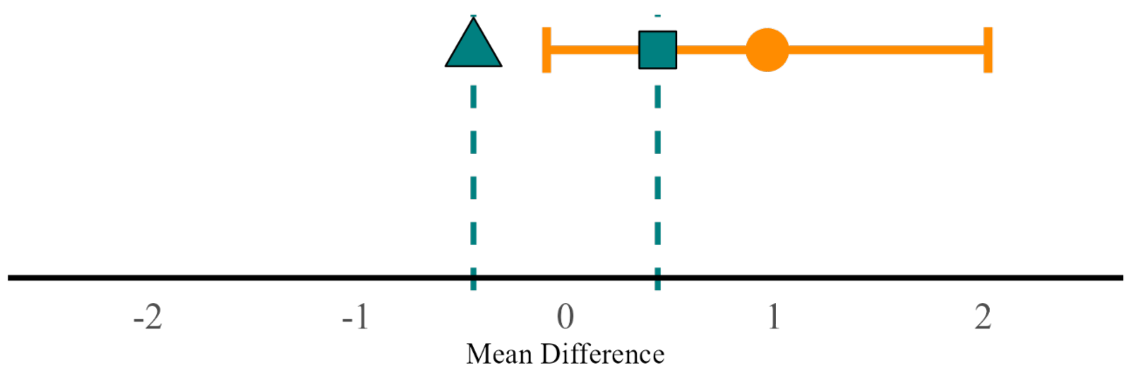


Figure B6

Equivalency Graph for Glideslope Deviation at 100 ft Above the Touchdown Zone

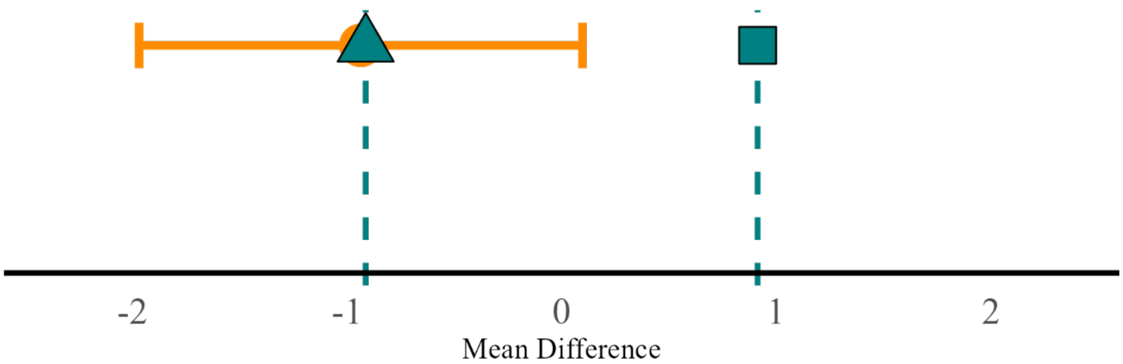


Figure B7

Equivalency Graph for RMS Airspeed Deviation During the Instrument Segment

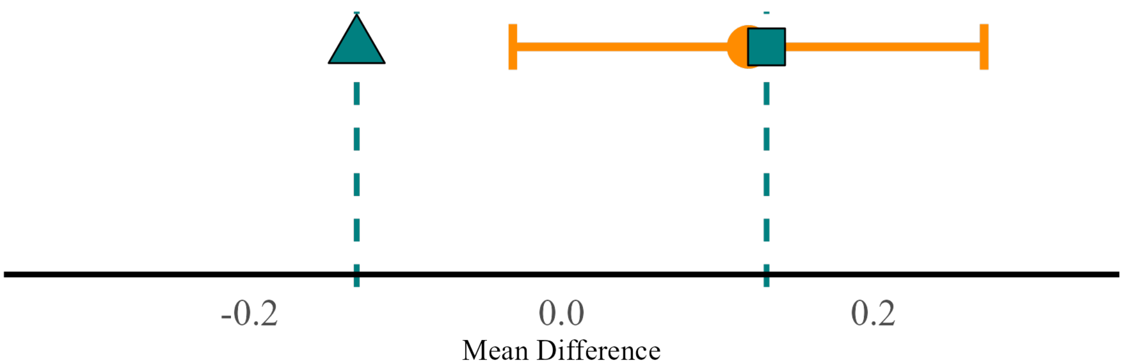


Figure B8

Equivalency Graph for Airspeed Deviation at the Decision Height

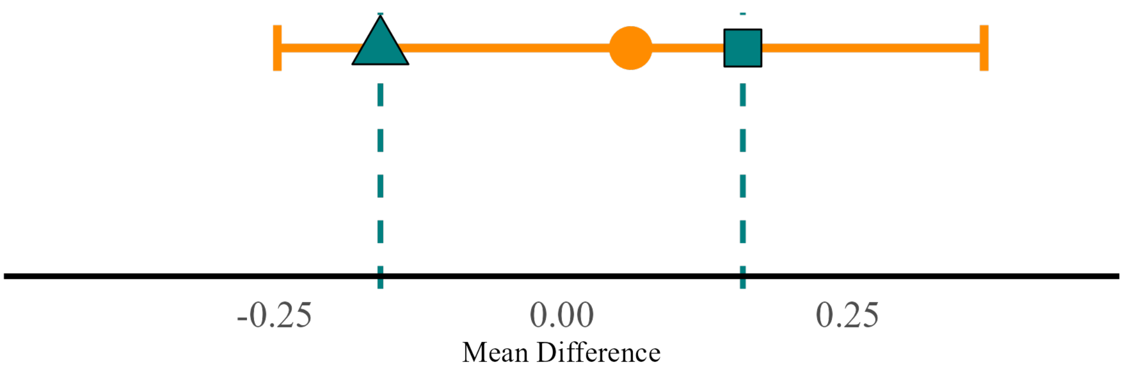


Figure B9

Equivalency Graph for Absolute Airspeed Deviation at 100 ft Above the Touchdown Zone

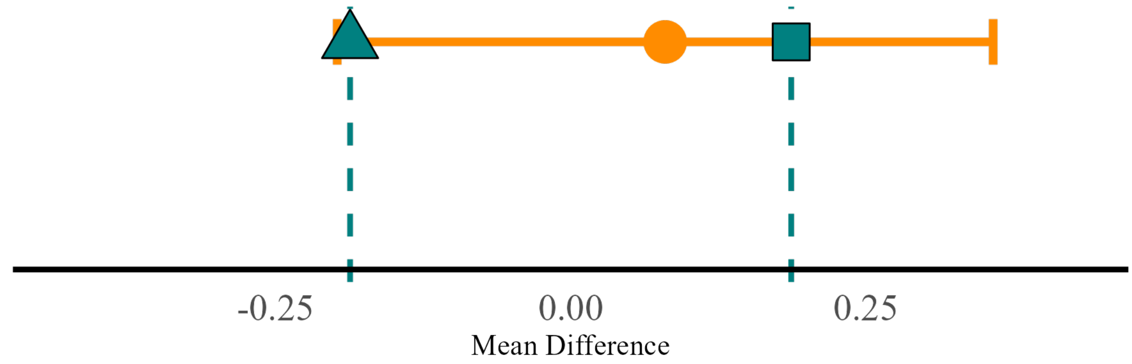


Figure B10

Equivalency Graph for Absolute Airspeed Deviation at Threshold Crossing

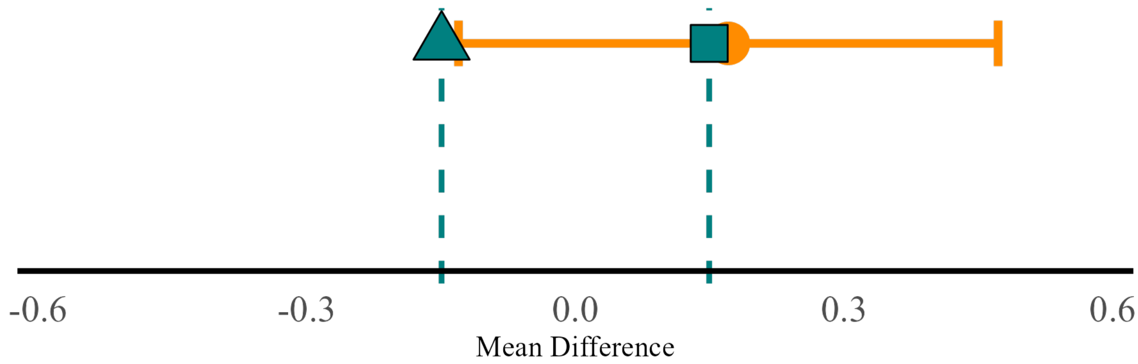


Figure B11

Equivalency Graph for Sink Rate before Touchdown

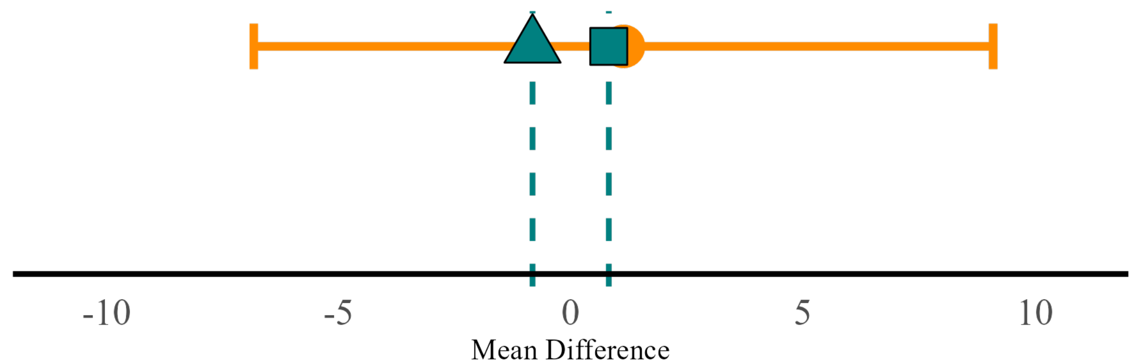


Figure B12

Equivalency Graph for Distance from Runway Threshold at Touchdown

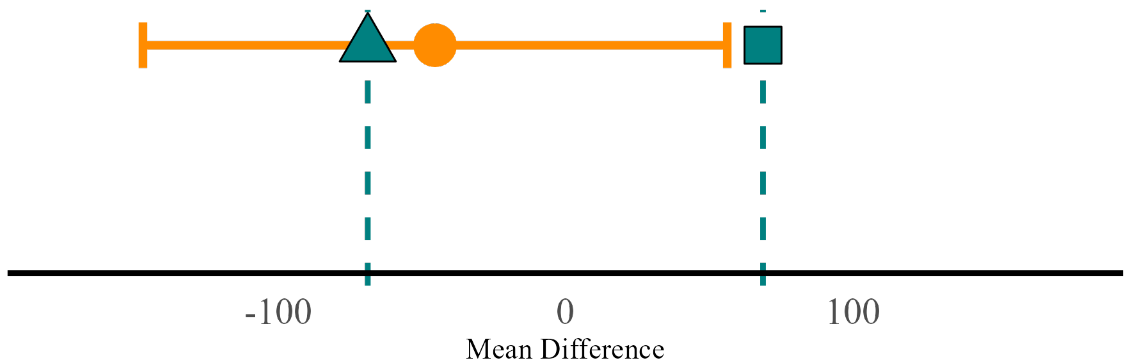


Figure B13

Equivalency Graph for Absolute Centerline Deviation at Touchdown