NASA-TP–2017–219603

Evaluation of a Head-Worn Display System as an Equivalent Head-Up Display for Low Visibility Commercial Operations

Jarvis (Trey) J. Arthur, III, Kevin J. Shelton, Lawrence J. Prinzel, III, Stephanie N. Nicholas, Steven P. Williams, Kyle E. Ellis, Denise R. Jones, Randall E. Bailey, Stephanie J. Harrison
Langley Research Center, Hampton, Virginia

James R. Barnes
Booz Allen, Hampton, Virginia

May 2017

NASA STI Program. . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI Program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI Program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collection of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.
- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.
- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI Program, see the following:

- Access the NASA STI program home page at `http://www.sti.nasa.gov`
- E-mail your question to `help@sti.nasa.gov`
- Phone the NASA STI Information Desk at 757-864-9658
- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA-TP–2017–219603

Evaluation of a Head-Worn Display System as an Equivalent Head-Up Display for Low Visibility Commercial Operations

Jarvis (Trey) J. Arthur, III, Kevin J. Shelton, Lawrence J. Prinzel, III, Stephanie N. Nicholas, Steven P. Williams, Kyle E. Ellis, Denise R. Jones, Randall E. Bailey, Stephanie J. Harrison
Langley Research Center, Hampton, Virginia

James R. Barnes
Booz Allen, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

May 2017

Acknowledgments

The authors would like to acknowledge the contributions of the technical review committee for their comments in improving this paper: Brian Baxley (chair), Kurt Swieringa, Ronald Daiker, and Michael Vincent. This work was sponsored by NASA's Aviation Safety Program (AvSP), Vehicle Systems Safety Technologies project (Mr. Paul Krasa, Program Manager). The authors gratefully acknowledge the contributions many NASA Langley technicians, engineers, programmers, management, and administrative support who are instrumental in these research projects.

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

Contents

List of Tables

List of Figures

Nomenclature

AC	Advisory Circular
ADS-B	Automatic Dependent Surveillance-Broadcast
AFFTC	Air Force Flight Test Center
AGL	Above Ground Level
ALSF2	Approach Lighting System with Sequenced Flashing Light-Model 2
ANOVA	Analysis of Variance
ATC	Air Traffic Control
CFR	Code of Federal Regulations
COTS	Commercial Off-The-Shelf
DA	Decision Altitude
DERP	Design Eye Reference Point
DH	Decision Height
DoD	Department of Defense
EFB	Electronic Flight Bag
EFVS	Enhanced Flight Vision System
EP	Evaluation Pilot
EV	Enhanced Vision
FAA	Federal Aviation Administration
FAF	Final Approach Fix
FLIR	Forward Looking InfraRed
FMS	Flight Management System
FOV	Field-Of-View
FTE	Flight Technical Error
GA	General Aviation
HAT	Height Above Threshold
HDD	Head-Down Display
HeLMR	Head Mounted Display Latency Measurement Rig

HMD	Helmet-Mounted Display
HUD	Head-Up Display
HWD	Head-Worn Display
ILS	Instrument Landing System
IMC	Instrument Meteorological Conditions
IQR	InterQuartile Range
IR	InfraRed
LaRC	Langley Research Center
LOE	Light-guide Optical Element
MANOVA	Multivariate Analysis of Variance
MASPS	Minimum Aviation System Performance Standards
MDA	Minimum Descent Altitude
NASA	National Aeronautics and Space Administration
NextGen	Next Generation Air Transportation System
ND	Navigational Display
OTW	Out-The-Window
PF	Pilot Flying
PFD	Primary Flight Display
PM	Pilot Monitoring
PTS	Practical Test Standards
RFD	Research Flight Deck
RMS	Root Mean Square
RMSE	Root Mean Square Error
RVR	Runway Visual Range
SA	Situational Awareness
SA-SWORD	Situation Awareness Subjective Workload Dominance
SART	Situational Awareness Rating Technique
SD	Standard Deviation

SMGCS	Surface Movement Guidance and Control System
SSQ	Simulation Sickness Questionnaire
SV	Synthetic Vision
TD	Traffic Diamonds
TLX	Task Load Index
VMC	Visual Meteorological Conditions
VSST	Vehicle Systems Safety Technologies

Abstract

Research, development, test, and evaluation of flight deck interface technologies is being conducted by the National Aeronautics and Space Administration (NASA) to proactively identify, develop, and mature tools, methods, and technologies for improving overall aircraft safety of new and legacy vehicles operating in the Next Generation Air Transportation System (NextGen). One specific area of research was the use of small Head-Worn Displays (HWDs) to serve as a possible equivalent to a Head-Up Display (HUD). A simulation experiment and a flight test were conducted to evaluate if the HWD can provide an equivalent level of performance to a HUD. For the simulation experiment, airline crews conducted simulated approach and landing, taxi, and departure operations during low visibility operations. In a follow-on flight test, highly experienced test pilots evaluated the same HWD during approach and surface operations. The results for both the simulation and flight tests showed that there were no statistical differences in the crews' performance in terms of approach, touchdown and takeoff; but, there are still technical hurdles to be overcome for complete display equivalence including, most notably, the end-to-end latency of the HWD system.

1 Introduction

The National Aeronautics and Space Administration (NASA) research, development, test, and evaluation (RDT&E) of flight deck interface technologies is being conducted to proactively identify, develop, and mature tools, methods, and technologies for improving overall aircraft safety of new and legacy vehicles operating in Next Generation Air Transportation System (NextGen). This work was part of the Vehicle Systems Safety Technologies (VSST) project of the Aviation Safety Program at NASA which conducts research of flight deck display technologies and concepts to potentially minimize the impact of weather and visibility on terminal area throughput and improve safety for these operations. The research objectives described in this paper were to obtain insight into the use of Head-Worn Display (HWD) systems as an equivalent display to a Head-Up Display (HUD), thus contributing towards the creation of a "better-than-visual" capability to enable NextGen equivalent visual operations [1].

NASA has conducted numerous studies evaluating the potential benefits of using HWDs, emphasizing surface operations [2–4]. HWDs are distinct from helmet-mounted displays in that they are small, light weight display devices that can be worn on the head without significant encumbrance. By coupling the HWD with a head tracker, unlimited field-of-regard can be realized and overlaid, conformal symbology can be used to improve performance and safety. This HWD system may create a "virtual HUD" concept [5–7].

In addition to overlaid symbology, imagery can be displayed on a HUD or HWD. The imagery can be generated using Synthetic Vision (SV) where the SV-viewpoint position and orientation can be defined via software. The terrain, visual flight references, and other obstacle or topographical information is contained in a database; thus, an unlimited field-of-regard is achieved since the SV scene is viewable from any virtual camera angle. (For sensor imagery, multiple tiled sensors or a "turreted" sensor would be required to realize a large field-of-regard.)

An alternative to rendering a virtual HUD on the HWD, a "Split" reference system [7] can be used in the rendering of symbology on the HWD. Non-conformal symbology (airspeed, altitude, etc.) can be display-referenced (i.e., drawn referenced to the glasses) rather than being space-stabilized. In other words, symbology such as the airspeed and altitude would be rendered fixed on the HWD screen; thus, would always be visible to the pilot regardless of where the pilot was looking. Conformal symbology (flight path marker, runway edge lines, runway extended centerline, etc.) and imagery would remain space-stabilized as with a virtual HUD concept.

This experiment explored the use of a HWD as an "equivalent display" to a HUD for the same operational credit. If this equivalence can be shown, then the HWD should receive the same operational credits as a HUD. Advisory Circular (AC) 90-106 defines criteria for an equivalent display [8].

> **Equivalent Display.** The regulations also make provision for an equivalent display. Specifically, §91.175 (m) states that the Enhanced Flight Vision System (EFVS) sensor imagery and aircraft flight symbology must be presented "...on a head-up display, or an equivalent display, so that they are clearly visible to the pilot flying in his or her normal position and line of vision and looking forward along the flight path ..."

In other words, an equivalent display must be some type of head-up presentation of the required information. A Head-Down Display (HDD) does not meet the regulatory requirement.

NASA has performed previous surface operations research using head-down displays, HUDs and HWDs [4]. This research has explored numerous operating paradigms and technologies, including the benefits of various display types versus paper charts, binocular versus monocular HWDs [9], color versus monochrome HWDs, and flight data presentation (symbology) on the display. These previous experiments with head-tracked HWD systems have been conducted in fixed-base simulators. This study is an extension of the head-up symbology research which explores HUD equivalence using a head-tracked HWD system in a motion simulator and flight test.

1.1 HUD Operational Credit

The Flight Safety Foundation identified significant safety benefits of head-up/HUD flight operations [10]. In addition to safety benefits, "operational credits" are now being derived from HUD equipage [11].

These HUD-unique credits include:

1. Fail-passive landing capability to 50-foot Decision Height (DH) and Runway Visual Range (RVR) as low as 600 feet using HUD-driven guidance through

approach, flare, landing, and roll-out (see Federal Aviation Administration (FAA) AC 120-28D [12]);

2. Low visibility takeoff minima of 300-foot RVR (as per AC 120-28D);

3. Special Authorization Category II minima on Type I Instrument Landing System (ILS) of 100-foot DH, 1,200-foot RVR (as per FAA Order 8400.13 [13]);

4. Reduction in Category II minima to 1,000-foot RVR (as per FAA Order 8400.13); and

5. Special Authorization Category I minima of 150-foot DH, 1,400-foot RVR in lieu of centerline and touchdown zone lighting (as per FAA Order 8400.13).

The HUD is the only display currently certified and approved for use as an EFVS. With an EFVS, a pilot may descend 100 feet below the published Decision Altitude (DA), DH, or Minimum Descent Altitude (MDA) from a straight-in instrument approach using an EFVS in lieu of natural vision. The EFVS operational credit (as per §91.175 (l) and (m)) explicitly expressed that the use of a HUD was an essential "characteristic and feature" of the EFVS operation.

In this experiment, the operational credits listed above and the additional operational credit afforded HUD operations with the simultaneous use of Enhanced Vision (EV) on head-up displays were explored. EV is an electronic means to provide a display of the external scene topography (the natural or man-made features of a place or region especially in a way to show their relative positions and elevation) through the use of an imaging sensor, such as a Forward Looking InfraRed (FLIR) or millimeter wave radar. Development of EV technology applications for commercial, business, and General Aviation (GA) aircraft was energized in January 2004 [8] when Title 14 of the US Code of Federal Regulations (CFR) §91.175 was amended such that operators conducting straight-in instrument approach procedures (in other than Category II or Category III operations) could operate below the published DA, DH or MDA when using an approved EFVS. An EFVS, in this application, is an integrated conformal display of EV and symbology shown on the pilot's HUD or equivalent display. In most atmospheric conditions, especially when natural visibility is reduced due to night, smoke, or haze, the EV provides a visibility improvement over natural vision, and it can be logically concluded that improvements in situation awareness (awareness of geographic position, of positioning on the runways and taxiways, and of objects, traffic, and other vehicles) are derived. This information may enable the flight crew (pilot) to more safely operate on the surface, including taxi, parking, and gate operations, or to conduct these operations in weather and visibility conditions for which this would normally be prohibited by federal regulations.

However, provisions for the use of an equivalent display were made. What constitutes an equivalent display is not explicitly defined, but by inference from CFR §91.175, the display must present "the required features and characteristics such that they are clearly visible to the pilot flying in his or her normal position and line of vision looking forward along the flight path." A critical component of EFVS performance is the integration of the "visual-like" imagery with symbology where the imagery is a display of the external scene from an imaging sensor such as a FLIR or millimeter wave radar. The primary reference for maneuvering the airplane is based on what the pilot sees through the EFVS and the HUD symbology. As such, the required external visual references must be continuously and distinctly visible and identifiable by the pilot.

1.2 HWD as an Equivalent Display

With many operational credits being provided by HUD operations, one possible avenue of HWD adoption across the NextGen fleet is by providing a "HUD-equivalent capability." The requirements for a HWD to meet a HUD-equivalent capability may be derived from FAA guidance material. For instance, under EFVS operations, these "essential features" of the HUD or equivalent display were described as follows [14]:

- The display should provide the EV image and spatially-referenced flight symbology so that they are aligned with and scaled to the external view (i.e., conformal rendering).
- The display should be located so the pilot is looking forward along the flight path (i.e., looking at and through the imagery to the out-of-the window view) to readily enable a transition from EFVS imagery to the out-the window view.
- The display should not require the pilot to scan up and down between a head down display of the image and the out-the-window view looking for primary flight reference information. This transition would otherwise be hindered by repeatedly re-focusing from one view to the other.

These requirements suggest that a HUD-equivalent display must provide conformal imagery; therefore, the HWD must use head-tracking to create a "Virtual-HUD" concept. The Virtual-HUD concept is not new. The F-35 and others are working toward making the HWD a HUD replacement [6]. However, achieving this capability for business and commercial aircraft is a formidable challenge [15].

The goal of this research is to evaluate a HWD system as an equivalent system to a standard flight HUD. If this equivalence can be shown, then the unique capabilities of the HWD — for example, unlimited field-of-regard head-up operations for piloted surface operations [4] — can be realized. The design challenge (and certification challenge) is to create this equivalent capability without increasing pilot workload, encumbrance, or obscuration of their normal vision. [16]

2 Simulation Experiment

A human-in-the-loop experiment was conducted to collect data to help quantify the characteristics that define an equivalent display. A secondary objective included the influence of traffic symbology on the HWD. The traffic symbology consisted of traffic icons which denoted the position of other aircraft based on their reported ADS-B position. The traffic icons were presented head-up in the interest of providing an intuitive traffic information display and specifically, during the taxi phase of the departure scenarios for the prevention of runway incursions.

Off-nominal scenarios introduce unexpected events to flight crews with the purpose of uncovering possible design issues in the system. Foyle and Hooey [17] describe the benefits for off-nominal scenario development. For this experiment, two off-nominal scenarios were conducted to gather data on the crew's reaction to a non-normal event and evaluate the crew's performance when using a HWD compared to a HUD.

2.1 Simulation Facility

This experiment was conducted in the NASA Langley Research Center (LaRC) Research Flight Deck (RFD) simulator on a motion base platform (Fig. 1). The RFD was configured to mimic the instrument panel of current state-of-the-art commercial transport aircraft, with four 10.5" vertical by 13.25" horizontal, 1280x1024 pixel resolution, color displays tiled across the instrument panel. Also, the RFD included a mode control panel, Flight Management System (FMS), control display units, and hydraulic-actuated side-stick control inceptors. A collimated Out-The-Window (OTW) scene provided approximately 200° horizontal by 40° vertical Field-Of-View (FOV) at 26 pixels per degree. Electronic charts and an aircraft moving map were provided on an Electronic Flight Bag (EFB). For this experiment, the EV system was a simulated FLIR camera fixed to the aircraft. The FLIR sensor aperture was placed 5.25 feet below the pilot Design Eye Reference Point (DERP), 1.5 feet to the right, and 6.5 feet forward, simulating an aircraft "chin" installation.

Figure 1. The Research Flight Deck simulator at NASA Langley Research Center.

2.1.1 Head-Up Displays

The HUD used in this experiment was a Rockwell Collins HGS-6700 installed for the left seat operator of the flight deck. The field-of-view of the HUD was 46° H by 34.5° V.

The HWD is shown in Fig. 2. A prototype head tracker was used to provide head orientation and was mounted on the left side of a pair of Lumus DK-32 glasses. The head tracker was a hybrid-inertial tracker with image processing to correct for inertial drift. The head tracker image processing used infrared, passive barcodes located at known locations in the flight deck to provide accurate head tracking. The Lumus glasses specifications are shown in Table 1 along with the HUD specifications for comparison. The Lumus eye-wear is a see-through, full color binocular display which utilizes patented Light-guide Optical Element (LOE) technology to generate an image that appears at "practical" infinity. For this experiment, only monochrome green symbology and imagery were displayed on the HWD so as to not introduce a confounding variable when comparing to the monochrome HUD.

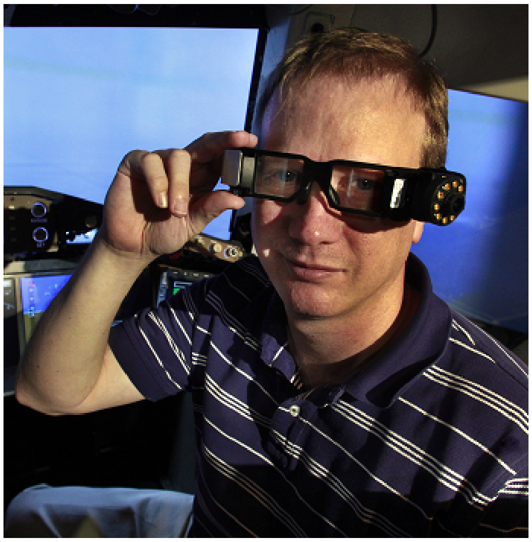

Figure 2. The HWD system used in the experiment.

Table 1. Display specifications.

	HWD	HUD
Resolution	1280 (H) x 720 (V)	1400 (H) x 1050 (V)
Field-of-View	35° H x 20° V	46° H x 34.5° V
Brightness	1000 fL	4000 fL
Image Focal Plane	Infinity	Infinity
Weight	0.20 kg	14 kg (combiner + overhead)

2.1.2 HWD/HUD Symbology

During simulated flight, the HWD symbology was designed to replicate typical HUD symbology for a commercial transport including a flare cue and runway outline, and other requisite EFVS symbology including a flight path angle reference cue, conformal guidance cue, flight path marker, and raw data (Fig. 3). At approximately 100 feet Above Ground Level (AGL), a flare cue would appear and provide guidance to the pilot for flaring the airplane. The flare cue was displayed based on a function of radar altitude.

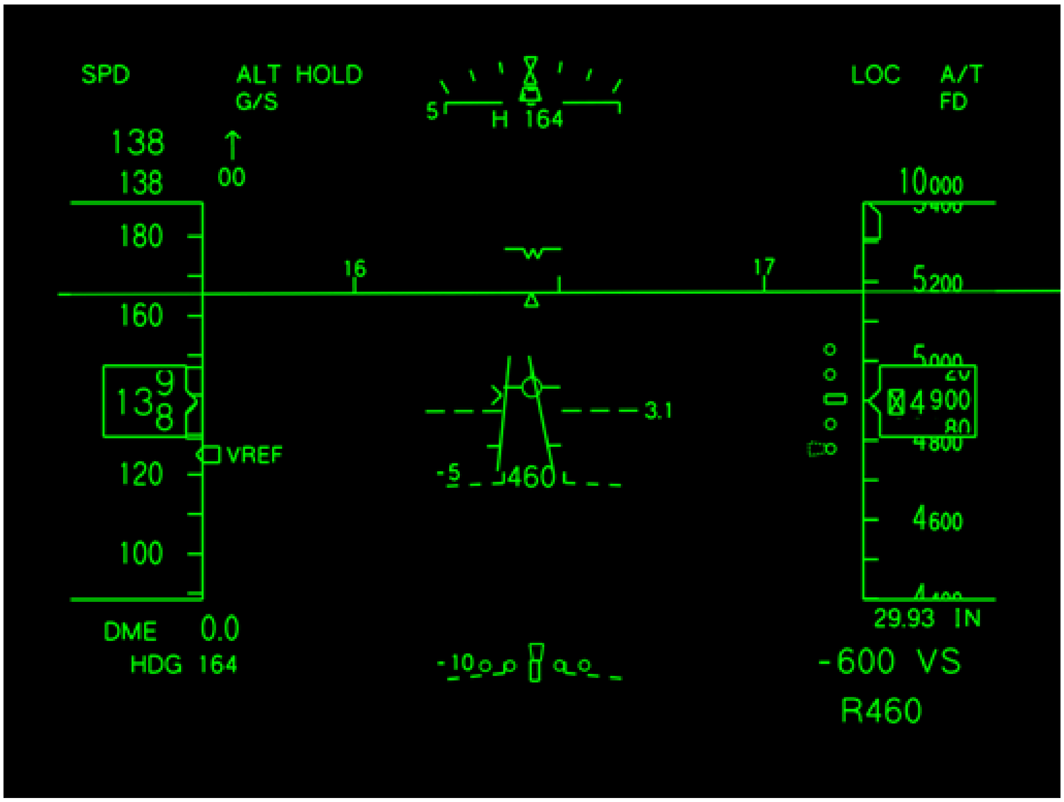

Figure 3. The approach symbology set for the HUD and HWD display concepts.

When the nose wheel was on the ground and the ground speed was less than 80 knots, the symbology would automatically transition to the taxi symbology (Fig. 4). The surface symbology set was developed at NASA from previous research [3]. These symbology sets were displayed on both the HUD and HWD. The taxi symbology set consisted of ground speed, heading, current taxiway the aircraft was on and the next taxiway on the cleared route. Above the next taxiway text, either a left or a right arrow was rendered to denote the direction of the next cleared taxiway turn. Near the bottom of the display was a raw data indicator showing linear deviation from the taxiway centerline. The deviation was scaled to represent ±25 feet. Traffic icons were rendered on head-up displays in a perspective format as unfilled "diamonds." Traffic diamonds were only displayed if the traffic position was within 1 nautical mile of ownship. This 1 nautical mile filter was used to reduce symbology clutter

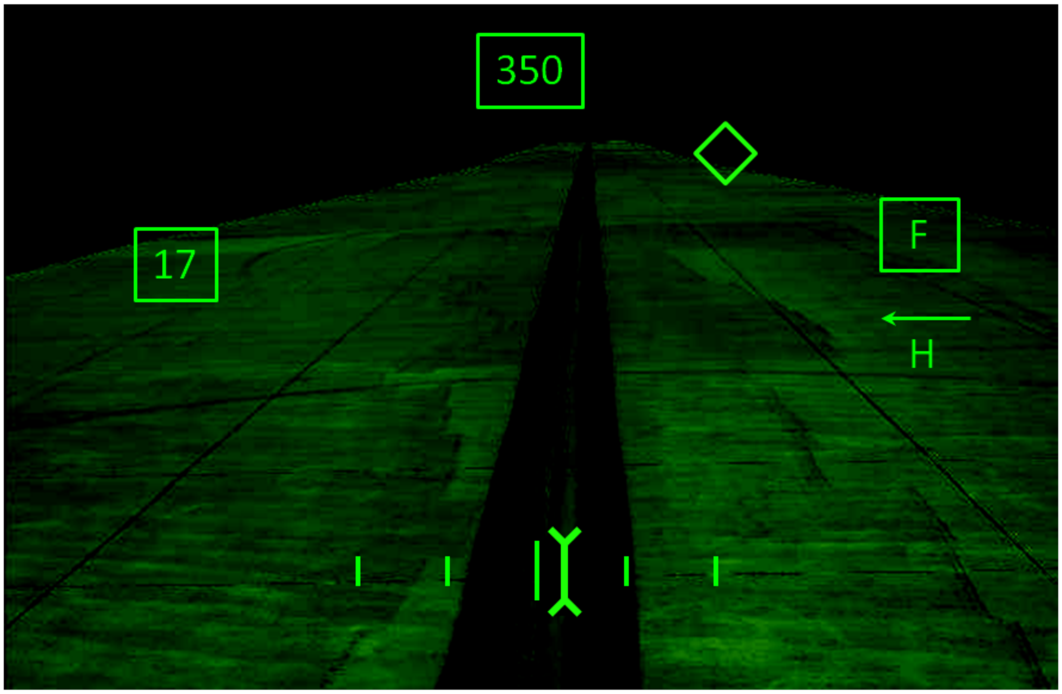

Figure 4. The surface symbology set for the HUD and HWD display concepts.

by only displaying traffic close to ownship. The traffic diamonds were rendered as a 30-foot diameter 3-dimensional object; therefore, the diamonds would appear larger as the distance between the traffic and ownship drew closer.

For departures, a typical takeoff symbology set was used, identical for both the HUD and HWD. The takeoff symbology set was very similar to the flight symbology with the addition of the ground localizer line to aid in centerline tracking during takeoff roll.

2.1.3 Head-Down Displays

The head-down displays consisted of 4 panels. The Pilot Flying (PF) displays (Fig. 5) consisted of a Primary Flight Display (PFD) on the left panel and a Navigational Display (ND) on the right. The Pilot Monitoring (PM) displays (Fig. 6) consisted of a ND with EV (i.e., FLIR) display on the PMs left panel and a PFD on the right. The EV display for the PM was present or absent depending upon the experimental condition. For scenarios where EV was displayed on the HUD or HWD, the PM would have a "repeater" EV head-down; otherwise, the display area was blank. The head-down EV repeater did not have overlaying symbology; it was raw imagery. Two EFBs were utilized (PF side and PM side) for various functions, including charts, checklists and displaying an airport surface map.

Figure 5. The pilot flying head-down displays.

Figure 6. The pilot monitoring head-down displays. The EV repeater is shown on the pilot monitoring navigation display in the upper left corner.

2.2 Enhanced Vision Simulation

The EV was simulated as a combined short-wave, mid-wave ($\sim$ 1.0 to 5.0 micron) FLIR sensor. The simulated camera was aligned with the HUD, so any image shift between the FLIR displayed on the HUD and the OTW was due only to installation parallax as described in Section 2.1. The image shift (i.e., error) due to camera parallax was half of the maximum error allowable for an EFVS in accordance with RTCA DO-315 [18], equating to a 2.5 milliradians image offset of a point located at a distance of 2000 feet.

2.3 Evaluation Pilots

Twelve commercial flight crews from various US airlines participated in the experiment. The Evaluation Pilots (EPs) were paired based upon their current employer to minimize inter-crew differences in Standard Operating Procedures and Crew Resource Management procedures. The Captain was the PF and sat in the left seat. Only the Captain had a HUD and wore the HWD for the experiment. As eyeglasses are not compatible with the HWD used in this experiment, pilots who required glasses for flying were excluded from participation. Accommodation for vision correction is a known issue [1] with HWD systems; however, it was not addressed in this experiment. The First Officer was the PM for the duration of the experiment; thus, crew members did not switch roles during the experiment.

All pilots held an Airline Transport Pilot rating. Captains had an average of 33 years experience with an average of 1800 hours of HUD experience, though 6 Captains had less than 1000 hours of HUD experience. First Officers had an average of 32 years experience. Of the 24 EPs who participated in the study, 4 pilots had over 500 hours experience with an EV system.

2.4 Evaluation Pilot Training

The EPs were given a 30-minute classroom briefing to explain the display concepts and the evaluation tasks for the experiment. After the briefing, a 1-hour training session was conducted to familiarize the EPs with the RFD simulator. Following this training, 2 hours of data collection was conducted for the approach runs followed by 2 hours of data collection for the departure runs. At the end of the day, a post-test interview was conducted to solicit the crew's comments on the experiment. The total duty time for an evaluation crew was approximately 8 hours.

2.5 Eye Tracking System

Eye and head tracking data was collected for the PF and PM using the Smart Eye® eye-head tracking system installed in the simulator. The HWD prevented reliable eye tracking (see Appendix A for the eye tracking analysis); however, head tracking with the oculometer system was not affected by the HWD.

2.6 Latency

Measuring total system latency is an important consideration in HWD applications [1, 19] because scene mismatch caused by latency effects can lead to motion/simulation sickness [20]. At this time, there is no standard for acceptable latency for HWDs though prior tests suggest that the latency should be less than 20 milliseconds [19]. All commercial or custom head mounted display systems that track the users head for the purpose of virtual or augmented reality applications encounter some latency. A basic HWD with head tracking system is comprised of 1) a near-to-eye display, 2) the head tracking system, 3) one or more symbology or image sources, 4) and the display/image processor [15]. Each element and the communication between them contribute a portion to the total latency.

For this experiment, the HWD system latency was measured using the Head Mounted Display Latency Measurement Rig (HeLMR) [19]. The HeLMR apparatus measures latency by slewing a HWD back and forth at a precise angular rate. By knowing this angular rate and measuring the degree offset between the camera image and the real world, the total system latency of the HWD system can be calculated.

A software error was discovered post-test in the implementation of the head tracker software. To isolate the head motion, the simulator attitude must be "canceled-out" of the head tracking calculations. The head tracker software was not receiving simulator attitude data. The result of this software error would appear as additional latency to the pilot as the tracker would correct via image processing without the added quickening of the platform data. Because of the nature of the test (straight-in approaches, no winds, no turbulence), there was little movement of the simulator platform and thus, this software error had little impact on the head tracker performance.

2.7 Methodology

Approaches and departures were simulated at the Memphis International Airport (FAA identifier: KMEM). The experiment data runs were grouped by Display Concept (HUD/HWD) within an operation block (Approach/Departure). The experiment was grouped by Display Concept to minimize the need for EPs donning and doffing the HWD between runs. For departures, half of the data runs contained Display Features (an EV image plus traffic diamond symbology) which were evenly distributed across the the head-up display type (either HUD, HWD-Virtual and HWD-Split). Table 2 shows the nominal run matrix for each crew. Table 3 shows the off-nominal runs spread across the 12 crews. After the last nominal approach and departure run, each crew experienced an off-nominal run which varied the display (HUD/HWD) across subjects.

During scenarios using the HWD, the HUD was stowed. EPs wore the HWD for approximately 45 minutes in duration within each HWD data collection block.

Table 2. Experiment matrix for nominal runs for each crew.

	HUD	**HWD-Virtual**	**HWD-Split**	**Display Features**
Approach	2	2	2	N/A
Departure	1	1	1	On (EV + traffic symbol)
	1	1	1	Off (no EV + no traffic symbol)

2.8 Evaluation Task

All expected procedures and appropriate protocols were briefed prior to the test for each crew, and training was provided to familiarize crews with operational procedures prior to data collection. The EFVS procedures used for this study were built

Table 3. Experiment matrix for off-nominal runs spread across the 12 crews.

	Approach Go-Around	**Departure Engine-Out**
HUD	4	4
HWD-Virtual	4	4
HWD-Split	4	4

around common practice in current EFVS operations and FAA requirements (CFR 91.175 (l) [8]).

The simulated weather conditions were 1000 ft RVR for the approaches and 300 ft RVR for the departures and surface operations. Both approach and departure scenarios simulated daytime conditions with no winds or turbulence. Approaches were manually flown by the PF with auto-throttles engaged. For approaches, the EV (i.e., FLIR) simulation was calibrated to show topographical objects within a range of approximately 2000 feet and light sources within a range of approximately 2400 feet. For departures, the FLIR was calibrated to show topographical objects within a range of approximately 600 feet and light sources within a range of approximately 1000 feet. A terrain database was developed for the KMEM area, which included all airport taxiways, runways, Surface Movement Guidance and Control System (SMGCS) visual aids and markings, prominent airport buildings, obstructions, signs, and airport terrain and cultural features. All approaches were straight-in to runways equipped with Approach Lighting System with Sequenced Flashing Light-Model 2 (ALSF2) runway approach lights. The simulator also used the appropriate database information to emulate the accurate location and appropriate radio frequencies of navigation aids, to coincide with published charts.

For approach scenarios, crews were briefed on their starting position (1000 ft AGL on final) and the approach runway. Crews were then briefed on the weather conditions and allowed to conduct any briefings or checklist before the data trial began. After the scenario began, crews were given a landing clearance with an expected high speed turn-off (if feasible). Once the aircraft was clear of the runway, the approach scenario ended.

For departure scenarios, the scenarios started at various points around the airport in the non-movement area. Crews were briefed on their starting position on the airport. At the start of the data collection trial, crews contacted the ground controller and received taxi instructions to the departure runway. If the PM failed to correctly read-back the proper taxi instructions, the taxi clearance was read to the crew until a correct read-back occurred. Upon reaching the runway holding position, crews were instructed by the ground controller to switch to the tower frequency at which time they would receive their departure clearance. Departure scenarios ended at an approximate altitude of 1000 ft AGL after takeoff.

Post-run questionnaires were given to both EPs after each scenario, and consisted of 1) a 3-part Situational Awareness Rating Technique (SART) [21] form, 2) an Air Force Flight Test Center (AFFTC) 7-point workload scale [22], 3) a NASA Task Load Index (TLX) workload rating [23], and 4) 10 questions addressing HWD equivalence, crew interaction, operational effectiveness, and EV usability. After each Display Condition group was completed, a Simulation Sickness Questionnaire (SSQ) [24] was administered. These questionnaires were given immediately after the end of each data trial.

2.8.1 Display Symbology for Approach Scenarios

The approach experiment matrix consisted of 1 independent variable: the Display Condition. The Display Condition consisted of 3 display types: 1) HUD, 2) HWD rendering a Virtual HUD (HWD-Virtual), and 3) HWD with split-referenced symbology (HWD-Split). Each display concept was replicated twice for each crew.

The HUD display condition was a typical HUD with EV imagery. The HWD-Virtual HUD Display Condition replicated the HUD display condition by utilizing the head-track HWD system such that HWD symbology and imagery overlaid the same positions as the HUD when the PF looked where an actual HUD would be (i.e., a “Virtual HUD”). The symbology and imagery was drawn using earth-reference and aircraft-reference stabilization (see Fig. 3). The HUD was stowed for all HWD Display Conditions.

The HWD-Split Display Condition consisted of the same symbology as the HWD-Virtual HUD display concept but included a mix of screen-referenced symbology and conformal symbology. In this condition, non-conformal symbology was drawn in the screen-reference space. For flight symbology, the non-conformal symbology consisted of the airspeed and altitude tapes, the roll scale, mode annunciations, heading indicator, and localizer and glideslope scales. For the surface symbology, the non-conformal symbology consisted of the boxed ground speed, the boxed heading indicator, the taxiway clearance text and the taxiway centerline deviation scale. Using screen-references, these non-conformal symbology elements were always drawn in the same HWD display location and the pilot’s head motion did not affect the rendering of these symbols. The flight path marker, pitch ladder, flight path angle reference cue, guidance symbology, and the EV imagery must remain conformal, so these conformal symbology elements were space-stabilized as in the HWD “Virtual HUD” condition.

After the completion of all of the 6 nominal approach runs, an off-nominal approach was conducted. This unannounced off-nominal data trial consisted of an Air Traffic Control (ATC) call to the crew to execute a go-around at an altitude of 100 feet AGL. Each of the off-nominal approach conditions were spread across the 3 Display Conditions with 12 crews; thus, each of the display types were replicated 4 times.

2.8.2 Display Symbology for Departure Scenarios

The experimental matrix for the departure runs consisted of 2 phases: 1) taxi to runway, and 2) takeoff and climb to an altitude of 1000 ft AGL. The independent variables for the taxi portion of the scenario consisted of 3 Display Conditions (HUD, HWD-Virtual, and HWD-Split) and 2 Display Features: 1) Baseline - no EV image and no Traffic Diamonds (TD); and, 2) EV+TD - surface symbology with a conformal EV image and TD symbology (see Fig. 4).

During taxi operations, the research ground symbology was used. Once the aircraft reached the runway, the surface symbology set transitioned to the takeoff symbology, and at positive rate of climb, the takeoff symbology transitioned to the airborne symbology set. The takeoff symbology was the same as the flight symbology with 2 exceptions: 1) the addition of ground localizer line; 2) the flight path marker was rendered differently. The flight path marker was caged vertically at the -2° pitch and while caged, was drawn with additional "legs" on the bottom half of the marker (see Fig. 7). The ground localizer line was symbology that consisted of a vertical line which was driven by the localizer to aid pilots in tracking the runway centerline on takeoff. This ground localizer line and the flight path marker legs were removed once the aircraft was airborne.

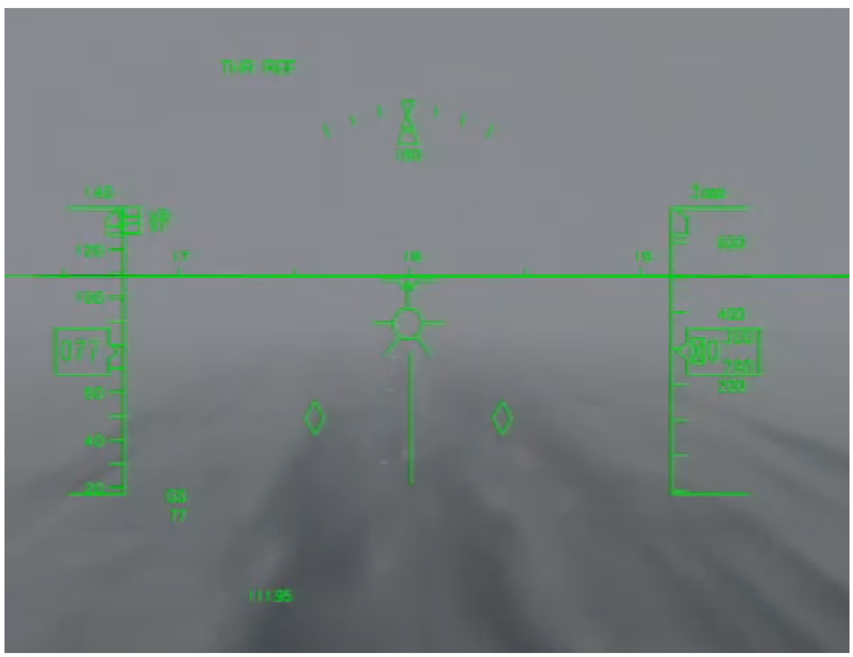

Figure 7. The head-up symbology on takeoff.

In addition to the nominal departure runs, an additional off-nominal departure run was conducted. The off-nominal departure event was an engine-out which occurred at 100 knots during the takeoff roll. The EPs were unaware of the impending engine-out event; the run was briefed the same as the nominal runs. For the departure off-nominal events, the augmented reality symbology was off (i.e., no EV or traffic diamonds). The off-nominal condition was spread across 12 crews with 2 Display Conditions (HUD or HWD-Virtual); thus, each Display Condition was replicated 6 times.

3 Simulation Results

Quantitative (i.e., aircraft state, navigational, systems interaction, eye tracking) data as well as qualitative (i.e., questionnaires, workload and situation awareness metrics, pilot opinion) responses were recorded and used in a detailed data analysis to determine if a HWD system had equivalent performance to a HUD. The legend for the box and whisker plot data which follows is shown in Fig. 8. Data that are greater than 1.5 times the InterQuartile Range (IQR) from the 25% quartile (Q1) or the 75% quartile (Q3) are considered outliers. N=125 would indicate the box plot represents 125 data points. The asterisk symbol is used to denote outliers for box plots with a small number of data points ($N <= 10$). For box plots representing a large amount of data points ($N > 10$) and have many outliers, a dot symbol is used to denote the outlier values. Some outliers are not shown for figure clarity (i.e., if included, the preponderance of data can make the figures unreadable). In those cases, the outliers were examined for relevancy and the analysis does not indicate any trends or issues in the data and their omission does not change the conclusions.

3.1 Quantitative Results

Descriptive statistics are located in Appendix B. Values are considered statistically significant for $p < 0.05$.

3.1.1 Flight Technical Error (FTE) on Approach

The quantitative flight path performance dependent measures reported were: Root Mean Square Error (RMSE) (Glideslope, Localizer, Sink Rate deviation) and Max Values (Glideslope, Localizer, Sink Rate deviation). For computing sink rate deviation, a nominal sink rate of 11.9 feet per second was used which was derived from a 3° glideslope with a ground speed of 135 knots with no winds.

An Analysis of Variance (ANOVA) was conducted on FTE for Localizer dot error and Glideslope dot error tracking performance from an altitude of 1000 feet AGL to 50 feet. The results found no significant effects for Localizer, $F(2, 69) = 0.341, p = 0.712$; or glideslope, $F(2, 69) = 0.409, p = 0.666$. The localizer, glideslope and sink rate data, collapsed across all pilots, are shown in Figs. 9, 10 and 11. For these data, outliers were removed. Table B1 in Appendix B shows the approach FTE descriptive statistics.

3.1.2 FTE on Instrument and Visual Segments

For comparison to previous research results [25], the approach analysis was further divided into 2 altitude segments; the Instrument Segment and the Visual Segment. The Instrument Segment was defined from 1000 to 200 feet Height Above Threshold (HAT) and the Visual Segment was defined from 200 to 50 feet HAT. These data are plotted in Fig. 12 (Localizer) and Fig. 13 (Glideslope) showing the data between the Instrument segment (1000 ft to 200 ft on the left-side of the figure) and the Visual Segment (200 ft to 50 ft on the right-side of the figure).

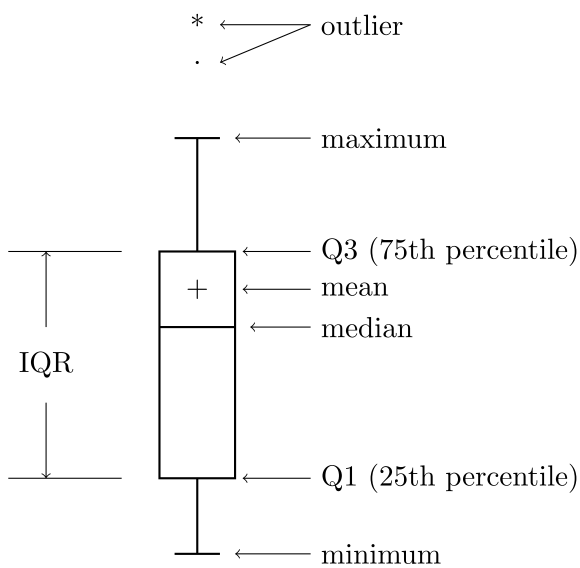

Figure 8. Legend for box plots (not to scale). Data that are greater than 1.5 times the IQR from the 25% quartile (Q1) or the 75% quartile (Q3) are considered outliers.

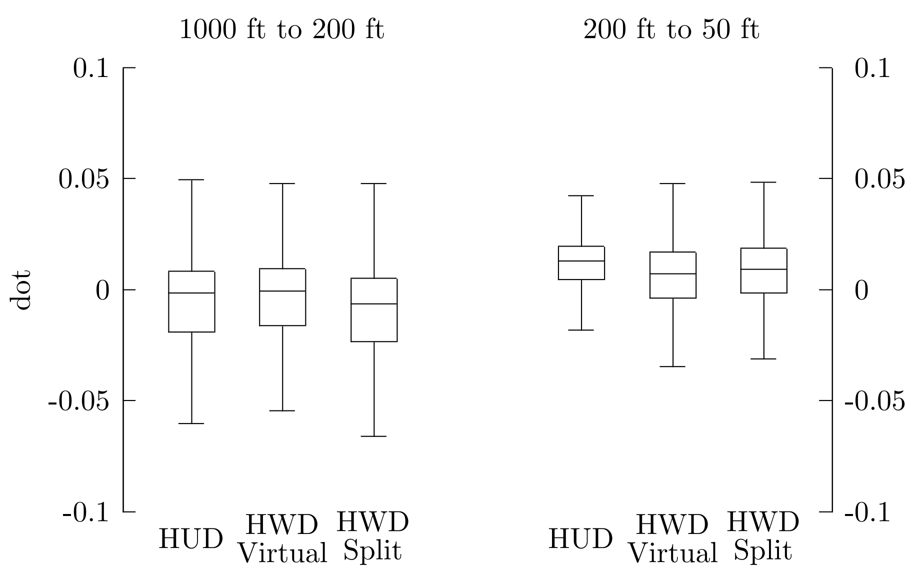

Figure 9. Localizer dot on approach (outliers removed).

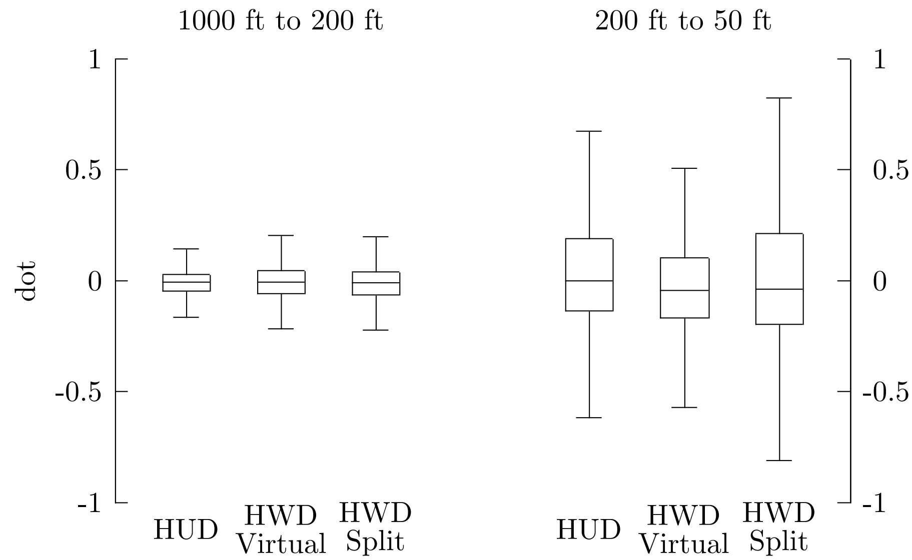

Figure 10. Glideslope dot on approach (outliers removed).

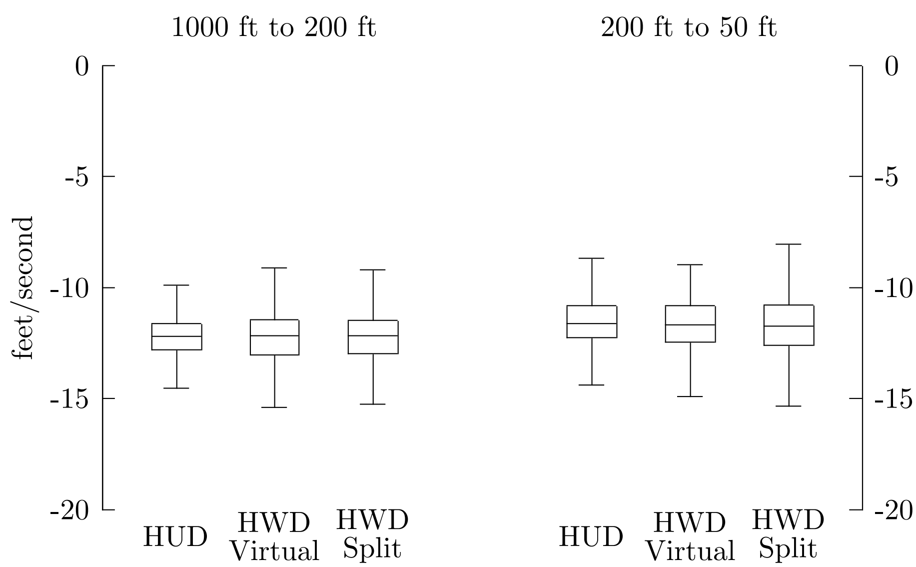

Figure 11. Sink rate on approach (outliers removed).

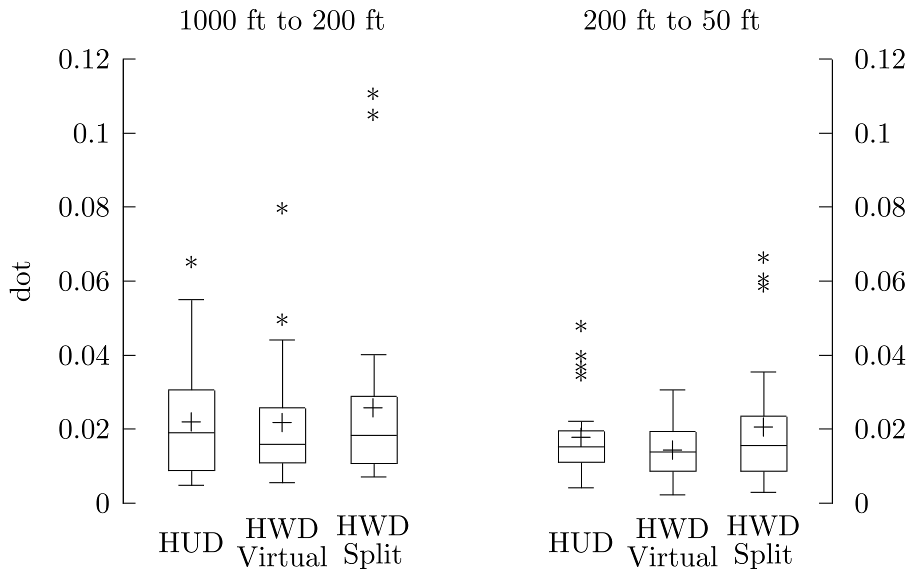

Figure 12. Localizer dot error (RMSE) on approach.

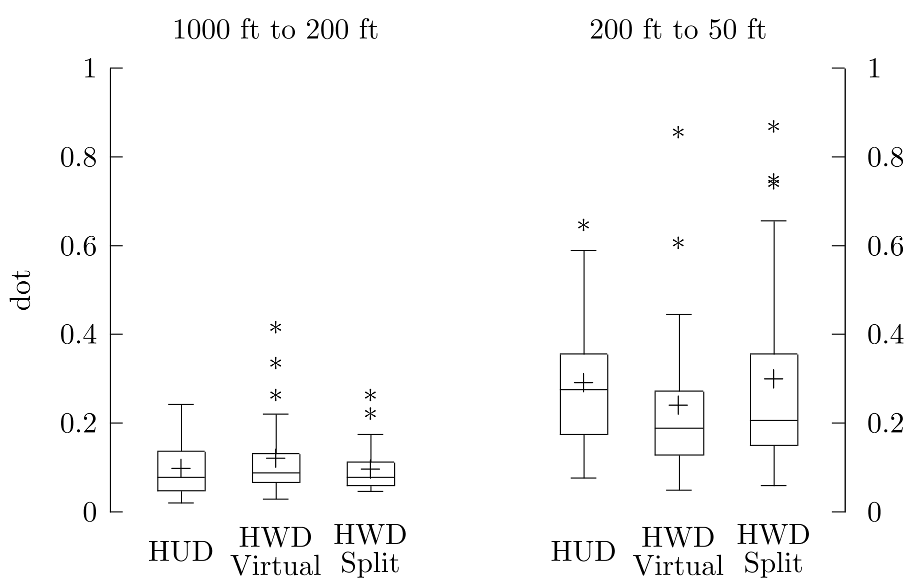

Figure 13. Glideslope dot error (RMSE) on approach.

An ANOVA was conducted on the RMSE dependent measures for the Instrument Segment. No significant results were found between the display concepts for Glideslope tracking, $F(2,81) = 0.284, p = 0.754$; Localizer tracking, $F(2,82) = 0.762, p = 0.470$; or sink rate deviation, $F(2,82) = 0.905, p = 0.409$. Table B2 in Appendix B presents the descriptive statistics for RMSE dependent measures.

Statistical analyses also failed to evince significant results for the Instrument Segment for maximum values of maximum localizer deviation (Fig. 14), $F(2,82) = 0.764, p = 0.469$; maximum glideslope deviation (Fig. 15), $F(2,82) = 0.279, p = 0.757$; or maximum sink rate (Fig. 16), $F(2,82) = 1.250, p = 0.292$. Table B3 in Appendix B presents the descriptive statistics for the maximum values for the approach dependent measures.

The same dependent measures were analyzed via ANOVA to examine the effect of the display concepts for the Visual Segment. The statistical results showed that the display concepts were not significantly different from each other in terms of the dependent measures of RMSE localizer, $F(2,69) = 1.358, p = 0.264$; RMSE glideslope, $F(2,69) = 0.674, p = 0.513$; or RMSE sink rate $F(2,69) = 0.707, p = 0.497$.

The ANOVA statistics for maximum values for these dependent measures also suggest equivalence across displays for maximum localizer deviation, $F(2,69) = 1.984, p = 0.145$; maximum glideslope deviation, $F(2,69) = 0.847, p = 0.433$; or the maximum sink rate $F(2,69) = 0.541, p = 0.585$. Tables B4 and B5 in Appendix B present the descriptive statistics for RMSE and maximum values for the visual segment, respectively.

3.1.3 Threshold Crossing Height Performance

The 3 display concepts were compared for quantitative performance at threshold crossing height (approximately 50 feet HAT). No significant differences were found for lateral deviation, $F(2,69) = 0.986, p = 0.378$; vertical deviation $F(2,69) = 0.064, p = 0.938$; or sink rate, $F(2,69) = 0.224, p = 0.800$. Figures 17, 18, and 19 show box plots of the deviation from the ideal 3° glide path at the 100 foot altitude point and the 50 foot altitude point on approach. Table B6 in Appendix B presents the descriptive statistics for threshold crossing height performance.

3.1.4 Touchdown Performance

The EPs were briefed to aim at a point on the approach 1000 feet down from the runway threshold. The crews were instructed to land within the touchdown zone, no closer than 200 feet from the threshold and no longer than 2700 feet from the threshold. The crews were also instructed to land as close as possible to the runway centerline.

These landing criteria were derived from performance standards required by Category III auto-land systems [12, 26]. Before data collection, crews were trained to land within the standard. The 3 performance categories used in this paper are 1) lateral distance from the centerline, 2) longitudinal distance from the threshold, and 3) sink rate at touchdown. Each of these performance categories have 3 levels:

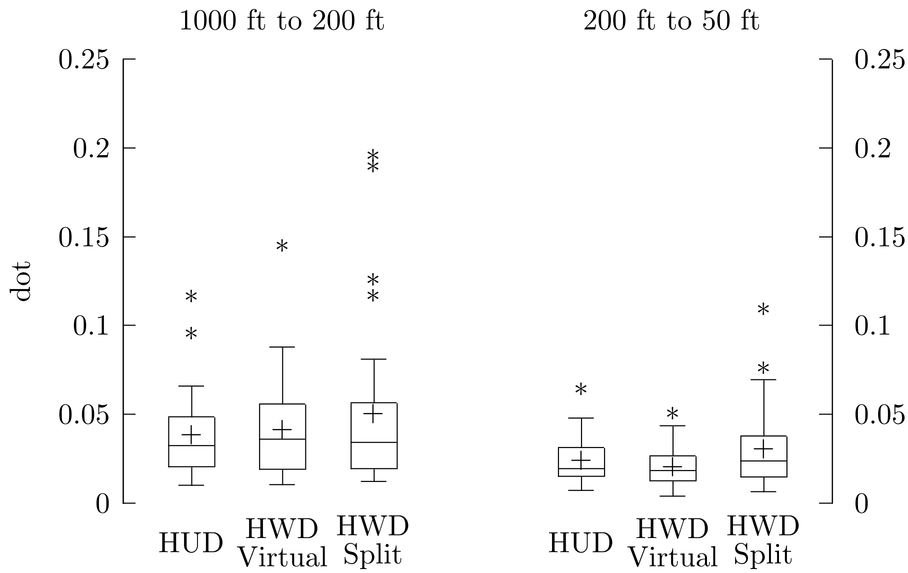

Figure 14. Maximum (absolute value) localizer dot error on approach.

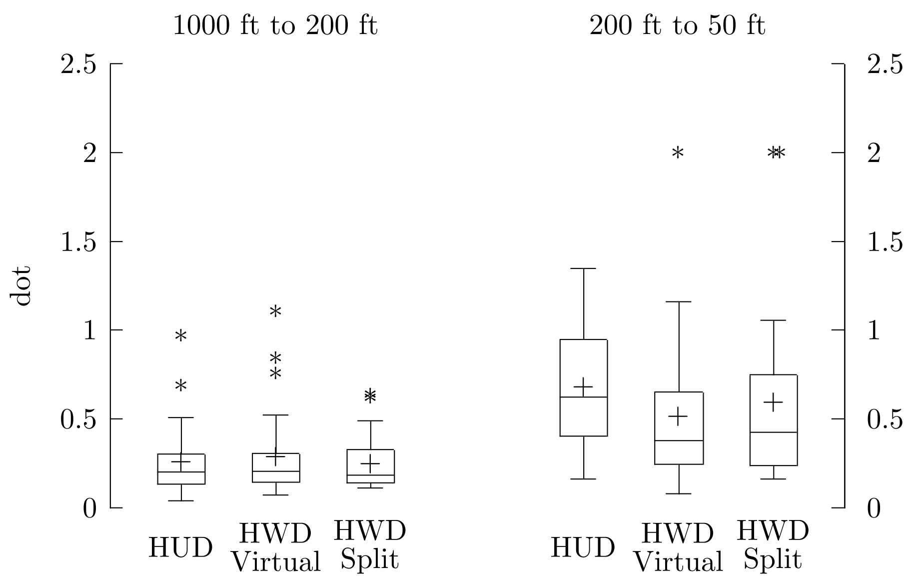

Figure 15. Maximum (absolute value) glideslope dot error on approach.

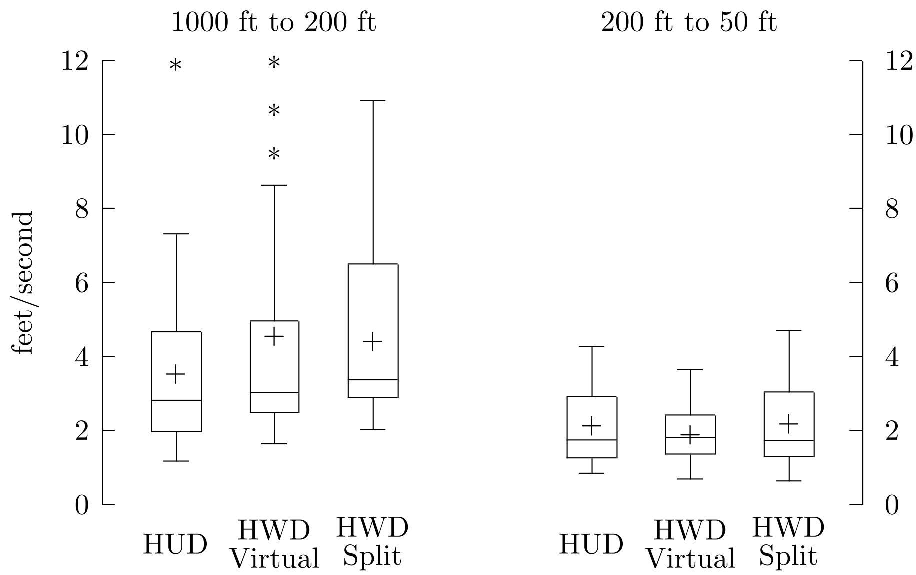

Figure 16. Maximum sink rate deviation on approach.

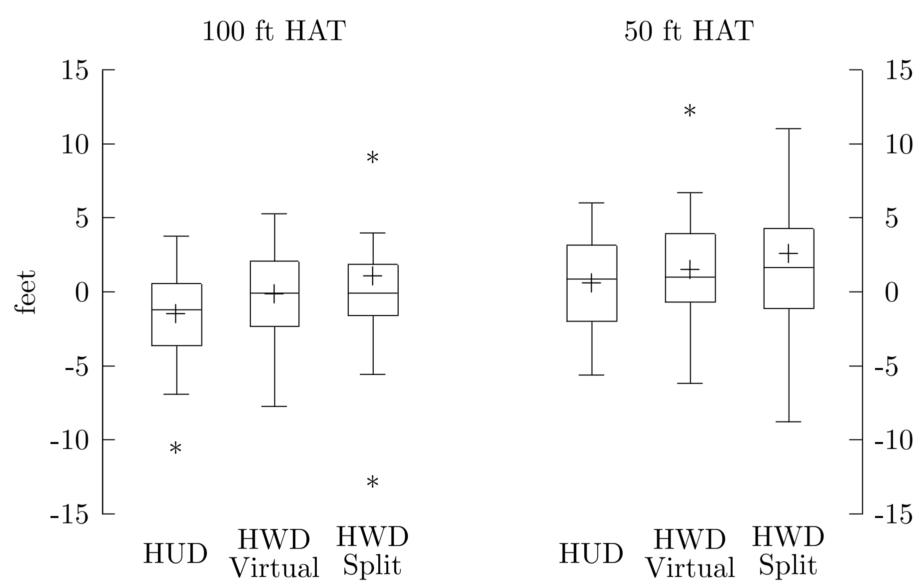

Figure 17. Lateral deviation in the visual segment on approach.

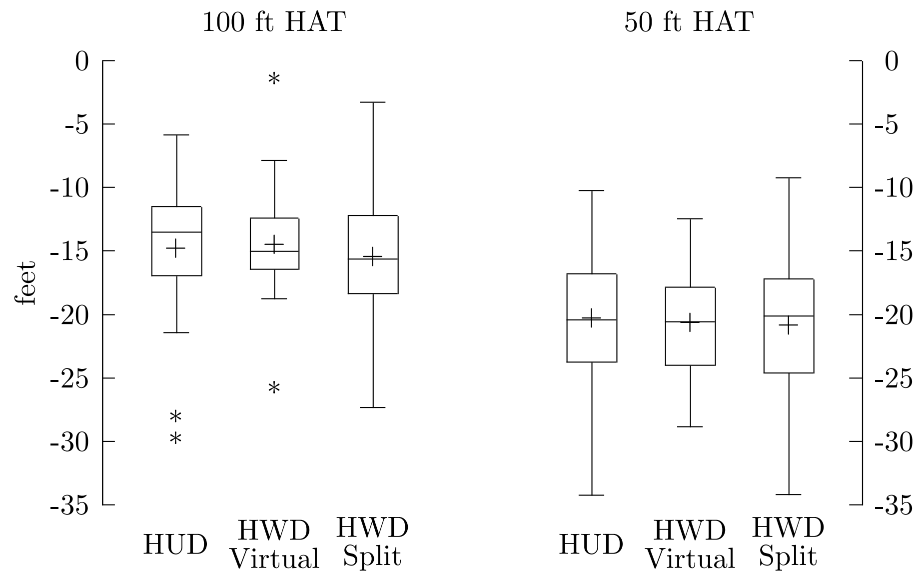

Figure 18. Vertical deviation in the visual segment on approach.

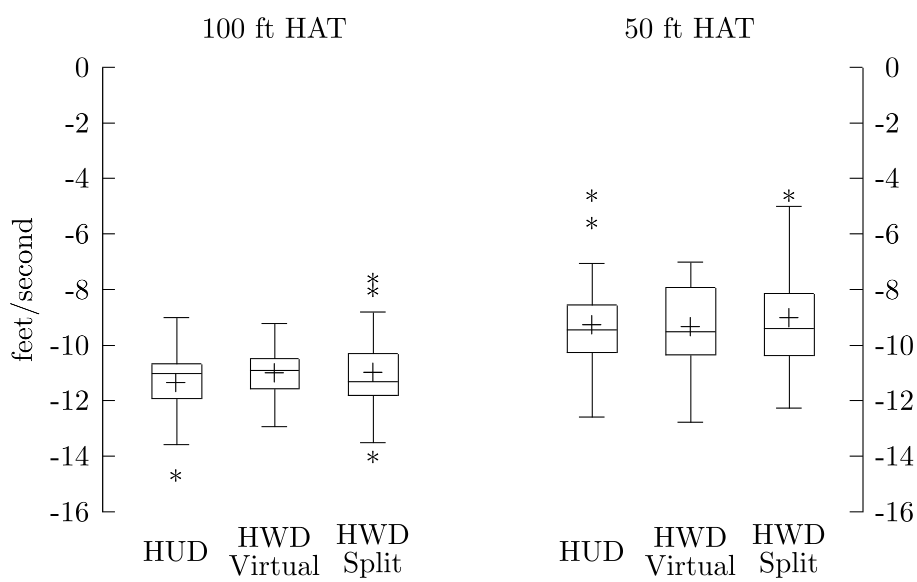

Figure 19. Sink rate on the visual segment on approach.

1) "desired", 2) "adequate", and 3) "not adequate". These performance values are defined in Table 4.

Table 4. Touchdown performance criteria.

	Desired	**Adequate**	**Not adequate**
Lateral	within 27 ft	between 27 and 58 ft	> 58 ft
Longitudinal	750 to 2250 ft	between 200 and 750 ft or between 2250 and 2700 ft	< 200 or > 2700 ft
Sink Rate	0 to 6 ft/sec	6 to 10 ft/sec	> 10 ft/sec

Across all crews, there were a total of 72 landings where the pilot was using the HUD or HWD. In terms of distance (lateral and longitudinal) from the aim point, all landings were "adequate" (Fig. 20) and 88% of the landings were in the "desired" zone.

An ANOVA was conducted on the landing performance statistics of longitudinal distance from threshold, lateral distance from centerline, and sink rate. For all these univariate F-tests, planned contrasts were conducted to evaluate the effect of display concept; the results failed to find any significant effects based on linearly independent pairwise comparisons among the estimated marginal means, ($p > 0.05$). Because the hypotheses were testing whether the HWD concepts were "equivalent" to the HUD, subsequent simple contrasts were conducted that compared the reference category of HUD to HWD-Virtual concept and to the HWD-Split concept. An ANOVA found no significant effects for longitudinal distance from threshold, $F(2, 69) = 0.105, p = 0.901$. Simple contrast measures were not significant between HUD compared to HWD-Virtual concept ($p = 0.701$) or the HWD-Split concept ($p = 0.685$).

Statistical analysis of the distance from the touchdown aim point was not significant, $F(2, 69) = 0.053, p = 0.948$. Simple contrast analyses revealed no significant differences between the HUD, the HWD-Virtual concept ($p = 0.773$) and the HWD-Split concept ($p = 0.988$).

For lateral distance from centerline, the results also found no significant findings across display conditions, $F(2, 69) = 1.589, p = 0.211$. Post-hoc simple contrasts evinced no significant effects for HUD compared to HWD-Virtual concept ($p = 0.151$) or HWD-Split concept ($p = 0.109$).

Sink rate (vertical speed) was also captured at the point of touchdown (Fig. 21). A total of 24 landings were performed with the HUD, 24 with the HWD-Virtual and 24 with the HWD-Split concepts across all EPs. The sink rate of 94% of the landings met either the desired or adequate criteria. Examination of maximum sink rate at touchdown evinces that 4 of 72 scenarios resulted in sink rates greater than 10 ft/sec for the HWD-Split (12.1 ft/sec; 10.2 ft/sec) and HWD-Virtual (10.8 ft/sec; 10.8 ft/sec), and these were all during scenarios with the same flight crew. Across all display concepts, this flight crew averaged 9.9 ft/sec (1.8 ft/sec Standard Deviation (SD)) using the HWD concepts compared to 7.8 ft/sec ("adequate") for the HUD scenarios (1.3 ft/sec SD).

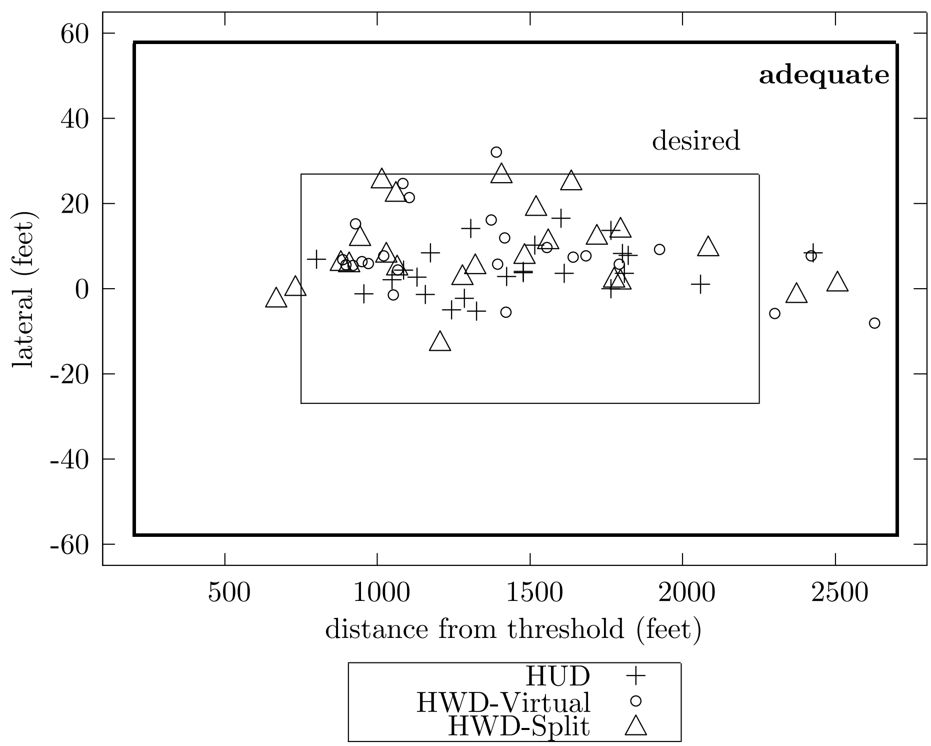

Figure 20. Touchdown point for all approach runs per display concept.

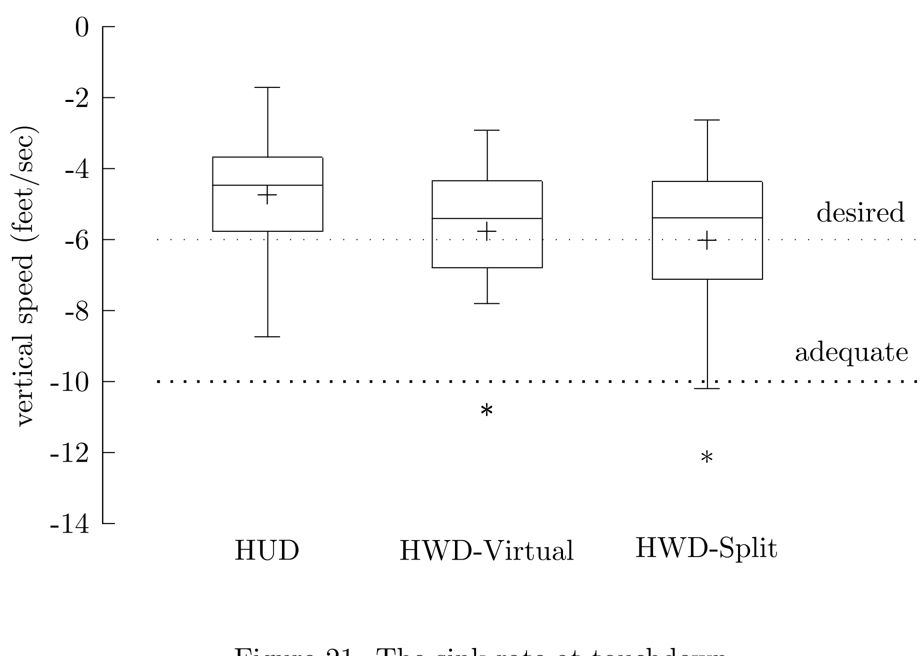

Figure 21. The sink rate at touchdown.

Alternately, of the 72 landings, 46 (64%) had a desired sink rate on landing. Of the 26 sink rates that were not in the desired range, 5 were with the HUD, 10 were with the HWD-Virtual HUD concept and 11 were with the HWD-Split concept. The results evinced no significant effects for sink rate at touchdown, $F(2, 69) = 2.678, p = 0.076$.

In addition to touch down and sink rate on landing, it is important to ensure that the orientation of the airplane on landing does not cause a wing or tail strike with the ground. Figure 22 shows all of the landings where within the maximum allowable pitch and bank angle limits; thus, there were no wing or tail strikes with the ground.

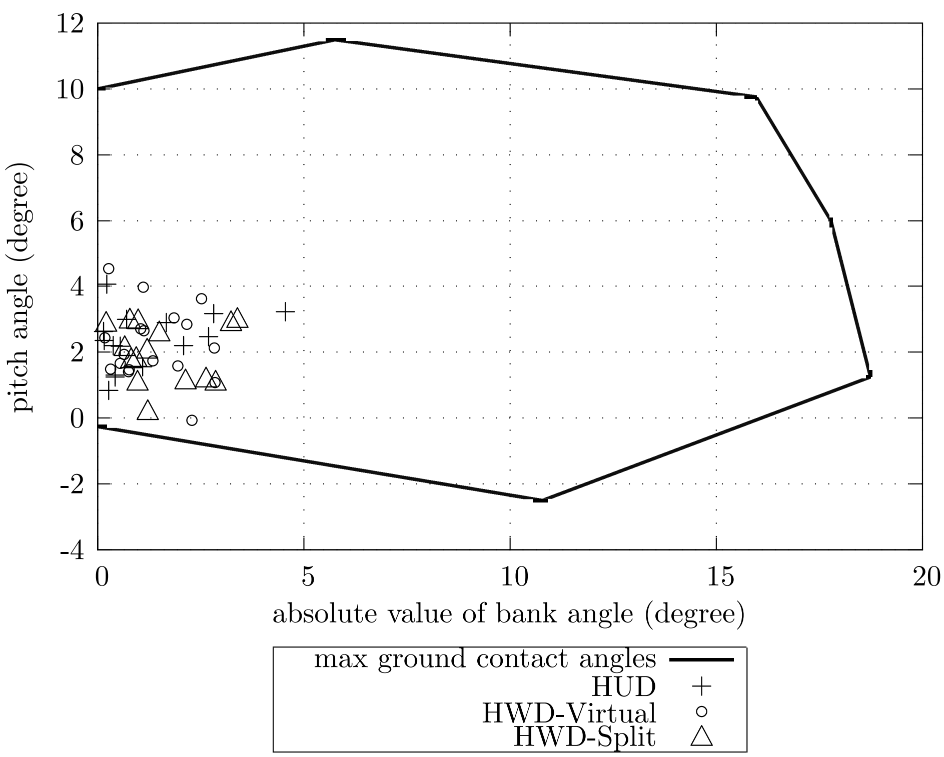

Figure 22. The airplane (ownship) orientation at touchdown.

3.1.5 Landing Rollout Performance

Lateral deviation from centerline statistics (maximum value, RMSE) were analyzed to evaluate how effectively the pilots could maintain centerline during rollout with the different EFVS HUD and HWD display concepts. Only lateral deviation measures are applicable for quantitative performance measurement during touchdown rollout. For lateral RMSE, there were no significant differences found between display concepts, $F(2, 69) = 0.644, p = 0.528$. No significant differences were also found for maximum value lateral deviation, $F(2, 69) = 1.244, p = 0.295$. Table B7 in Appendix B presents the descriptive touchdown rollout statistics.

3.1.6 Off-nominal: Flight Path During Go-around

For the last run on the approach block, ATC called for a go-around at 100 feet AGL. Figure 23 shows a plot of altitude versus time for all of the go-around maneuvers. All go-around runs ended at 1000 feet AGL between 95 and 105 seconds and are not considered operationally different between the Display Concepts. Because of the low number of observations, there was not enough statistical power to conduct parametric analyses on the data. However, no operationally significant differences of the data shown in Fig. 23 were found as function of Display Concept.

3.1.7 Taxi Speeds

For departure scenarios, average taxi speed was calculated when the aircraft was first above 1.0 knots ground speed, and continued until the hold short line at the departure runway.

A Multivariate Analysis of Variance (MANOVA) was conducted on the correlated dependent measures of maximum taxi speed, average taxi speed, and average taxi time for the independent variables of Display Condition (HUD, HWD-Virtual, HWD-Split) and Display Features (Baseline, EV+TD). The MANOVA was not significant for display, $F(6, 156) = 0.519, p = 0.793$; or features, $F(3, 77) = 1.384, p = 0.254$. Figure 24 shows the average taxi speed per Display Condition and Display Features.

The omnibus F-test failed to reveal significant findings ($p > 0.05$) for Display Condition for maximum taxi speed, $F(2, 79) = 0.785$; for average taxi speed, $F(2, 79) = 0.285$; or average taxi time, $F(2, 79) = 0.731$. Simple contrasts between the reference category of HUD compared to HWD-Virtual or HWD-Split concepts for all dependent measures were not significant ($p > 0.05$). For comparison of no advanced features and no enhanced vision (Baseline) to enhanced vision and traffic diamonds (EV + TD) further yielded no significant findings for maximum taxi speed, $F(1, 79) = 3.174, p = 0.079$; average taxi speed, $F(1, 79) = 1.517, p = 0.222$; or average taxi time, $F(1, 79) = 1.096, p = 0.298$.

3.1.8 Taxi Errors

During the course of the experiment, crews deviated from their cleared taxi route (i.e. made a wrong turn) a total of 7 times. Four of the errors were with the HWD display condition and 3 errors occurred with the HUD condition. Note that during the taxi operations, there was a single symbology set used for all display conditions. The 4 errors with the HWD consisted of one crew going past a hold short line without proper clearance, and the remaining 3 errors were crews making a wrong turn. The 3 errors with the HUD consisted of one crew going past a hold short line without proper clearance, and the remaining 2 errors with the HUD consisted of crews turning onto taxiways which deviated from the cleared taxi route.

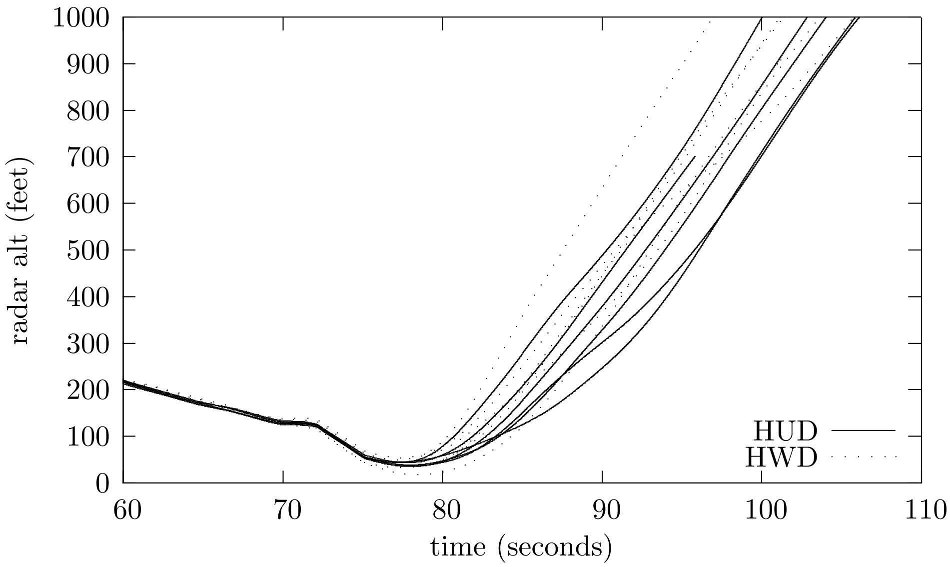

Figure 23. Altitude plot over time for all of the go-around runs.

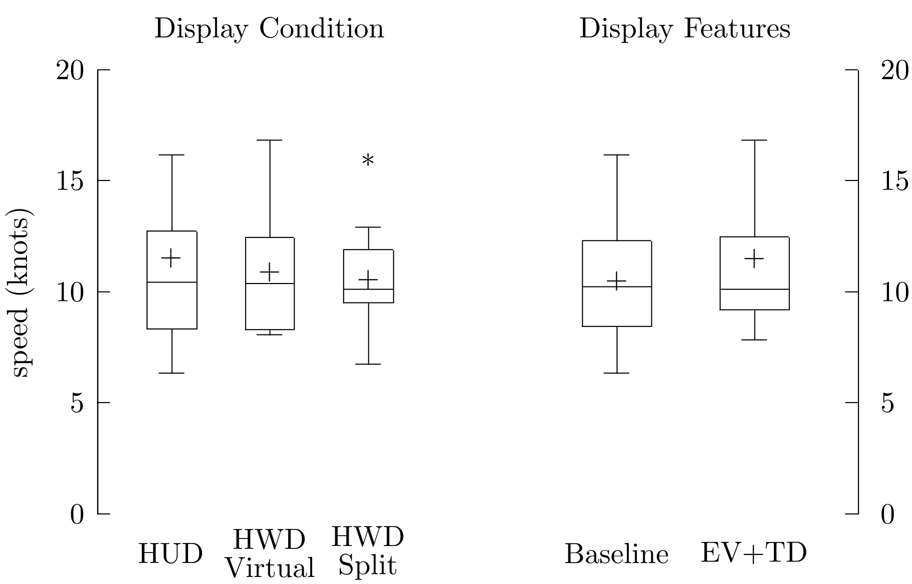

Figure 24. Average taxi speed per display and features for departures.

3.1.9 Centerline Tracking on Takeoff Roll

Statistical analyses were conducted on the centerline tracking during takeoff roll for the dependent measures of centerline localizer Root Mean Square (RMS), centerline maximum localizer deviation, and time during takeoff roll. For this analysis, the takeoff roll was defined to be when the aircraft was on the runway and ground speed was between 30 knots and 128 knots. The ANOVA failed to reveal significant effects for centerline localizer RMS, $F(2, 69) = 1.282, p = 0.282$; centerline maximum localizer deviation, $F(2, 69) = 1.712, p = 0.144$; or time during the takeoff roll, $F(2, 69) = 0.709, p = 0.619$. The takeoff data is shown if Fig. 25.

3.1.10 Off-nominal: Lateral Deviation During Engine-out

For the off-nominal departures, an engine-out event occurred at 100 knots airspeed. The outer edge of the main gear is 17 feet from the aircraft centerline, thus for the aircraft to remain on the 150-foot wide runway, ownship must be within ± 58 feet of the runway centerline. These ± 58-foot limit lines are shown as bold lines in Fig. 26. Eleven of the 12 crews were able to safely stop the aircraft on the 150 foot wide runway. For the single crew that stopped off the runway, the pilot erroneously applied reverse thrust. Figure 27 shows the aircraft airspeed starting at the time of the engine-out.

3.1.11 Eye Tracking Analysis

The head-pitch and the head-yaw data collected from the PFs in the experiment is plotted in Fig. 28 and 29, respectively. All of the head-pitch values reported by the oculometer system with a head position quality greater than 0.8 are grouped by Display Condition (HUD, HWD-Virtual and HWD-Split) and scenario (Approach/Departure). Figs. 28 and 29 represent about 10 hours of head tracking data and 1.8 million data points (HUD: $N \approx 800,000$; HWD-Virtual: $N \approx 500,000$; HWD-Split: $N \approx 500,000$). The lower number of points for the HWD Display Conditions was caused by the HWD obstructing the eye tracking cameras, resulting in fewer data points with sufficient eye tracking quality.

The data in Figure 28 show a median value of 7 degrees versus a median value of 5 degrees for the HWD concepts. Prior to data collection beginning for a crew, the Smart Eye system was calibrated to each pilot without them wearing the HWD. From the stand-alone tests, the data showed that the eye tracking system would not reliably track that pilots' eyes; but, the head rotation data was unaffected with the exception of a head-pitch shift. The entire range of head-pitch points on the HWD box plots is biased down by the approximate amount observed in the stand-alone tests. Therefore, the pilots' scanning tendencies were not a result of the pilots having differing behavior with the HWD; but rather, the HWD affects the Smart Eye system in such a way the head is perceived to be pitched lower than is actually observed.

Results showed that pilots' head pitch and yaw movements were not affected by the display condition. Further, in post-test interviews, pilots did not mention any differences in scanning between the HUD or HWD.

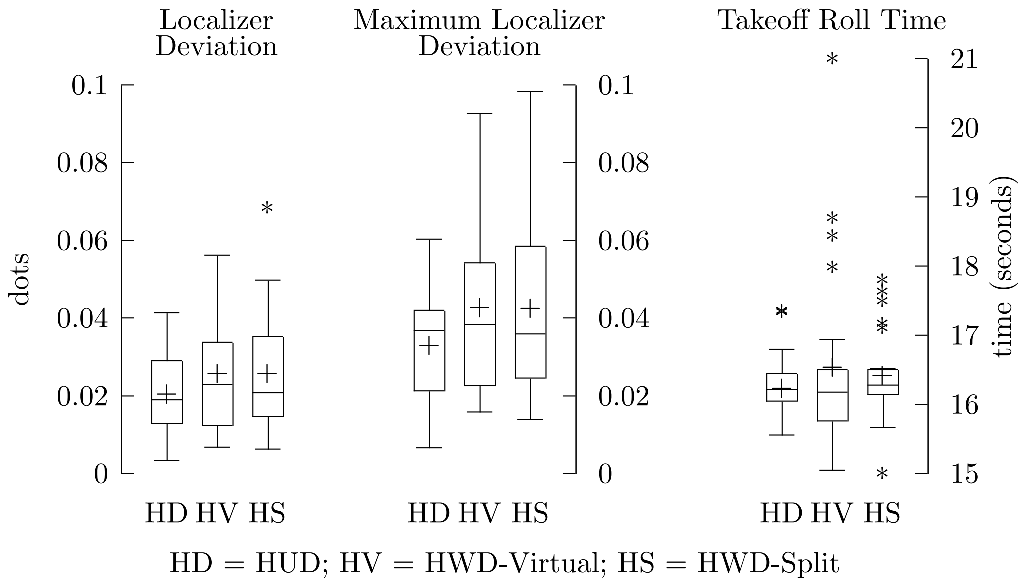

Figure 25. RMS of localizer deviation, maximum localizer deviation and time for takeoff roll.

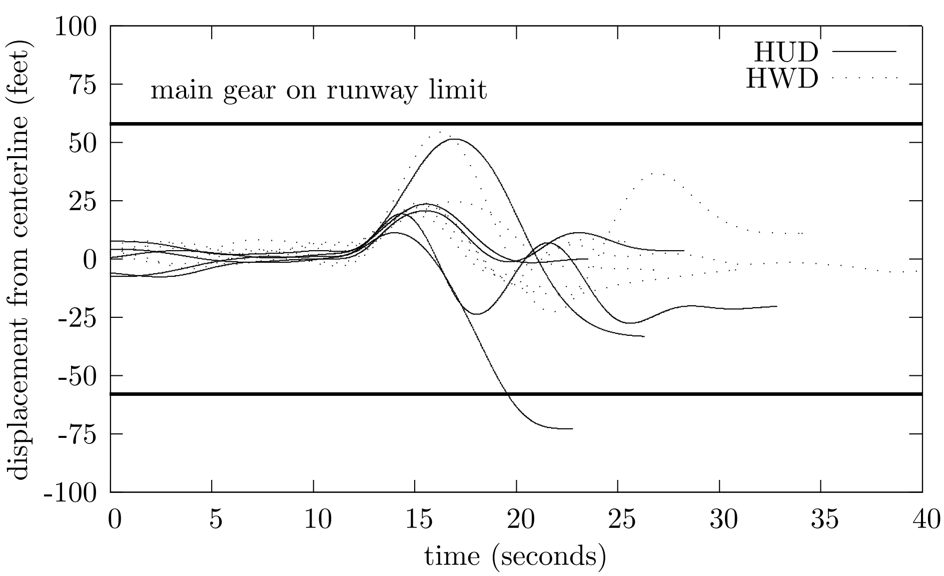

Figure 26. Centerline tracking for for the engine-out run. Data outside the main gear limit lines (two bold lines) indicate ownship off of the runway.

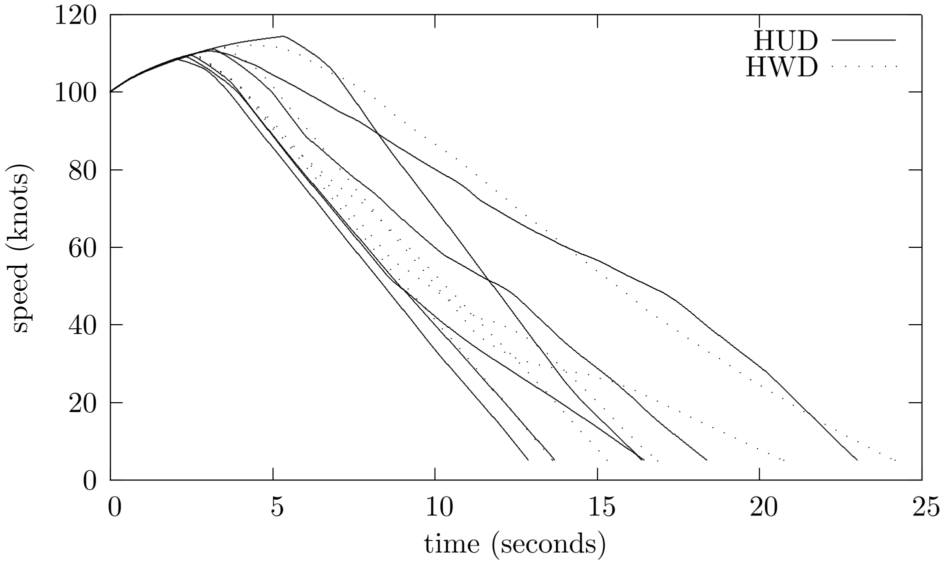

Figure 27. Deceleration plot for the engine-out run.

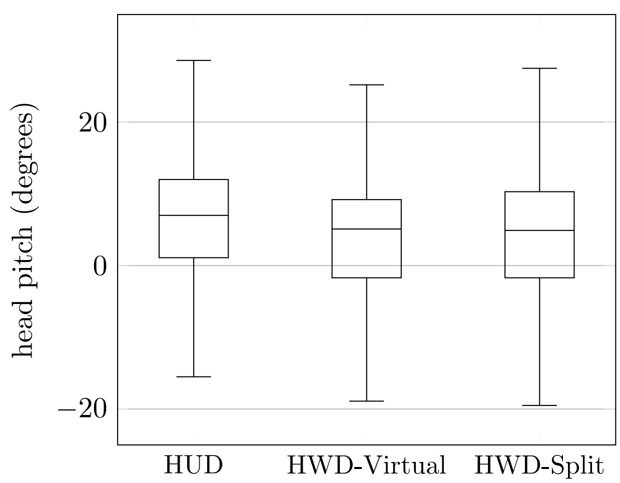

Figure 28. Head pitch values reported from the SmartEye oculometer collapsed across all PFs.

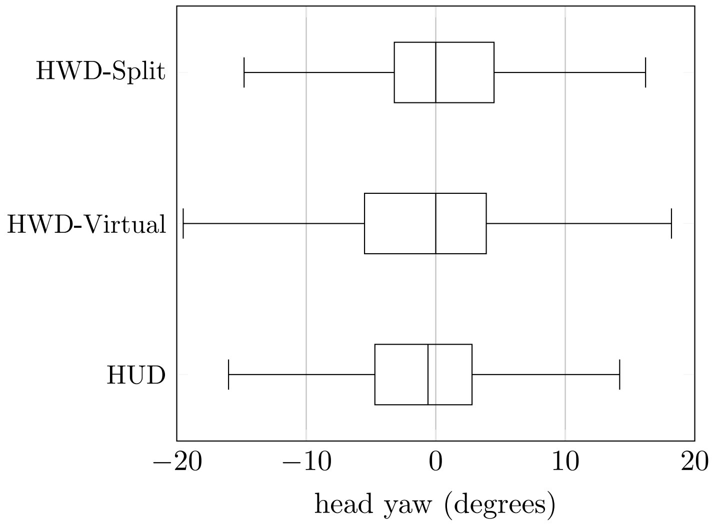

Figure 29. Head yaw values reported from the SmartEye oculometer collapsed across all PFs.

3.1.12 Latency

The latency measurements for the HWD used in this experiment are plotted in Fig. 30. The average total latency was 86 milliseconds. Others have concluded that the helmet-mounted display latency requirements are: 50 milliseconds preferred, 100 milliseconds marginal, 150 milliseconds unacceptable [27].

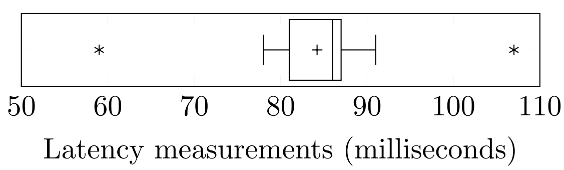

Figure 30. Latency measurement of the HWD system.

3.2 Qualitative results

3.2.1 Situation Awareness

A 3-part SART [21] (Appendix F.1) was administered after each run. The SART provided an assessment of the Situational Awareness (SA) based on the pilot's subjective opinion of three dominant components: demand on the pilot's resources, supply of resources, and understanding of the situation. Pilots rated their perception of the impact of these components using scales from 1 to 7. A total SART score was derived using the formula: $SA = Understanding - (Demand - Supply)$. The range of scores from the application of the formula is from -5 for extremely low SA to 13 for extremely high SA. Figure 31 shows the total SART scores broken-out by Display Condition and Approach/Departure.

No significant differences were found between the PF and the PM for any of the dependent measures ($p > 0.05$), so the analysis was collapsed across role.

Analysis of the SART results indicated that pilots did not report any significant differences in SA across display concepts for either the approach and landing scenarios ($F(2, 69) = 0.879, p = 0.420$); or the departure scenarios ($F(2, 69) = 0.735, p = 0.483$).

3.2.2 Workload

Workload was assessed via the AFFTC 7-point subjective workload scale [22](Appendix F.2) and the NASA TLX [23](Appendix F.3). The AFFTC 7-point scale consisted of a single number to represent overall workload where 1 represents "Nothing to do; No system demands" and 7 represents "Overloaded; System unmanageable; Essential tasks undone; Unsafe." The pilots' responses for the AFFTC workload ratings are shown in Fig. 32. No significant differences were found between the PF and the PM for any of the dependent measures ($p > 0.05$) so the analysis was collapsed across role.

The NASA TLX consisted of 6 scales associated with mental, physical, and temporal demand, performance, effort, and frustration level. The TLX scores across the 6 components were averaged to determine an overall workload rating (Fig. 33). Note that the performance component of the NASA TLX scale was reversed scored to compute the total (average of all NASA TLX components) score. A paired comparison was not done between the NASA TLX components.

For the mental workload results collected during the approach scenarios, no significant results were found for either the NASA TLX, $F(2, 69) = 0.481, p = 0.620$, or AFFTC, $F(2, 69) = 0.724, p = 0.488$. For the departure scenarios, the ANOVA results for mental workload evinced no significant differences for NASA TLX, $F(2, 69) = 0.905, p = 0.195$; or AFFTC, $F(2, 69) = 1.672, p = 0.195$.

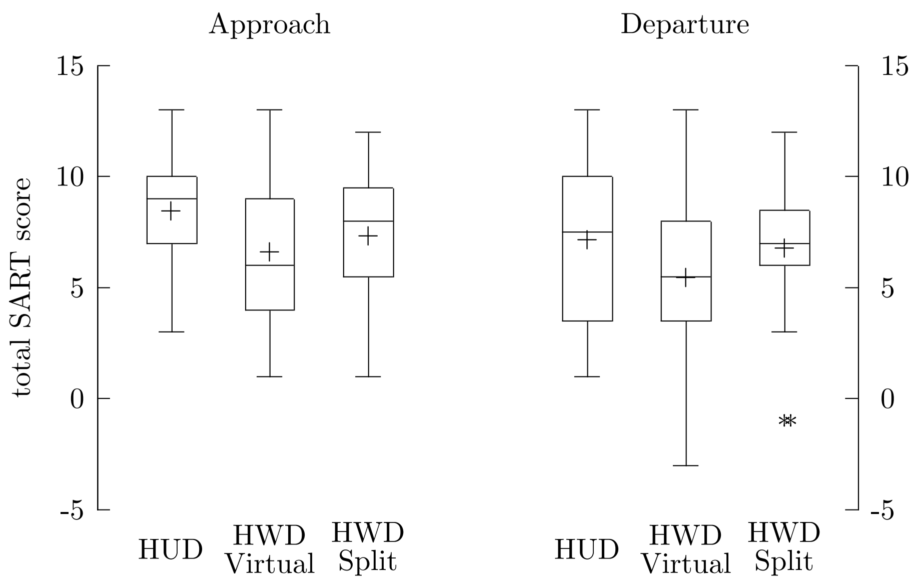

Figure 31. SART scores for the PF.

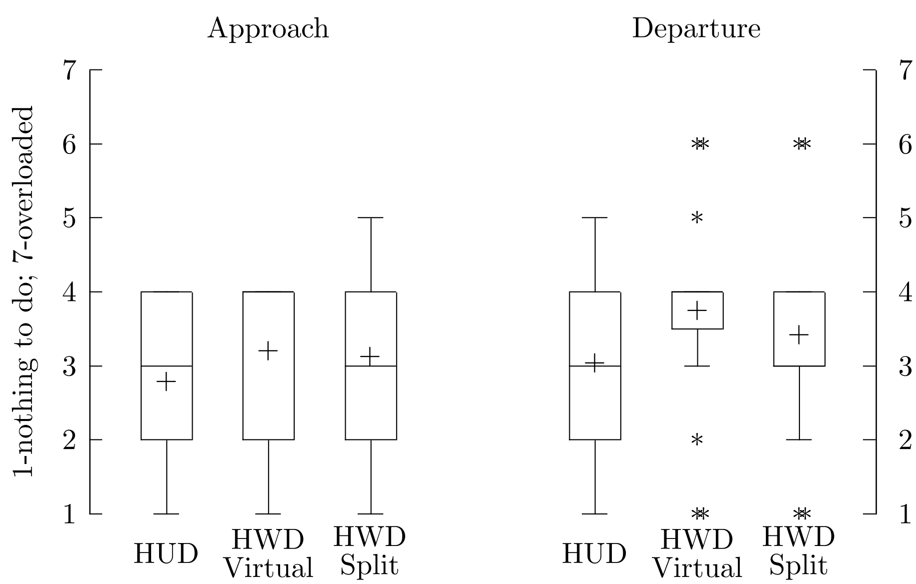

Figure 32. AFFTC workload rating by the PF.

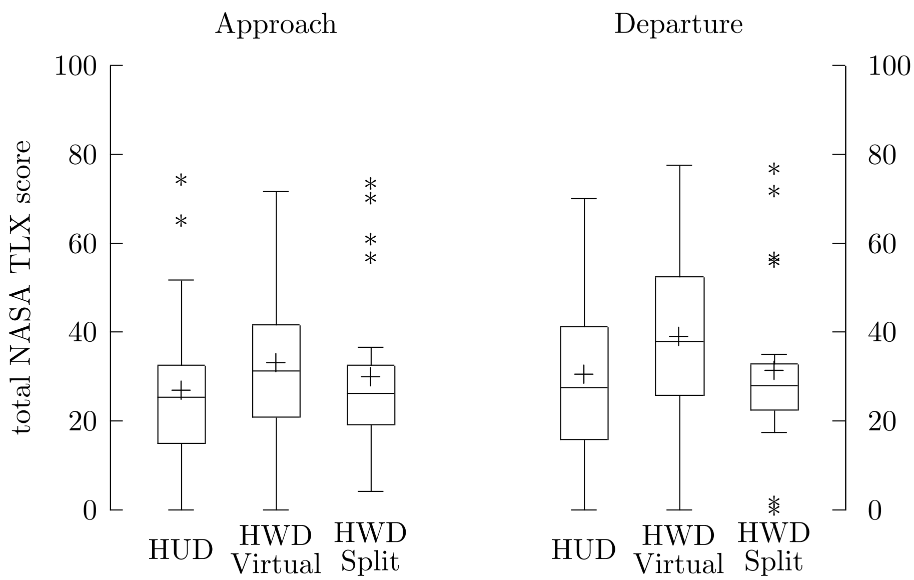

Figure 33. NASA TLX rating (total score) for the PF.

3.2.3 Simulation sickness

One concern with HWD systems is latency-induced sickness. Crews were given a SSQ [24] throughout the day (see Appendix F.5). The SSQ is a list of 16 symptoms (general discomfort, fatigue, nausea, etc.) which the crews were asked if they were experiencing at that moment. If they did experience a symptom, they were asked to rate the severity of the symptom as slight, moderate or severe. A total of 120 SSQs (for all 12 crews) were administered with one SSQ given at the beginning of the day and one SSQ given at the end of the day. The remaining SSQs were given at the end of the Display Condition block.

Of all the SSQs administered, only 6 (5%) had scores of non-zero (see Table 5). The 6 scores were equally distributed among the 3 Display Conditions and all were "slight symptoms." Responses of "none" are not reported in Table 5.

The SSQ symptom ratings can be grouped into 4 scoring categtories: Nausea, Oculomotor, Disorientation and a Total Score. Normally, these scores are calculated per pilot to gauge how a simulator affects a pilot's well-being. For the purposes of this paper, the SSQ scores were computed on a per display basis to compare reported symptoms across the Display Concept. These scores are calculated based on the formula provided by Kennedy [24] and reported in Table 6.

The ratings show that while the HUD had more Nausea related symptoms (General Discomfort, Stomach Awareness, Burping), pilots reported more Oculomoter (General Discomfort, Fatigue, Headache, Eye Strain) and Disorientation (Fullness of Head) related symptoms with HWD system. These slight variations in the small number of reported symptoms resulted in an almost equivalent SSQ Total Score for each Display Concept.

Table 5. Frequency of responses to the SSQ per Display Condition.

	HUD	HWD-Virtual	HWD-Split
General Discomfort	1	1	1
Fatigue	1	1	1
Headache	0	1	1
Eye Strain	0	0	1
Fullness of Head	0	1	1
Stomach Awareness	3	0	0
Burping	0	1	1
Totals	**5**	**5**	**6**

Table 6. SSQ scoring per Display Condition.

	HUD	HWD-Virtual	HWD-Split
Nausea-related	38.16	19.08	19.08
Oculomotor-related	15.16	22.74	30.32
Disorientation-related	0	13.92	13.92
Total Score	**22.44**	**22.44**	**26.18**

3.2.4 Post-Run Questionnaire

After each data trial, a questionnaire (Appendix F section F.4) was given to crews who were asked for their level of agreement to 10 statements on a 7-point Likert scale where a rating of 1 was "strongly disagree," 2 was "disagree," 3 was "slightly disagree," 4 was "neither agree nor disagree," 5 was "slightly agree," 6 was "agree," and 7 was "strongly agree." The 10 statements on the questionnaire were:

A. I was aware of ownship position.

B. I was aware of traffic and other vehicles during operations.

C. The display concepts were effective for maintaining SA.

D. The display concepts were effective for management of mental workload.

E. The display concepts contributed to communication effectiveness (ATC and crew).

F. The display concepts promoted effective crew resource management, coordination, and cohesion.

G. The display concepts contributed to perceived safety.

H. The display concepts were effective for detection of potential surface conflicts.

I. If applicable, the display flown was equivalent for use during the approach/departure as the HUD.

J. The display concepts provided for adequate visual references and awareness (for approach, in terms of flight path, altitude, runway, landing zone; for departure, in terms of maintaining centerline and runway heading).

The post-run questionnaire was analyzed for approach and departure scenarios during both nominal and off-nominal trials. For both the nominal and off-nominal approach scenarios, no significant effects were found for the individual run questionnaire items or the grouped scaled constructs (see Table 7) formed by combination of the individual run questionnaire questions.

Table 7. Grouped construct definition.

Group Construct	**Post-run Question**
Hazard Awareness	A, B, G
Attention Management	C, D, H
Communication Efficacy	E, F
Operation Equivalence	I, J

For the approach, no significant differences were found for any post-run questionnaire items for display comparisons ($p > 0.05$). The statistical analyses for the post-run questionnaire for the nominal and off-nominal departure scenarios for display condition revealed no significant effects ($p > 0.05$). The crews' responses to each of the post-run statements are shown in Appendix C.

One objective of the departure scenarios was to evaluate the addition of traffic symbology and enhanced vision to determine whether these display features would significantly enhance traffic and hazard awareness. Pilot responses to the post-run question "I was aware of traffic and other vehicles during operations" is shown in Fig. 35. The data was demarcated between scenarios with either (a) no display features (Baseline) or (b) display features which consisted of enhanced vision and traffic diamonds (EV+TD). The statistical results evince that the presence of enhanced vision and traffic diamonds for any Display Concept significantly enhances traffic and other vehicle awareness (Question B), $F(1, 70) = 7.671; p = 0.006$; perceived safety (Question G), $F(1, 45) = 4.33; p = 0.043$; and detection of potential surface conflicts (Question H), $F(1, 45) = 9.337; p = 0.004$. Flight crews also rated "hazard awareness" to be significantly greater under EV+TD scenarios than baseline scenarios, $F(1, 70) = 12.456; p = 0.001$.

Flight crews did not rate the display concepts, for the EV+TD to baseline comparison, as significantly different for attention management, $F(1, 70) = 3.098, p = 0.083$; communication efficacy, $F(1, 70) = 0.780, p = 0.380$; or operational equivalency, $F(1, 70) = 0.830, p = 0.365$. No other post-run questionnaire items were found to be significant between EV+TD and baseline conditions ($p > 0.05$).

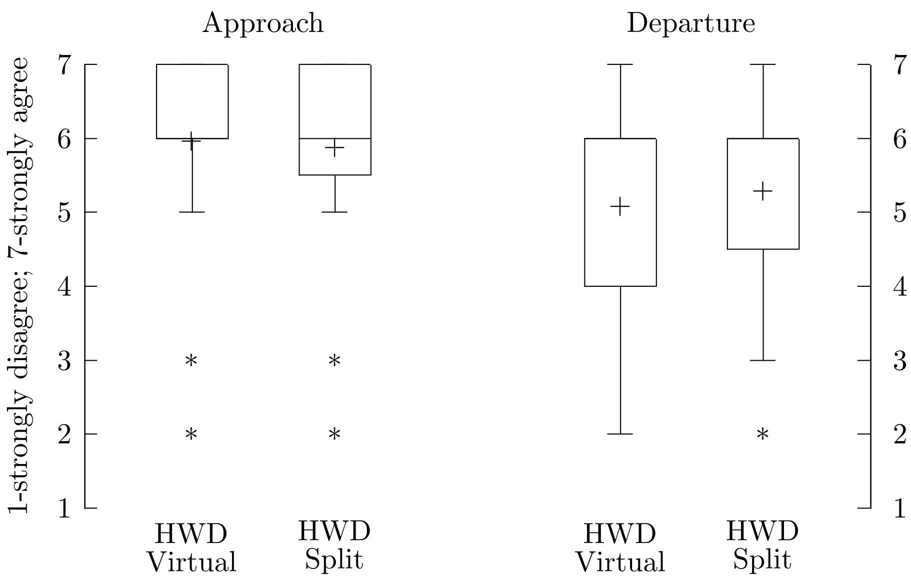

Figure 34. PF ratings of HWD equivalence to the HUD.

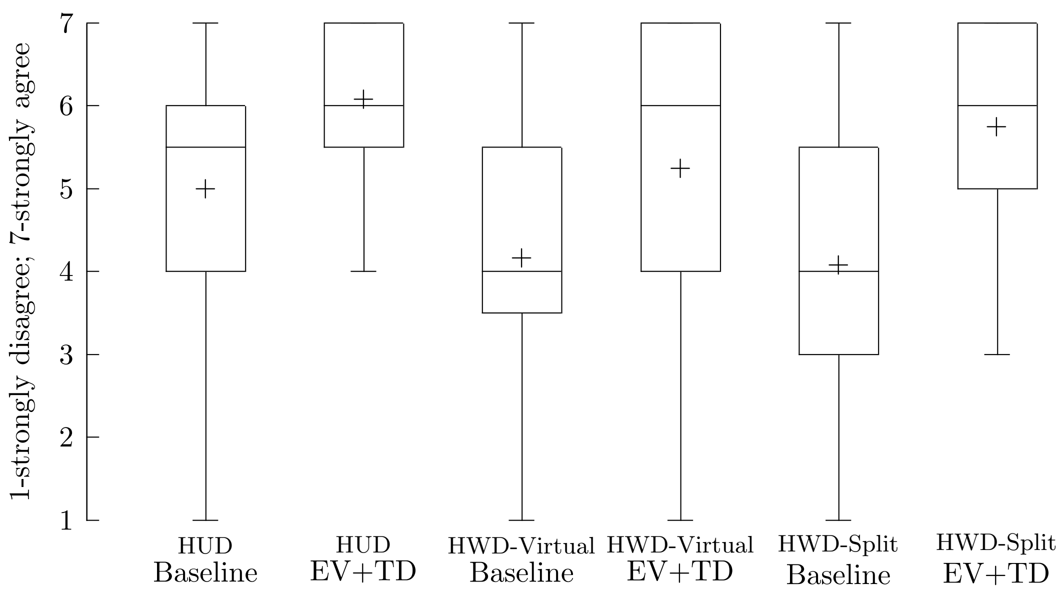

Figure 35. Post-run Question B: Traffic awareness ratings per augmented reality condition by the PF.

3.2.5 Post-Test Questionnaire

During a semi-structured verbal debrief session, the PFs were asked several questions (see Appendix F.6) that were designed to elicit their responses and ratings for various research objectives. This session generally lasted between 30 and 45 minutes and included items commented upon in the questionnaires, additional issues the pilots noticed during the runs, specific items the researchers had noticed during that particular crew's scenarios, and general comments concerning this experiment. Pilot ratings to each question are presented in Appendix D.

The PFs were asked to provide pairwise ratings of "display equivalence" between HUD and HWD concepts (see Post-Test questions 3, 4, and 5). Geo-means were calculated based on the ratings and subsequent parametric statistics [28, 29] were conducted on these means. The non-significant interaction of display (HUD, HWD-Virtual, HWD-Split) and the operation (Approach, Departure) suggest that flight crews rated the HUD, HWD-Virtual, and HWD-Split to be equivalent in terms of "operator use" during both the approach and departure scenarios, $F(4, 44) = 1.062, p = 0.387$.

Based on the ratings provided for Post-Test questions 3, 4, and 5, pilots were asked for improvements to the HWD system if it was not completely equivalent to a HUD on approach or departure. The pilots' comments are listed below.

- The field-of-view of the HWD is too small.
- A pilot controlled declutter switch that could remove the EV imagery and/or symbology would be desirable for the HWD. Having a pilot manually adjusting the brightness on the visual segment is not tenable. A "flip-up" type of HWD would be useful as a declutter method.
- The HWD system needs to be optimized (reduce the latency and more comfortable to wear).
- I preferred the HWD-Virtual to the HWD-Split concept because the HWD-Split appears to have more jitter. The jitter in the HWD-Split concept seems unsafe.
- I would like the HWD to be much lighter and more ergonomic.
- The HWD-Virtual concept needs to be stabilized (reduce latency).
- For the surface symbology, add a 'distance to go' to the next taxiway.
- I'm concerned about attention capture of the HWD because it is always "in your face."

Pilots were asked (see Post-Test question 6), "In your opinion, could a head-worn display (HWD) replace a head-up (HUD) display?" The pilots' comments are listed below.

- Yes, but the HWD used in this experiment needs to be optimized.

- Yes, but reduce the lag (total system latency) in the HWD. At times, I couldn't focus on the EV image in the HWD but had no problems with the EV image on the HUD. I would like more field-of-view for the FLIR especially for turns in surface operations.
- Yes, but the ergonomics of the HWD used in this experiment need improvement: weight and comfort, especially.
- Not with the current system. In my opinion, the lag combined with turbulence would make the HWD unusable.
- Yes, but improve the ergonomics and the image stability.
- Yes, but needs optimization.
- Yes, but improve comfort.
- Yes, they are equivalent.
- Yes, improve on ergonomics.
- Yes.

Pilots were asked (see Post-Test question 7), "What did you like/prefer about the HWD compared to the HUD (what were its advantages, if any)? What did you dislike/not prefer about the HWD compared to the HUD?" The pilots' comments are listed below.

- No advantages for the HWD. The weight and being on the face are a disadvantage.
- HWD advantage: you can easily look around past the eye box of the HUD. HWD disadvantage is comfort; you don't wear a HUD.
- The HWD-Split concept allows the pilot to see airspeed and altitude all the time. The disadvantage of the HWD is the comfort.
- The HWD advantage was for surface operations, you can look around and still have taxi information which you can't do with a HUD.
- The HWD-Split allows you to see some conformal symbology (velocity vector, pitch ladder, EV image) but always see the non-conformal symbology (airspeed, roll indicator, altitude).
- With the HWD, you have an unlimited field-of-regard.
- The HWD is better (than the HUD) for ground (surface) operations.
- The HWD unlimited field-of-regard allows finding traffic much easier compared to a HUD.
- With the HWD-Split, you can always have some critical flight information.

- The HWD is better for traffic awareness because of the "look-around" capability.

- The HWD has an unlimited field-of-regard which aids in traffic awareness.

To determine the fidelity of the simulation facilities used to collect the approach and departure data, flight crews were asked (see Post-Test question 13) to rate the simulator quality compared to 1) the pilot's airline simulators, 2) real-world, and 3) quality of the simulated EV. The mean rating on a scale of 1 (very poor) to 7 (excellent) for all qualities was approximately 5 and 95% of all responses for all qualities were 4 or greater.

4 Discussion of Simulation Results

The main goal of this study was to determine if a HWD system could provide equivalent performance as a HUD. Approaches, surface operations and departures were performed by EPs in low visibility conditions using either the HWD or the HUD. The same symbology set was used on both the HUD and the HWD as to not introduce a confounding factor in the analysis. Even though the Lumus display glasses are color capable, only monochrome green was used to match the HUD. The results showed that there were no statistical differences in the crews' performance in terms of the approach, surface operations, departures and off-nominal scenarios. Further, there were no statistically significant differences between the HUD and HWD in pilots' responses to questionnaires. Although these results showed that there were no statistical differences in the crews performance, there are still technical hurdles to be overcome for complete display equivalence including, most notably, the end-to-end latency of the HWD system. Also, comfort, boresighting, brightness and declutter controls, field-of-view, and HWD weight factors must be considered.

The crews were asked if the HWD (after flying the HUD) was equivalent for use during the approach/departure as the HUD. The data, shown in Fig. 34, show that the HUD and HWD were subjectively rated as equivalent and in general, were rated as "agree" for the approach and "slightly agree" for the departure. The brightness settings of the HWD were difficult to adjust in real-time; thus pilots would wait until the next run to adjust brightness. On departures, some pilots commented that the takeoff symbology on the HWD, combined with the lower visibility (300-foot RVR vs 1000-foot RVR on approach) and the EV image would make it difficult to track the centerline on takeoff.

On approach, there were no significant differences in the crews' touchdown performance measures when using either the HUD or the HWD. For vertical speed, the average sink rate at touchdown trended higher for the HWD concepts than for the HUD though the average sink rate for each display configuration was in the "desired" performance range. However, there were twice as many sink rates out of the "desired" range with a HWD concept compared to the HUD. Pilots commented the HWD used in this experiment greatly reduced peripheral vision. A widely held belief is that the loss of peripheral vision along with degraded visibility (1000 ft RVR) may have contributed to the slightly higher sink rates [30]; however, Kramer et al. [31] showed field-of-view of the head-up display may play a larger role in touchdown performance than peripheral cues. Though a head-tracked HWD has an unlimited field-of-regard, the HWD display device had a fixed field-of-view about 10 degrees horizontal and 15 degrees vertical smaller than the HUD. Future studies would be required to determine if the smaller field-of-view of the HWD and/or the combination of the loss of peripheral vision of the HWD tends to cause higher sink rates.

There were no significant differences in flying the straight-in approach in 1000-foot RVR. Pilots were able to track the localizer and glideslope regardless of the Display Condition. Pilots did comment that for the HWD approaches, they felt compelled to keep their heads as still as possible. This is directly related to the HWD total system latency (85 milliseconds) as any slight head movement would manifest itself into an apparent symbology oscillation/misalignment. While others have concluded that the acceptable range of latency is between 50 and 150 milliseconds, the total system latency may need to be on the order of 20 milliseconds for pilot acceptability [19] for HUD equivalence on a high-resolution, large field-of-view display. Though there were no statistical differences between the Display Conditions, future studies should investigate the acceptable limits of total system latencies for head-tracked HWD systems throughout the various phases of flight.

Simulation sickness was not an issue with either the HWD or the HUD. The frequency of symptoms reported was evenly distributed among the 3 display concepts. Further, no simulation sickness symptom was reported with a severity greater than 'slight'.

For errors made by crews during taxi, the number of errors was almost evenly split between the HWD (4 errors) and the HUD (3 errors). Considering that there were twice as many HWD condition runs as HUD, this would suggest that the display device alone is not a factor in the crew's propensity for committing an error. As the symbology set was consistent across the displays, this would suggest that crews need more state information during low visibility surface operations to maintain SA. Some pilots commented that their SA would be improved if the surface symbology contained a distance to the next cleared taxiway.

Off-nominal scenarios were used to gain insight in crews performance with the HWD during non-normal events. For all of the go-around runs, pilots were able to reach an altitude of 1000 feet in as little as 17 seconds and no greater than 27 seconds, a 10 second range regardless of Display Condition. For the engine-out on departure, most pilots were able to safely stop the aircraft on the runway within a range of 12 seconds regardless of the Display Condition. Pilots commented that executing the go-around or the rejected takeoff with the HWD was similar to having a HUD. Given the crews' comments and their ability to complete the off-nominal task within seconds of each other regardless of display concept, the display variance is not considered operationally significant.

Comparing the HWD-Virtual to the HWD-Split concepts, there were no statistical differences between the display conditions in terms of performance or subjective comments. Some pilots preferred the HWD-Split as critical aircraft state information was easily readable because it was fixed on the glasses; however, they felt the conformal symbologies (the EV image and the velocity vector) on the HWD-Split concept appeared to have more "movement" compared to the HWD-Virtual concept even though the conformal symbologies were displayed the same on both the HWD-Virtual and HWD-Split concepts. The perception of more movement is from the conformal symbology components continually being corrected to the real-world due to head movement and latency while 2-D symbology components remained in a fixed location on the HWD glasses. In addition, pilots commented that with the HWD-Split concept, symbology clutter situations can arise which can be distracting to the pilot. For example, the velocity vector symbol could become obscured if the pilot's head orientation was such that it overlaid on top of one of the screen fixed symbologies.

If the total system latency of the HWD could be reduced to near zero, the HWD-Split condition could be less attractive for certification as the clutter/obscuration issue would have to be researched and resolved. However, several pilots commented that they liked the HWD-Split configuration because it provided aircraft state information wherever they were looking. The need and/or utility of having information on the HWD even when they are not looking straight ahead for commercial operations needs to be investigated. A low latency HWD-Virtual concept would appear to be a HUD in terms of symbology and functionality.

One issue raised was the certification of the bore-sighting procedure. A HUD system is bore-sighted when it is installed on an aircraft and should remain in alignment. An equivalent HWD system would need to provide feedback to the flight crews as to the integrity of the alignment and head-tracker health.

The HWD was rated statistically the same as a HUD in terms of situation awareness and workload. However, one statistically significant result was ratings for the post-run question "I was aware of traffic and other vehicles during operations." Though Display Condition (HUD, HWD-Virtual, HWD-Split) was not statistically significant, the presence of EV imagery and traffic symbology was statistically significant. As expected, the addition of EV and traffic symbology allowed crews to monitor traffic not visible out-the-window because of visibility conditions. Pilots' subjective comments also support that the EV imagery and the traffic symbology increased their SA.

The quantitative and qualitative analysis support the position that a HWD system can be equivalent to a HUD at least for the low visibility operations conducted in this experiment. Pilot comments also supported the hypothesis that a HWD is equivalent to a HUD; however, many pilots qualified their comments with the need for improvements to the HWD system used in the test. Comfort and optimization were two common concerns with the HWD. The weight of the HWD system used in the experiment was acceptable for short periods but there was concern that it was too heavy for longer periods (an hour or greater). The HWD system was also slightly unbalanced, due to the placement of the head tracker, which was countered by using an easily adjustable strap to hold the HWD firm to the pilot's head. This strap, while effective for stabilizing the HWD on the pilot's head, had a tendency to create "hot spots" on the pilots' nose and ears. Thus, the reason many pilot's commented on the need for a more ergonomic design.

5 Flight Test

Following the simulation test, a flight test was conducted which focused on HUD equivalence as described in AC 90-106 §91.175 (m) by simulating Instrument Meteorological Conditions (IMC) to the EP flying with the HWD system. This proof-of-concept flight test evaluated the maturity of the HWD technology undergoing development and evaluation at the NASA Langley Research Center and generate data to support industry and government guidance for HWD development for commercial and business aircraft applications. The HWD system described earlier in the simulation experiment (section 2.1.1) was used in the follow-on flight test (Fig. 36). As in the simulation experiment, the HWD used monochrome HUD-type symbology and imagery in the forward-looking, bore-sight direction to be analogous with a certified HUD with EFVS capability; that is, a virtual HUD or HUD-equivalent concept. In addition to nominal HUD symbology, an extended zero-pitch line with heading tick marks and traffic symbology were displayed on the HWD.

A FLIR camera was not available for the flight test; thus, a 640x480 pixel monochrome visible light camera was used to simulate enhanced vision imagery. Further, a HUD was not installed on the research aircraft; thus, highly experienced test pilots with HUD experience participated in the flight test and provided their opinion of how the HWD compared to a HUD. For safety considerations, all flights were conducted in daytime Visual Meteorological Conditions (VMC). IMC conditions were simulated using a Commercial Off-The-Shelf (COTS) IMC training device clipped-on to the HWD, which blocks the out-the-window view but allows the pilot to view the head-down instruments.

5.1 Test Aircraft

The HWD evaluations were flown on board the NASA Langley Beechcraft King Air (BE-200) aircraft (Fig. 37). The BE-200 is a corporate-sized twin turbine aircraft that can be flown single pilot. Approach speed and procedures were similar to commercial operations. The flight test crew included: a NASA safety pilot in the left seat, the EP in the right seat, a system operator, a test director, and up to 2 observers in the cabin.

5.2 HWD Flight Test Symbology

The flight symbology set (Fig. 38) was typical HUD symbology for a commercial/business transport. The flight symbology consisted of a flight path marker (also referred to as the velocity vector), an airspeed dial with ground speed, a pitch ladder with a -3° flight path angle reference line, localizer and glideslope scales, conformal runway outline with touchdown tick marks and an extended centerline, roll scale, heading display, and altitude indicator. All flight symbology was rendered in monochrome green with the exception of two symbology elements: 1) a 360° zero-pitch line rendered in white (see Figs. 39 and 40); and 2) traffic diamonds rendered in cyan (see Figs. 38 and 40). The 360° zero-pitch line was an extension of the zero-pitch line on the pitch ladder and extended across 360° of azimuth.

Figure 36. The HWD worn by the evaluation pilot.

Figure 37. The King Air research aircraft at NASA Langley.

Located next to the EPs sunvisor was a panel that consisted of 2 rotary knobs that controlled the HWD brightness and a push button for boresighting. The first rotary knob allowed the EPs to control the overall brightness of the HWD, and the second rotary knob was used to control the brightness of simulated FLIR imagery. When the push button was pressed, an alignment pattern would appear on the HWD which the EP would line up with a holographic-projected sight mounted in a fixed, known position on the glareshield. Once the EP aligned the HWD alignment pattern with the holographic sight, the EP would press the push button a second time to boresight.

During surface operations, a reduced symbology set was used (Fig. 41). When the nose wheel was on the ground and the ground speed was less than 80 knots, the flight symbology set would automatically transition to the surface symbology on the HWD. The surface symbology set consisted of ground speed, heading, current taxiway the aircraft is on, and the next taxiway on the cleared route. The display of FLIR imagery during surface operations was the independent variable in the flight test.

Traffic icons were displayed to denote the positions of actual aircraft based upon reported Automatic Dependent Surveillance-Broadcast (ADS-B) surveillance data. The traffic in the flight test was incidental and was not part of a rehearsed scenario. Traffic icons were rendered on the HWD in a perspective format as unfilled, cyan-colored, 2-dimensional diamonds. On the surface, traffic diamonds were only displayed if the traffic position was within 1 nautical mile of ownship. This 1 nautical mile filter was used to reduce symbology clutter by only displaying traffic close to ownship. The traffic diamonds were rendered as a 30-foot diameter 3-dimensional object; therefore, the diamonds would appear larger as the distance between the traffic and ownship drew closer.

5.3 Evaluation Pilots

Seven test pilots served as EPs and were recruited based on the following selection criteria:

Figure 38. The flight symbology set overlaid on a simulated FLIR image. The blue diamond symbology depicted ADS-B traffic.

- Each EP held an Airline Transport Pilot rating or equivalent;
- Each EP was trained or holds equivalent experience and served as a test pilot for the government (Department of Defense (DoD), NASA, or FAA) or commercial aircraft company;
- Each EP had HUD and/or Helmet-Mounted Display (HMD) experience, having flown at least 100 hours of HUD or HMD, pilot-in-command operations;
- Each EP had EV/EFVS experience, either military, general aviation, or commercial;
- Each EP had 20/20 visual acuity; if correction required, only correctable by use of contacts (no glasses as they were not compatible with the HWD tested).

The EPs were from industry (corporate and commercial), military, and NASA. Average flying experience was more than 20 years, greater than 9000 total flight hours, and greater than 1000 hours of HUD experience.

5.4 Evaluation Pilot Training

The EPs were given a 45-minute classroom briefing to explain the display concepts and the evaluation tasks for the experiment. After the briefing, the HWD was donned by the EP to ensure proper fit prior to boarding the airplane. Following the fitting session, EPs attended a safety briefing. At the end of the day, a post-test

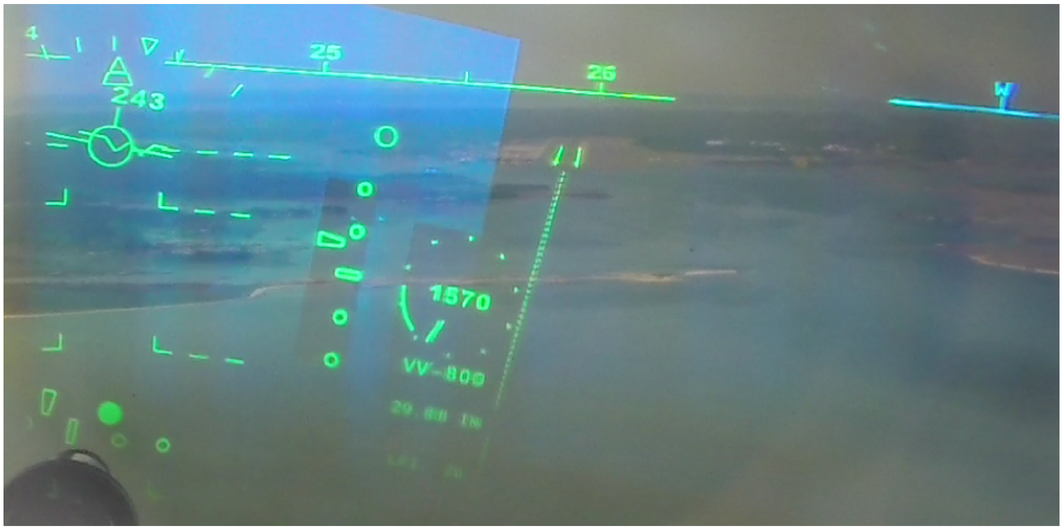

Figure 39. This figure is a screen capture of the pilot looking to the right of the Virtual HUD in VMC. The runway outline and extended centerline (dashed line) are shown in the middle of the figure to the right of the altitude dial. A partial view of the off-boresight 360° zero-pitch line is shown in white in upper right.

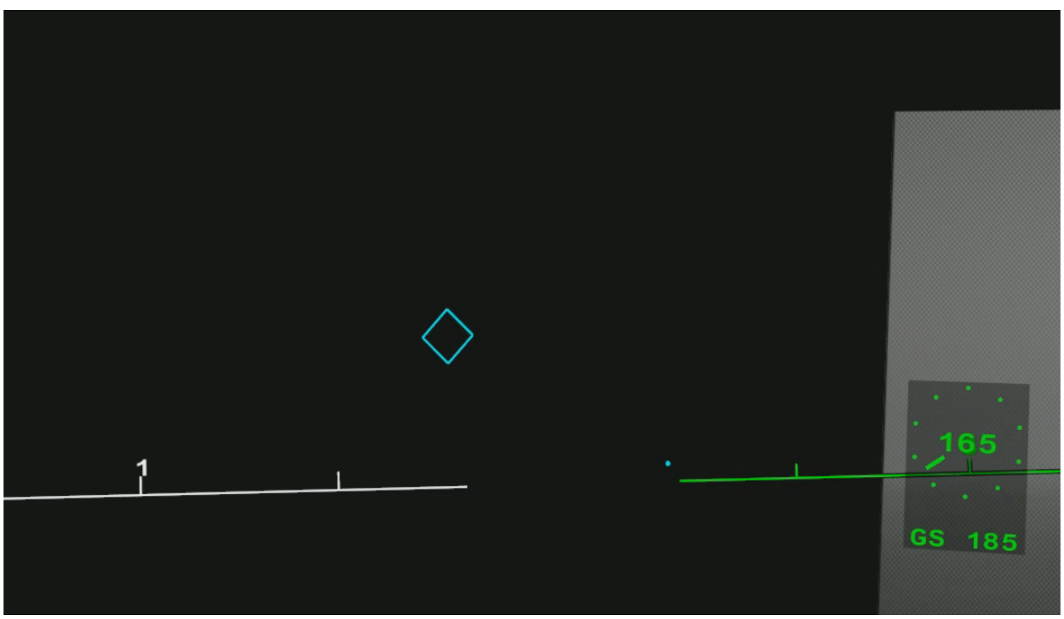

Figure 40. This figure is a screen capture of the pilot looking to the left of the Virtual HUD in IMC (notice the airspeed indicator of the Virtual-HUD on the right-side of the figure). The blue diamond symbology shows ADS-B traffic not in the forward field-of-view. The off-boresight 360° zero-pitch line is shown in white.

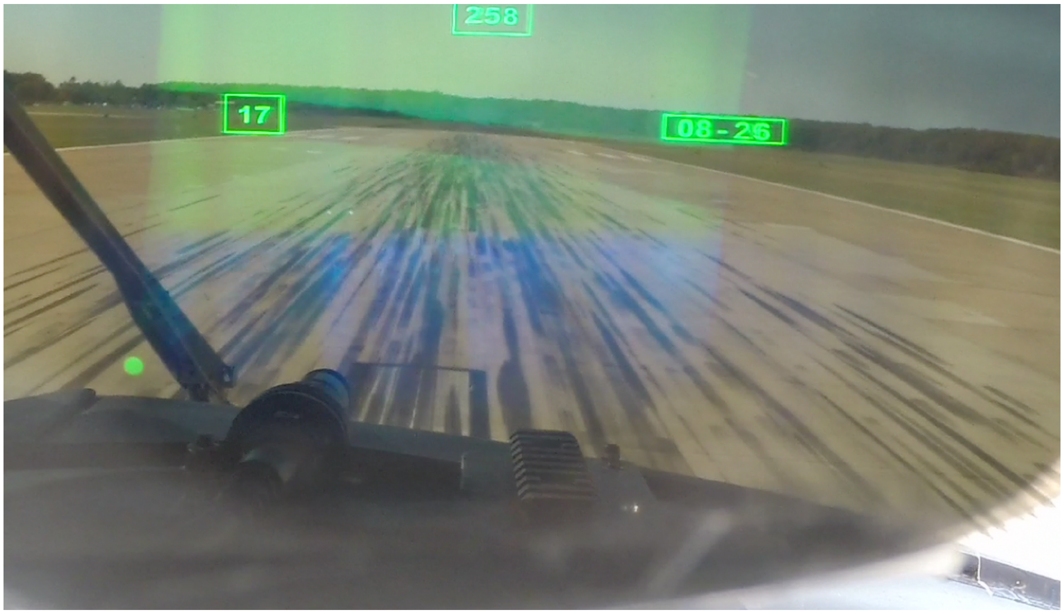

Figure 41. The surface operations symbology set. The ground speed was boxed on the left (17 in this figure), the magnetic heading was boxed on at the top of the display (258°) and the current taxiway or runway boxed on the right (runway 08-26 in the figure). The simulated FLIR is shown as the translucent green rectangle behind the symbology. The green filled circle in the lower left indicated nominal head tracking.

interview was conducted to solicit the EPs comments on the flight with the HWD. The total duty time for an EP was approximately 6 hours.

5.5 Methodology

Flight test operations included surface operations and ILS approach scenarios designed to evaluate HWD performance and structured following existing SAE International HUD performance requirements and Minimum Aviation System Performance Standards (MASPS) for EFVS (RTCA DO-315). There was 1 independent variable for the surface operations; the presence/absence of enhanced vision imagery (simulated FLIR ON/OFF). There were 2 independent variables for the approach phase of the flight test; the EPs presence/absence of natural vision (IMC/VMC) and the display of enhanced vision imagery (simulated FLIR ON/OFF). The visibility was set to either VMC or simulated IMC, by use of a flip-down view-limiting device identical to the traditional clip-on instrument training aid. This device blocked the evaluation pilots view out the window but still allowed full use of the HWD and head down instruments. For VMC data trials, clip-on sunglasses were used to increase the contrast of the HWD symbology and imagery. A typical data flight lasted approximately 1.5 hours of flight time. The partial factorial test matrix is shown in Table 8. Note that the 'surface operation/IMC/FLIR on' condition was not tested as it had the unrealistic effect of completely blocking the EP's out-the-window view. In other words, for surface operations in low visibility conditions, at least some portion of the airport environment remains visible with the unaided eye.

Table 8. Experiment run matrix for each EP.

	VMC	**VMC**	**IMC**
Phase of flight	**FLIR ON**	**FLIR OFF**	**FLIR ON**
Surface Operations	1	1	
Approach	1	1	1

5.6 Evaluation Task

The task for EPs was to operate the aircraft as the "pilot flying" in a 2-person crew. The safety pilot acted as the "monitoring pilot" and handled the radio and ancillary tasks as well as executed tasks at the request of the EP. The evaluation of several aspects of the HWD were requested including: general head tracker operation during maneuvers, system latency, symbology and imagery conformance, and display optical performance in variable lighting. The testing was conducted in four phases for each EP's flight:

1. Ground Test: On-Ramp, after engine-start, donning and bore-sighting, the EP evaluated the system performance of the HWD and gave an initial impression (see Appendix F, section F.7) of the device while the aircraft was still in the chocks.

2. Surface Operation Trials: Two surface operation trials were conducted in VMC: the initial taxi-out, and the return taxi-in after the flight. The FLIR was randomly enabled for either the taxi-in or taxi-out. Run questionnaires were administered after each trial (see Appendix F, sections F.1, F.2 and F.8).

3. Approach Flight Trials: All approaches were either to Runway 26 at Langley Airforce Base (FAA identifier: KLFI) or Runway 25 at Newport News / Williamsburg International Airport (FAA identifier: KPHF). For the research aircraft, the localizer and glideslope were not available on the data bus; thus, the localizer and glideslope indications were generated via the display software using the GPS-based aircraft position data (see Appendix E). Three approaches were conducted by the EPs in VMC and simulated IMC, with and without FLIR (see Table 8). Run conditions were randomized across EPs. Run questionnaires were issued after each data collection trial. (see Appendix F, sections F.1, F.2 and F.8).

4. Post-Test: Following completion of the in-aircraft testing, post-test surveys and interviews were conducted to capture pilot comments, ratings, and assessment (see Appendix F, section F.9 and section F.10). The EPs were given this questionnaire in a semi-structured interview format asking their ratings

on comfort, optical performance, bore-sighting procedures, brightness, color shifts, distortions, conformality, FLIR usage, eye strain, headaches, perceived safety, safety compared to HUDs, and equivalency to a HUDs.

Post-run questionnaires after the surface operations and approach trials consisted of: 1) a 3-part SART [21] form (see Appendix F, section F.1), 2) an AFFTC 7-point workload scale [22] (see Appendix F, section F.2), and 3) 10 questions which asked the EPs their agreement (1-strongly disagree to 7-strongly agree) with statements on readability, comfort, usability, imagery, symbology, field-of-view, obscuration, impairment, and conformality (see Appendix F, section F.8).

6 Flight Test Results and Discussion

The primary purpose of the flight test was to demonstrate the use of HWDs as an equivalent display to a HUD during actual aircraft operations; however, the flight test also yielded dependent measures that provided descriptive statistics and quantitative and qualitative statistical results. From these evaluations, the maturity of the technology will be assessed.

6.1 Quantitative Results

6.1.1 Flight Technical Error on Approach

A number of testing criteria has been examined for use in evaluation and qualification of FTE. AC 120-29A [32], Appendix 2, Paragraph 6.2.1 provides a means for evaluation of approach performance. The AC defines minimally acceptable performance for approaches from the Final Approach Fix (FAF) (e.g., 1000-foot HAT) to 200-foot HAT in terms of localizer ($<$ 2/3 dots or 1/3 full scale deflection) and glideslope tracking ($<$ 1 dot or 1/2 full scale deflection). Another useful objective measure of FTE are Practical Test Standards (PTS) which delineate minimally acceptable performance as $<$ 3/4 full scale (1.5 dot) localizer and glideslope scale deflection between the FAF and decision height (FAA-S-8081-4D) [33].

FTE data was collected on the approach from an altitude above ground of 1000 feet to approximately 300 feet. Between 200 and 300 feet, the data trial ended, and the safety pilot took control of the aircraft. The data in Fig. 42 shows that the EPs were able to track the localizer within +/-0.2 dots. The box plot in Fig. 43 illustrates the maximum localizer deviation (absolute value) across all runs in both the visual conditions (IMC/VMC) and with both FLIR conditions (on/off).

The EPs' lateral performance was well within acceptable standards. The localizer data showed the EPs flew within a 0.1 of a dot. EPs stated the extended centerline not only helped in lateral line-up on final, but it was a great situational awareness tool on the turn to final. Pilots could track the extended centerline throughout the turn due to the unlimited field-of-regard of the HWD.

The data in Fig. 44 shows the glideslope for all EPs for the approach. The box plot in Fig. 45 shows the maximum glideslope deviation (absolute value) across all runs in both the visual conditions (IMC/VMC) and with both FLIR conditions (on/off). Both acceptable performance limits ($<$ 1 dot or 1/2 full scale deflection) are shown in Fig. 45 - the PTS and the AC 120-29 - and illustrate that the maximum glideslope deviation across all runs in both the visual conditions (IMC/VMC) and with both FLIR conditions (on/off) meet the PTS and the AC standards. However, there was 1 outlier for the IMC FLIR condition. This outlier occurred during an approach with moderate turbulence which the pilot commented the turbulence was increasing workload and making it difficult to precisely follow the guidance symbology. Overall, the data shows outstanding FTE performance.

Quantitative results show that EPs were able to fly within accepted flight technical error criteria while using the HWD even on the first approach with no training or practice runs. From the qualitative results and the EP comments, important issues came to light. Several EPs encountered light to moderate turbulence during

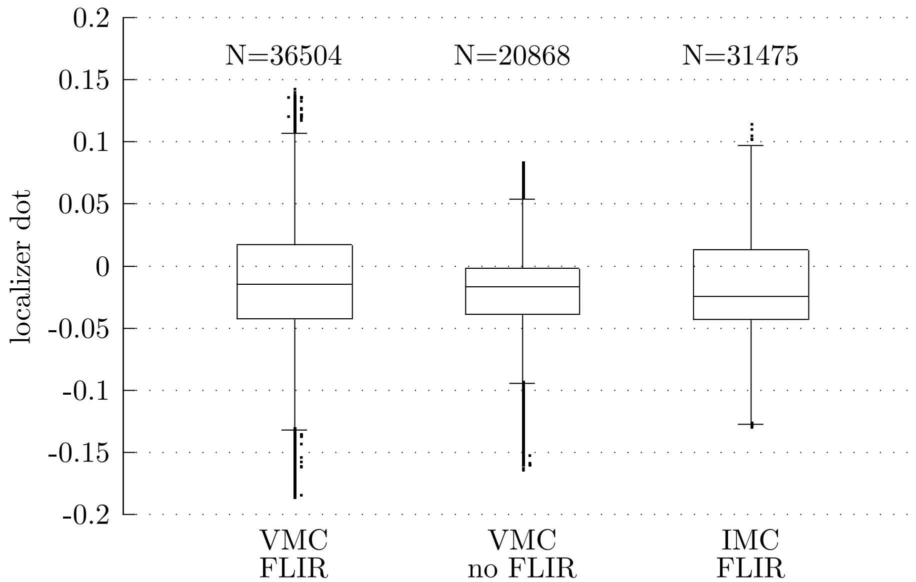

Figure 42. Localizer on approach collapsed across all EPs.

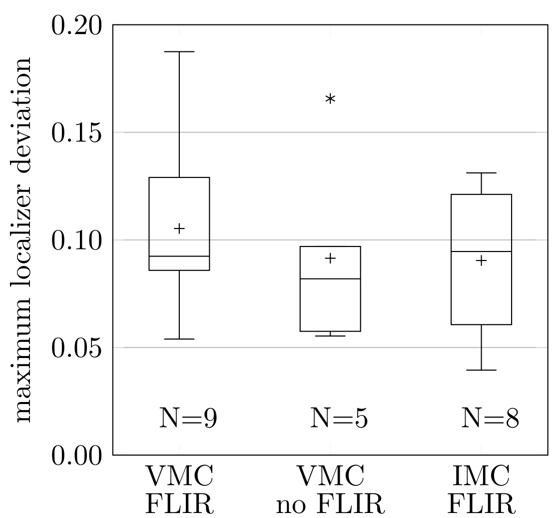

Figure 43. Maximum localizer deviation (absolute value) for each approach broken-out by Display Condition.

approaches which combined with system latency and created a "jittery, bouncy" display that was difficult to read and follow. With head-tracked head-up displays, image stabilization is critical for pilot readability and acceptability. Turbulent conditions only exacerbate the image jitter. Allowing that test pilots are able to overcome many system limitations and still fly with low flight technical error, additional system improvements are necessary.

6.2 Qualitative results

The data collected also consisted of pilot response data gathered after each trial (post-run) and post-flight. The post-run data included measures of situation awareness, mental workload, and system attributes (comfort, usability, conformality, HUD equivalency, etc.). Data were collected for surface operations; however, due to the small number of trials, the data is reported but not analyzed.

6.2.1 Situation Awareness

A 3-part SART [21] questionnaire was administered after each run to assess SA. Figure 46 shows the SART scores across conditions; visual conditions (IMC/VMC), FLIR conditions (on/off), and operation type (surface operations/approach).

An ANOVA analysis of SART scores on approach failed to find a significant effect for FLIR (On, Off), $F(1,18) = 0.761, p = 0.394$; or Visibility Condition (IMC, VMC), $F(1,18) = 2.424, p = 0.137$.

The lowest SART scores, including the one point indicated by the asterisk in the plot as an outlier (i.e., outside 1.5 times the IQR of the data), were from 1 flight with moderate turbulence. This so-called outlier was from the first approach with the HWD. The EP commented that SA increased with more use of the HWD system throughout the flight. In general, the EPs commented that the addition of the FLIR for the IMC condition was useful; but, FLIR for the VMC condition wasn't necessary and was distracting both for approaches and surface operations.

6.2.2 Workload

Mental workload was assessed using the AFFTC questionnaire [22]. The mean workload estimation rating across all visibility and FLIR conditions was a '3' which represents "Moderate Activity; Easily Managed; Considerable Spare Time." The AFFTC ratings for FLIR and visibility conditions for the approach and surface operations are shown in Figure 47.

The statistical results showed no significant effects for FLIR, $F(1,18) = 0.795, p = 0.384$; or Visibility Condition, $F(1,18) = 3.231, p = 0.089$. For the 2 labeled as outlier (asterisks) ratings with the VMC FLIR condition on approach, the EPs commented that the FLIR image was not properly aligned. This misalignment was likely due to image latency causing the FLIR image to "swim" with respect to the real world. Further, one EP commented that localizer scale was too low which increased the workload by having to tilt the head down in order to view the scale.

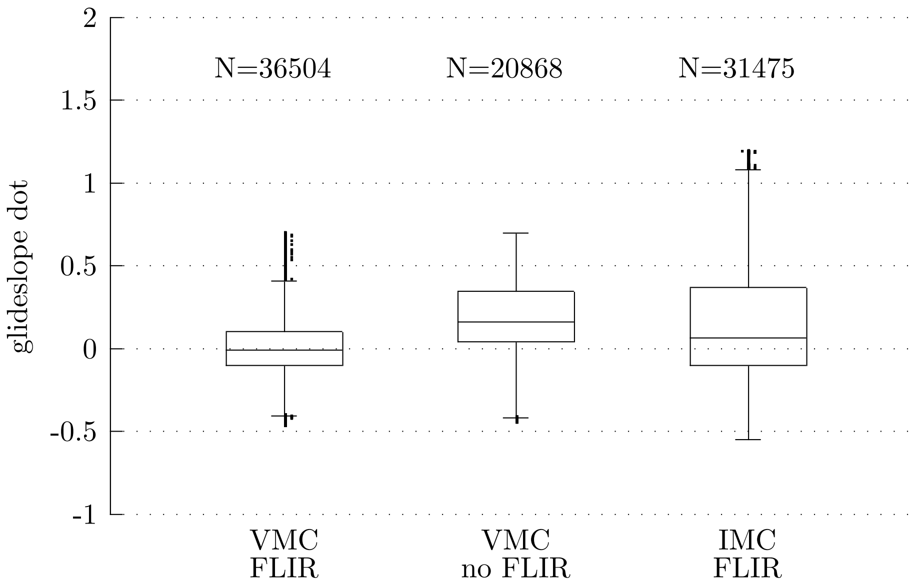

Figure 44. Glideslope on approach collapsed across all EPs.

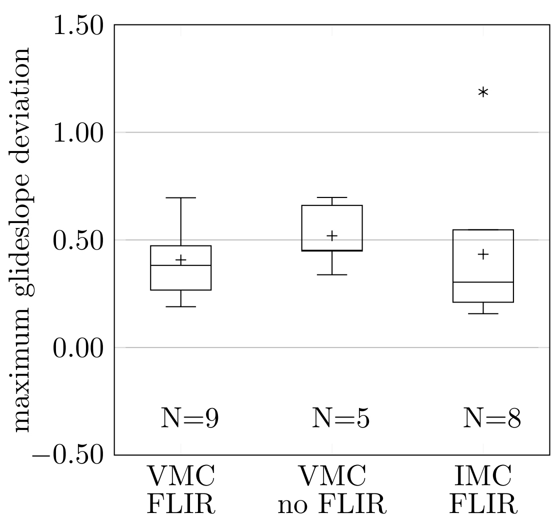

Figure 45. Maximum glideslope deviation (absolute value) for each approach for each Display Condition.

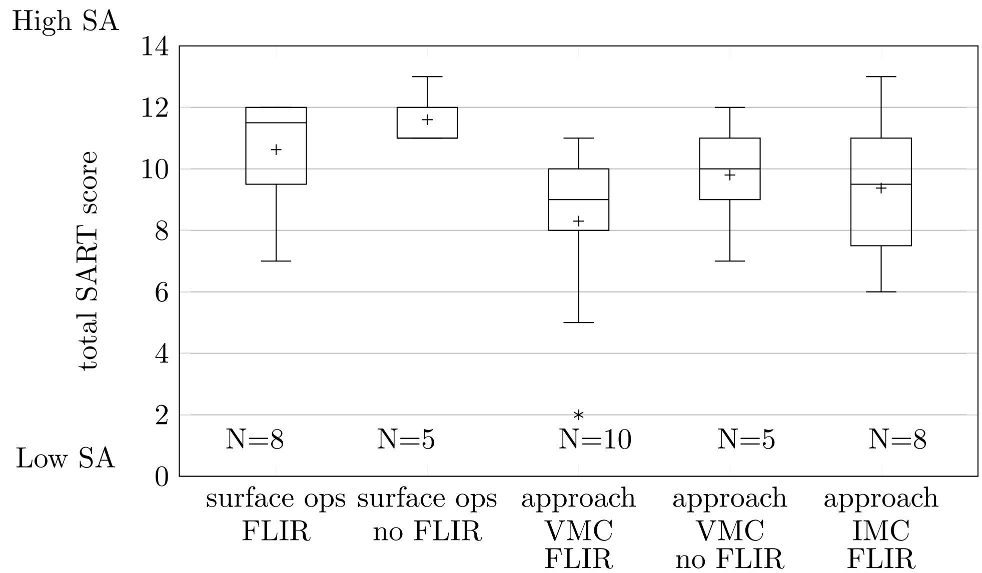

Figure 46. Total SART score.

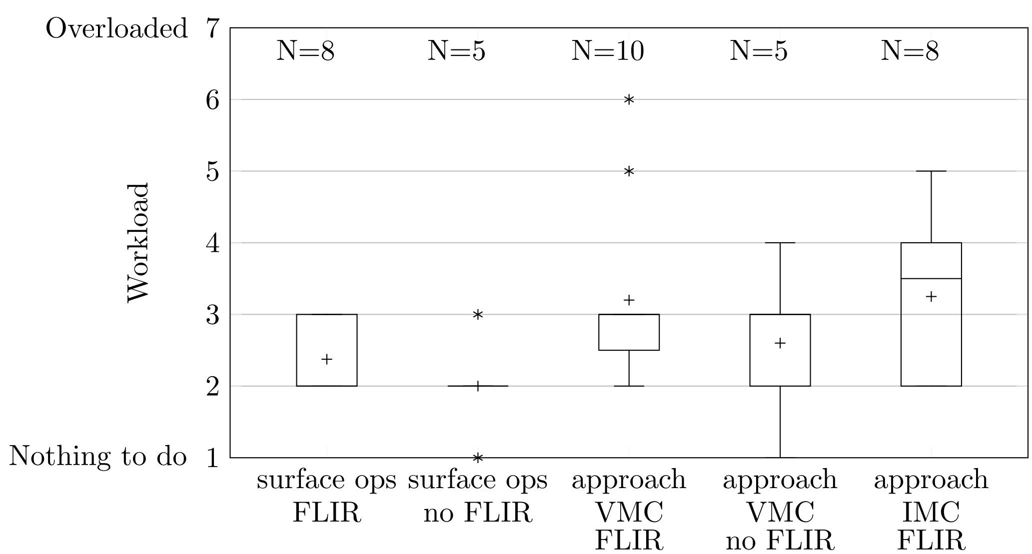

Figure 47. Workload ratings.

6.2.3 Post-Run Questionnaire

Following each data trial, EPs were asked their level of agreement (1 - strongly disagree to 7 - strongly agree) to 9 statements. The post-run questionnaire is in Appendix F, section F.8. The ratings given by the EPs for the approach and surface operation trials are shown in Figs. 48 through 56. No statistically significant results were found ($p > 0.05$) for the post-run questionnaire except for ratings of symbology clarity.

The results of the post-run statements are as follows:

- *A: The HWD caused me to experience eye strain.* Eye strain was not prevalent although the ratings spanned from 'strongly disagree' to 'agree' with means trending towards 'slightly disagree' for the approach trials and 'disagree' for the surface operations trials (Fig. 48).

  EP Comments: Bumps and latency issues cause the eyes to refocus leading to eye strain.

- *B: The HWD caused me to experience headaches.* In general, the data indicated that the HWD did not cause any headaches for the EPs. The mean rating for all display conditions was 'disagree' (Fig. 49). There were 4 ratings labeled by the asterisks as "outliers" had ratings of 'slightly agree.' In this case, the pilots obviously experienced slight discomfort and headaches when using the HWD.

  EP Comments: EPs who rated 'slightly agree' for headaches commented that the turbulence caused eye strain which contributed to slight headaches. One EP commented that the support nose pieced on the HWD created a "hot spot" over time and contributed to a slight headache.

- *C: The HWD was comfortable to wear.* For the qualitative assessment of comfort, the ratings spanned the entire range of the scale. Although the data tended toward the 'comfortable' side, the median rating was neutral. The HWD was found on occasion to be comfortable and for others, they strongly disagreed and found the HWD uncomfortable in that the HWD created a "hot spot" on the nose caused by the added weight of the head tracker mounted to the HWD (Fig. 50).

- *D: The HWD symbols were easy to read (terms of clarity).* The qualitative ratings for symbol clarity indicated generally excellent HWD optical performance. The ratings spanned from 'slightly agree' to 'strongly agree' with the mean rating between 'agree' and 'strongly agree.' One negative response occurred related to the earlier comment that bumps and latency caused the eyes to refocus and thus affected the readability on the initial data run (Fig. 51).

- *E: The HWD video/imagery was easy to read (terms of clarity).* The qualitative ratings for imagery clarity also indicated generally excellent HWD optical performance. The ratings spanned from 'slightly agree' to 'strongly agree' with an average rating of 'agree' for all FLIR conditions. The negative responses were from EPs who commented that turbulence was the main cause in the readability of the imagery (Fig. 52).

- *F: The HWD field-of-view was acceptable to perform task and operation.* The qualitative ratings for the field-of-view were acceptable with the ratings spanning from 'slightly agree' to 'strongly agree' with 1 rating (identified as an "outlier" in the figure) for each display condition. For the outliers on approach where the EPs disagreed that the HWD field-of-view was acceptable, the EPs experienced a large cross wind that caused the altitude symbology to obscure the runway. For the extreme surface operations so-called "outlier" ratings, the EPs commented the horizontal FLIR field-of-view was too small for turning maneuvers as FLIR imagery was not viewable in turns (Fig. 53).

- *G: The HWD did NOT obscure or impair my ability to see traffic.* The qualitative ratings spanned from 'disagree' to 'strongly agree' with the means tending towards 'slightly agree.' One EP consistently rated the HWD as obscuring the outside world due to the nature of having a device in front of the eyes; however, the EP qualified that statement that such obscuration is not unique to the HWD used in this flight test (Fig. 54).

- *H: The HWD concept provided usable and sufficient visual cues to safely perform the task/operation.* The qualitative ratings spanned from 'slightly disagree' to 'strongly agree' with the means tending towards 'agree.' The ratings from 'strongly disagree' to 'neutral' were from an EP that felt the loss of peripheral vision and the FLIR imagery in the background made the symbology difficult to read (Fig. 55).

- *I: The HWD symbology and imagery was conformal (i.e., aligned and scaled) to the outside world.* The qualitative ratings spanned from 'slightly disagree' to 'strongly agree' with the mean tending towards 'slightly agree'. The EPs commented that their ratings from 'disagree' to 'neutral' were from turbulence or ground vibration causing a misalignment of the symbology and/or FLIR imagery (Fig. 56).

6.2.4 Paired Comparisons

The EPs were asked to provide pairwise ratings of situation awareness using the Situation Awareness Subjective Workload Dominance (SA-SWORD) [29] and "display equivalence" for flight guidance and SA between the HWD and both a HUD (based on past HUD experience since a HUD was not used in this flight test) and a head-down PFD. The paired comparison questionnaire is in Appendix F, section F.10. Geo-means were calculated based on the ratings and subsequent parametric statistics [28]. The result for the SA-SWORD was a significant main effect, $F(2, 12) = 29.377, p = 0.001$. Post-hoc within-subject contrasts revealed that pilots significantly rated the HWD as better for situation awareness (for approach operations) than both the HUD and baseline display standard. The HUD was also rated significantly better than the head-down display.

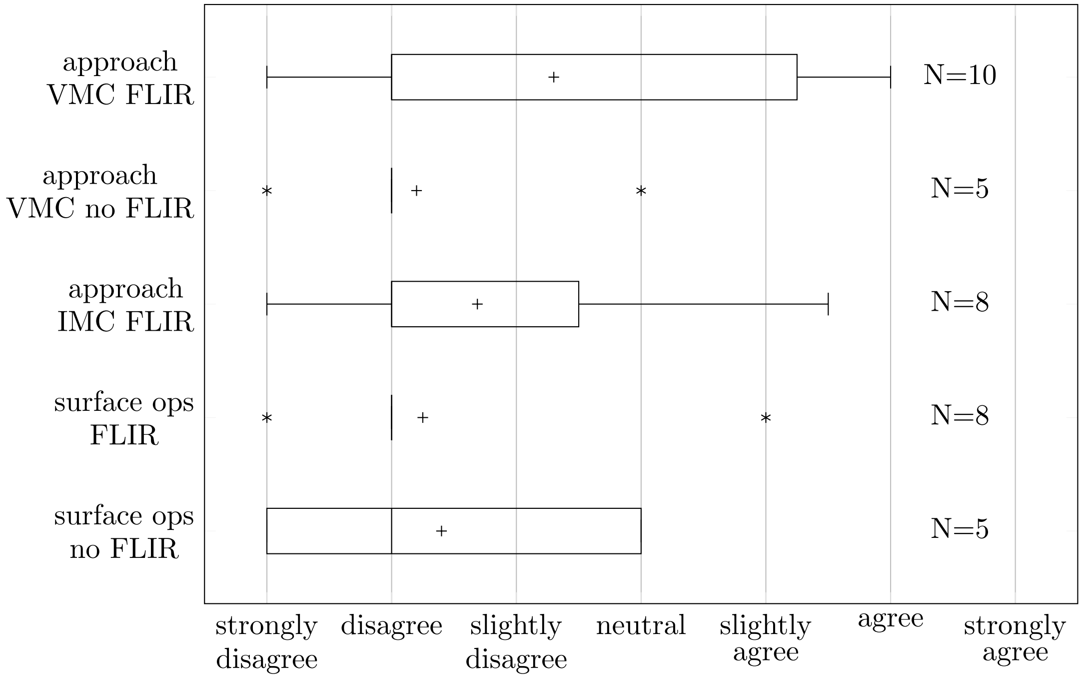

Figure 48. Post-run statement A: The HWD caused me to experience eye strain.

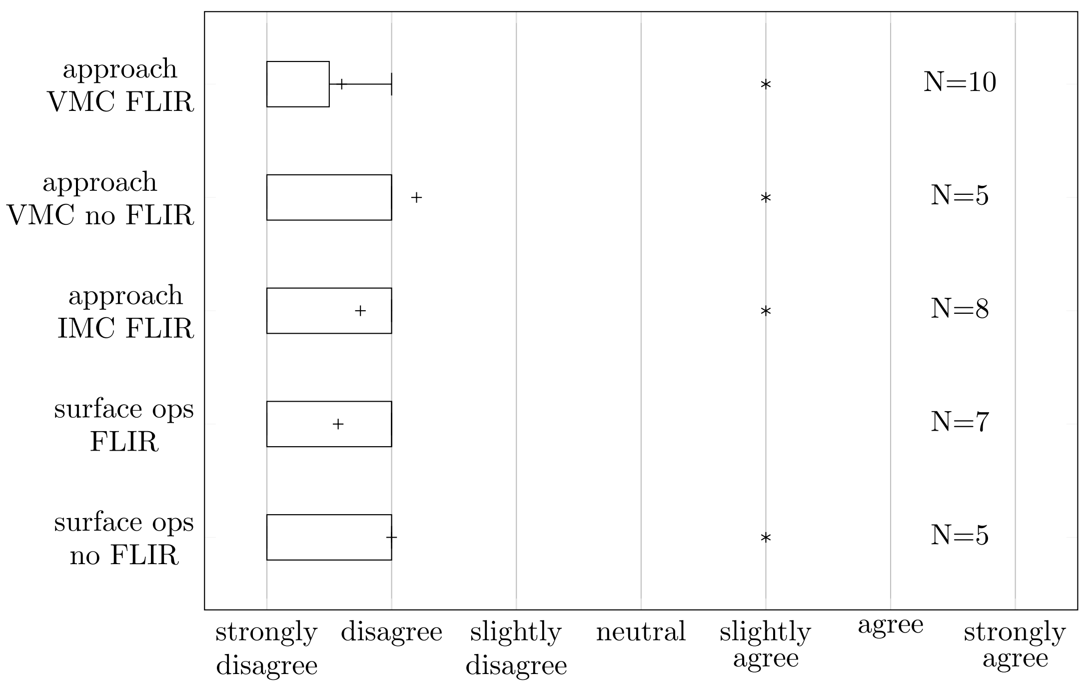

Figure 49. Post-run statement B: The HWD caused me to experience headaches.

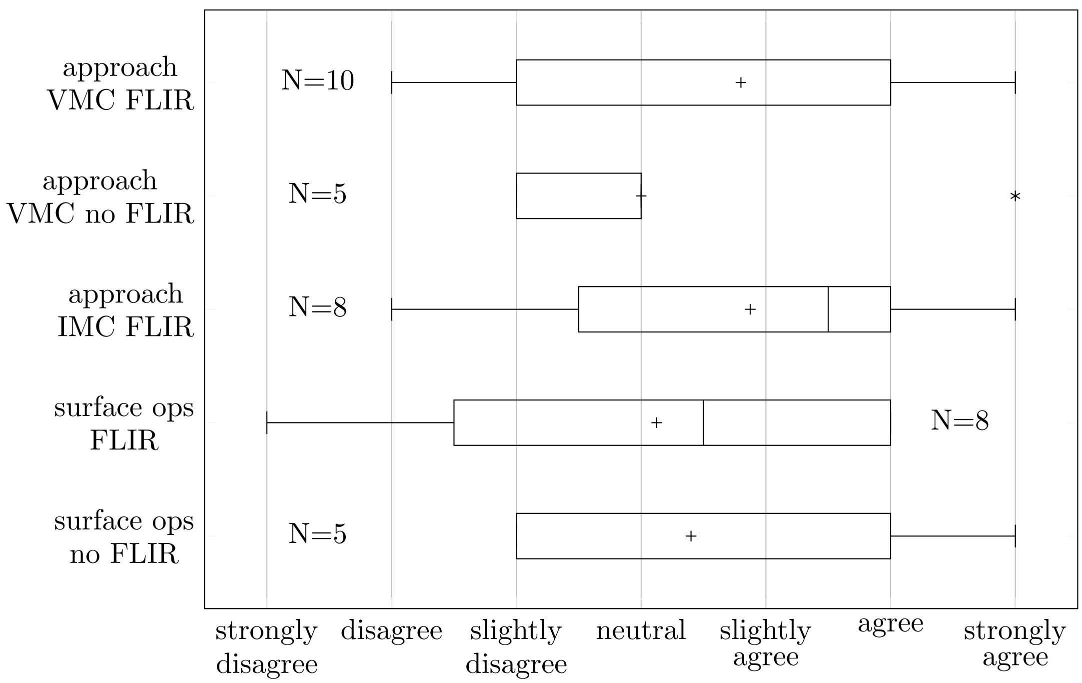

Figure 50. Post-run statement C: The HWD was comfortable to wear during the task/operation.

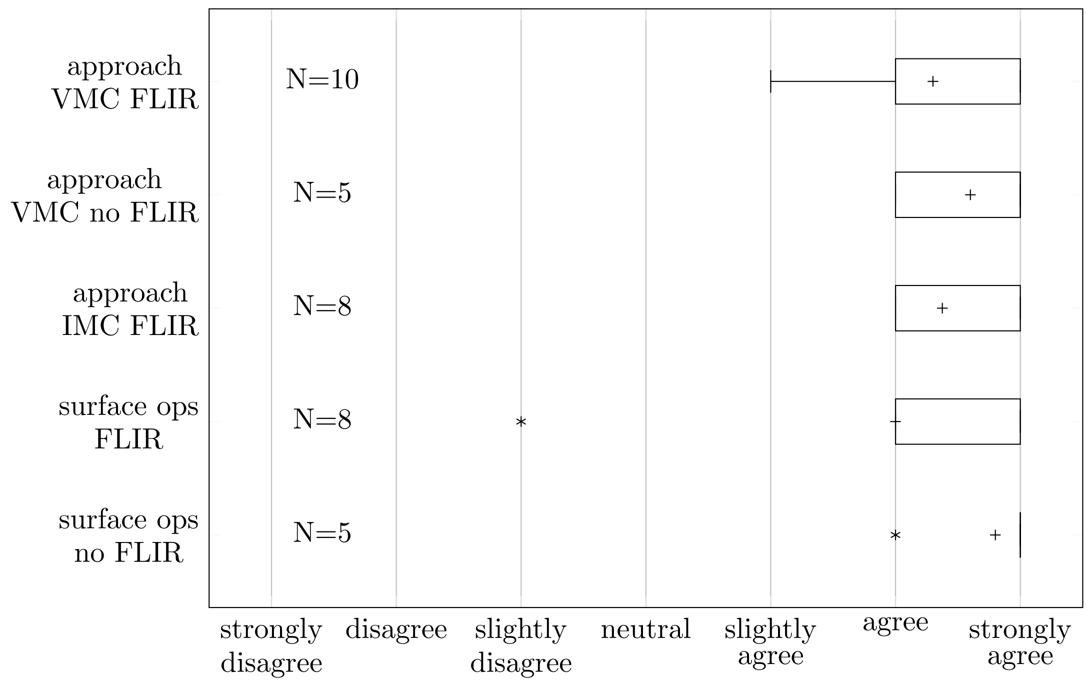

Figure 51. Post-run statement D: The HWD symbology was easy to read (in terms of clarity).

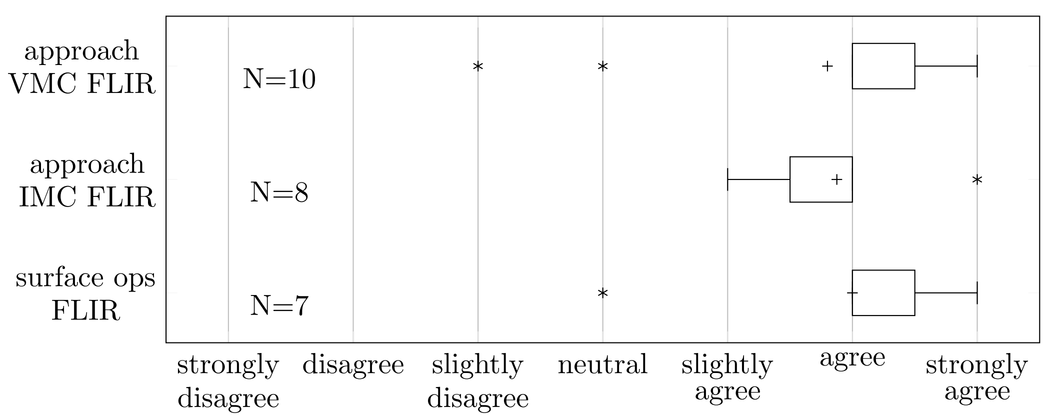

Figure 52. Post-run statement E: The HWD video/imagery was easy to read (in terms of clarity).

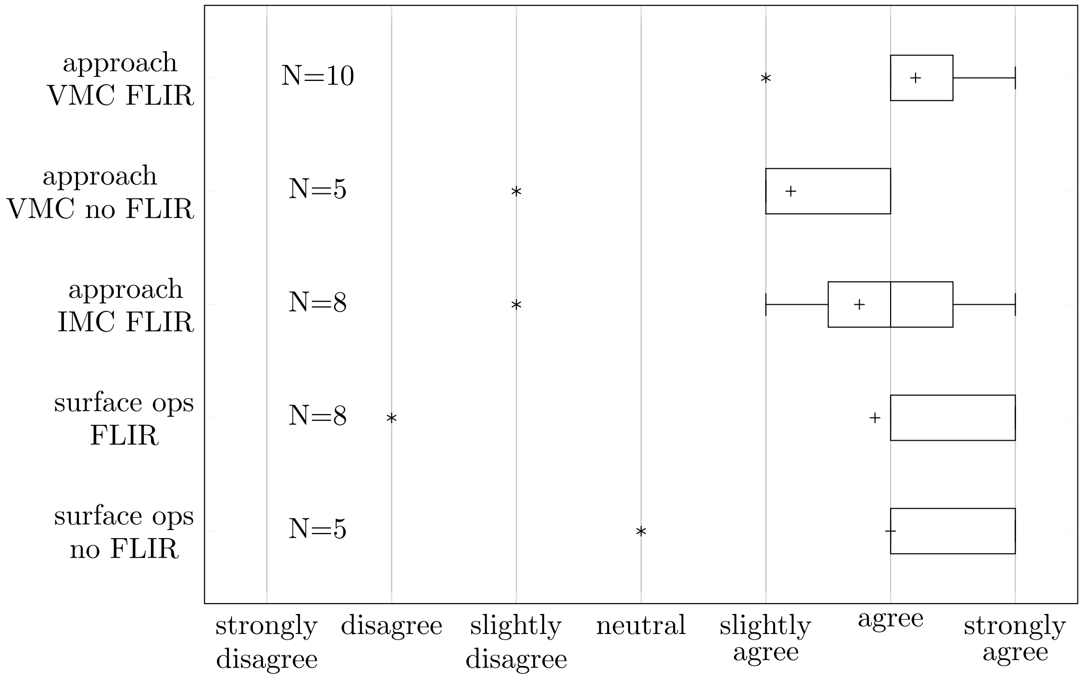

Figure 53. Post-run statement F: The HWD concept field-of-view was acceptable to perform the task and operation.

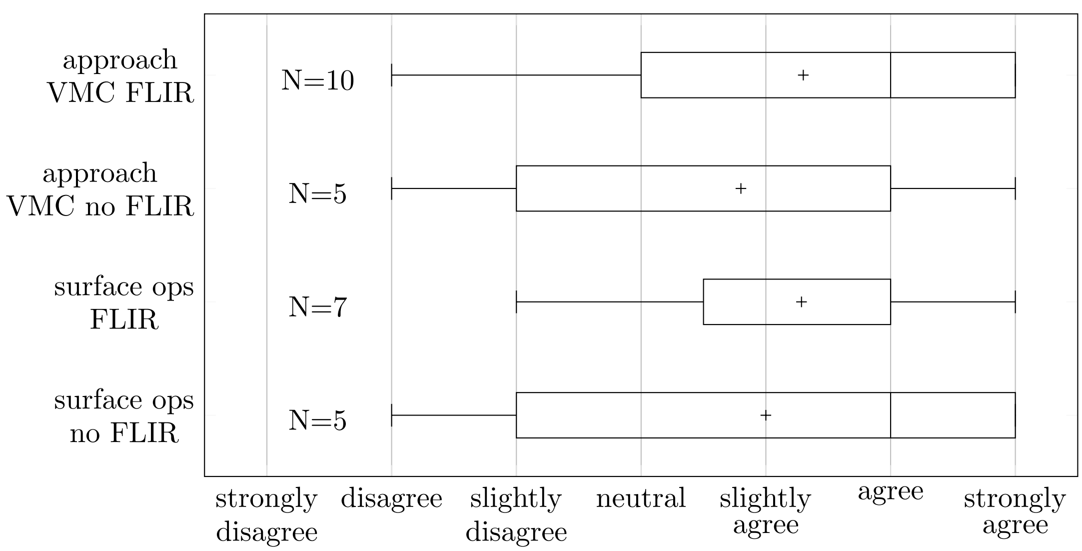

Figure 54. Post-run statement G: The HWD did NOT obscure or impair my ability to see traffic or other vehicles.

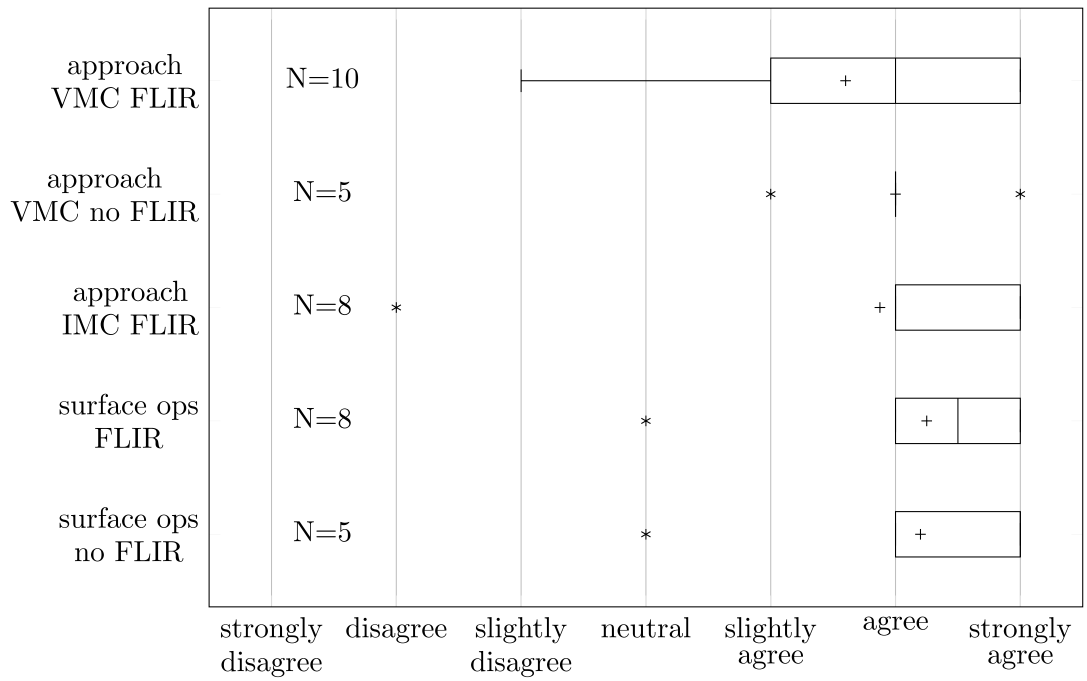

Figure 55. Post-run statement H: The HWD concept provided usable and sufficient cues to safely perform the task/operation.

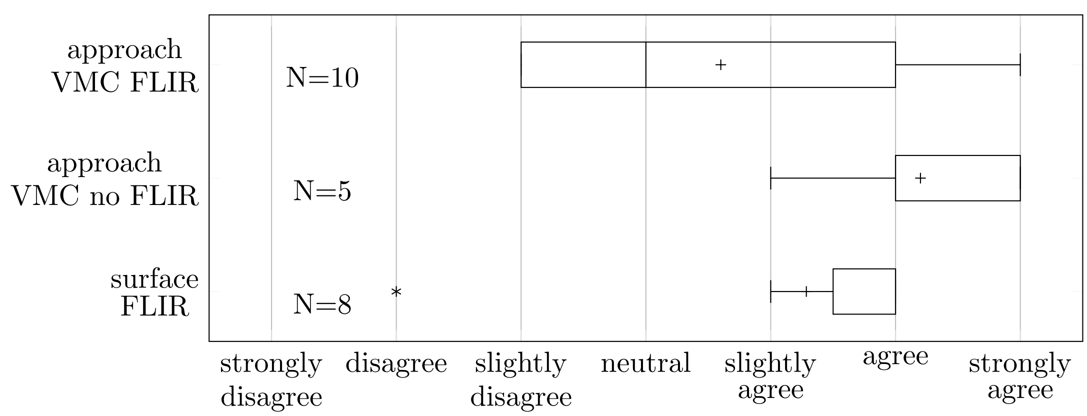

Figure 56. Post-run statement I: The HWD symbology and imagery was conformal (i.e., aligned and scaled) to the outside world.

6.2.5 Ground-Test Questionnaire

Upon boarding the test aircraft, the EPs donned the HWD and gave their initial impressions. The EPs were given the questionnaire shown in Appendix F, section F.7. The results of the EPs ratings are shown in Figs. 57 through 61.

The results of each ground questionnaire statement is as follows:

- *A: Please rate the overall comfort of the HWD.* The average rating for comfort was 'comfortable.' There were 2 ratings, indicated by asterisks as points which are outside of 1.5 times the IQR, with: 1 rating of 'uncomfortable' and 1 rating of 'extremely comfortable' (Fig. 57).

  EP Comments: Difficult to get on but comfortable. Comfortable. Heavier than sunglasses. Hardly notice wearing the glasses (HWD) except a small amount of pressure on the nose. (The HWD creates) pressure on the bridge of the nose.

- *B: Please rate the clarity of the symbology.* The average rating for clarity was 'extremely readable' with no score below 'readable' (Fig. 58).

  EP Comments: Current heading is positioned somewhat high on the display. When the HWD is pitched down on the head, (there is) better clarity of the heading symbology. The HWD has some ghosting. Fonts could be less bold.

- *C: There was a discernible color shift of the external world due to the HWD.* Ratings by the EPs ranged from 'disagree' to 'agree' with an average of 'neither agree nor disagree' (Fig. 59).

  EP Comments: Slight color shift is almost imperceptible. Color shift less than on a normal HUD.

- *D: There was a discernible distortion of the external world due to the HWD.* Ratings by the EPs ranged from 'disagree' to 'agree' with an average between 'slightly disagree' and 'neither agree nor disagree.' (Fig. 60).

  EP Comments: No issues. When pitching head up and down there is a small amount of jitter. There is "ratcheting" in the image.

- *E: I experienced significant glare while using the HWD.* All but 1 pilot disagreed with the statement that the HWD caused significant glare. There was 1 rating that agreed that the HWD caused glare (Fig. 61).

  EP Comments: No glare issues. Small amount of glare near the bottom of the display, mostly out of the field-of-view.

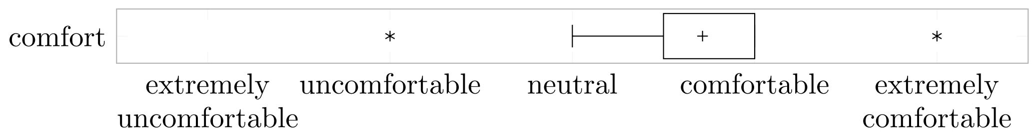

Figure 57. Ground-test statement A: EPs ratings of overall comfort of the HWD before flying. $N = 7$

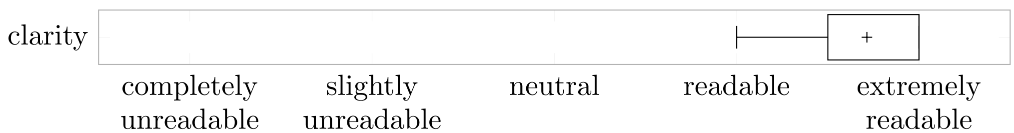

Figure 58. Ground-test statement B: EPs rating of the clarity of the symbology before flying. $N = 7$

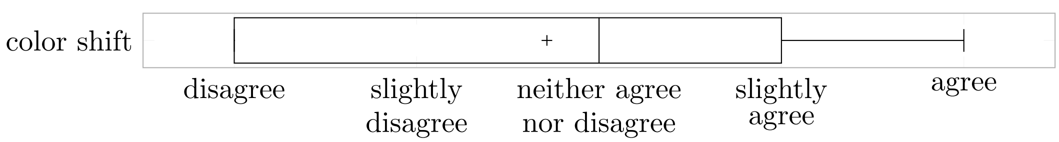

Figure 59. Ground-test statement C: EPs rating of the color shift due to the HWD before flying. $N = 7$

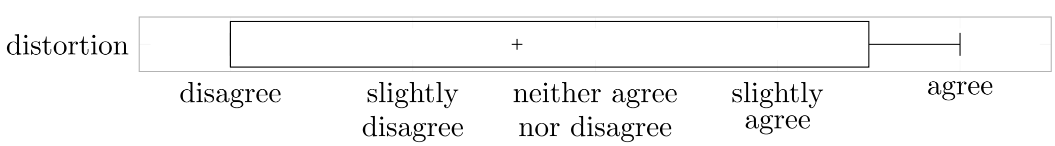

Figure 60. Ground-test statement D: EPs rating of the distortion due to the HWD before flying. $N = 7$

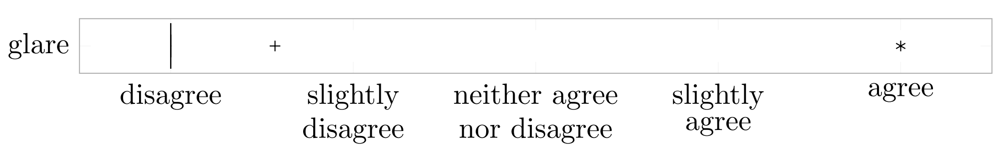

Figure 61. Ground-test statement E: EPs rating of the glare using the HWD before flying. $N = 7$

6.2.6 Post-Test Questionnaire

Immediately following the flight, a questionnaire was administered to the EPs. The results of the EPs ratings are shown in Figs. 62 through 69 and discussed below.

- *F: Please rate the overall comfort of the HWD.* The comfort ratings spanned from 'uncomfortable' to 'comfortable' with an average rating of 'neutral' (Fig. 62).

  EP Comments: The HWD was fine for most of the flight but became uncomfortable at the end. The nose piece felt "snug" towards the end of the flight. No worse than sunglasses when I fly. The head band has to be snugged to hold the HWD (stable) which created a "hot spot" on the nose. I had to tilt the HWD forward on the face to see full symbology. I had to rotate my head down to see the localizer scale. After an hour, the only "hot spots" noticed were the bridge of the nose and just behind the ears. Comfortable for about 30 minutes, then the bridge of the nose started to hurt. Pressure on the nose after 30 minutes that was persistent but could be dismissed when consumed with a task.

- *G: Please rate the effectiveness of the bore sighting procedures.* EPs rated the effectiveness of the boresighting procedure between 'difficult' and 'extremely easy' with the mean rating of 'easy' (Fig. 63).

  EP Comments: First time was difficult to acquire the (boresight) target, but was able to accomplish the task in subsequent boresights by closing one eye. Boresight seemed to have varying success rate. Hard to hold alignment during boresight procedure. Holographic (boresight target) hard to find and jitter in the system made it hard to calibrate. Easy but vibrations/turbulence made it hard to be accurate. Bouncing image was a little difficult to boresight. No issues.

- *H: Please rate the effectiveness of the brightness controls.* EPs rated the ease of use of the brightness controls between 'easy' and 'extremely easy' with 1 outlier rating of 'neutral' as indicated by the asterisk (Fig. 63).

  EP Comments: I didn't use them. I didn't evaluate them. The rotary switches were intuitive; however, the location required the pilot to remove hand from the throttles and visually search for the knobs.

- *I: The HWD's vertical field-of-view was sufficient to provide flight guidance information throughout the intended operational envelope.* EPs ratings spanned the entire range of responses with a mean rating of 'slightly agree' (Fig. 64).

  EP Comments: I had to slightly depress my viewing angle to see the localizer scale. I constantly had to look low for localizer. Adequate but would have liked more. I had to tilt my head down a few degrees to see the localizer guidance on the instrument approaches. In level flight, the field-of-view was fine; however, I needed to tilt the glasses to see the localizer guidance. Likewise, on approach, the glasses needed to be tilted down on glideslope to see the localizer.

- *J: There was a discernible color shift of the external world due to the HWD.* All EPs rated HWD causing a color shift as 'disagree.' (Fig. 65)

  EP Comments: Color shift analogous to wearing sunglasses. Didn't really notice a color shift.

- *K: There was a discernible distortion of the external world due to the HWD.* The EPs rated HWD causing a distortion between 'disagree' and 'slightly agree' with a mean rating of 'slightly disagree' (Fig. 65).

  EP Comments: Imagery (FLIR) was not affected by quick head movements. Quick head movements did affect alignment between symbology/imagery and outside world. When it (HWD symbology and imagery) doesn't match (the world) exactly, it becomes distracting.

- *L: The HWD's lateral field-of-view was sufficient to provide flight guidance information throughout the intended operational envelope.* EPs rated HWD causing a distortion between 'disagree' and 'slightly agree' with a mean rating of 'slightly disagree' (Fig. 64).

  EP Comments: I would like it to be slightly larger. I would like to have attitude information as well as airspeed and altitude when off-boresight.

- *M: The simulated FLIR was helpful in increasing situation awareness during tasks.* EPs rated helpfulness of the FLIR between 'slightly agree' and 'agree' with a mean rating of 'agree' (Fig. 66).

  EP Comments: It was useful for Night/IMC but would not be for day VMC. Especially (useful) in low visibility and I would like to have declutter options. Depends on the task: in day VMC, it (simulated FLIR) was distracting; but, it was awesome for IMC. Especially (useful) for the IMC data runs.

- *N: The HWD system caused me to experience eye strain.* The EPs rated the HWD casuing eye strain the entire range between 'disagree' and 'agree' with a mean rating of 'neutral' (Fig. 67).

  EP Comments: At the end of the flight, I noticed a very slight amount of eye strain but no more than I experienced using a helmet mounted system. The latency of the HWD system greatly added to the "work" to keep the symbology focused. I experienced slight eye strain that I attribute to not wearing glasses and trying to focus on cockpit instrumentation and reading questionnaires. I experienced eye strain close-in with turbulence. I blink less so eyes get dry (contacts) so they get sore. I experienced eye strain towards the end of the flight. It became more noticeable probably due to mismatch between the real world and the HWD symbology and imagery.

- *O: The HWD system caused me to experience headaches.* The EPs rated the HWD casuing headaches between 'disagree' and 'slightly agree' with a mean rating of 'slightly disagree' (Fig. 67).

  EP Comments: Due to the pressure of the nose clip on the bridge of the nose. Because of the pressure of the nose clip over time.

- *P: Please rate your overall perceived safety of the HWD.* All EPs rated the perceived safety of the HWD as 'safe' (Fig. 68).

  EP Comments: I believe my experience with HUDs and FLIR makes using this device very safe. Pilots without similar experience might require some training. I would rate the HWD (used in this flight test) "extremely safe" if the HWD did not block the peripheral vision of the pilot. I believe you could train (pilots) to the differences with a HUD. Improve the ergonomics and stabilize the symbology in turbulence. Lateral field-of-view is restricted. Traffic symbology was enhancing. Continuous (off-boresight) out the window viewing was enhancing.

- *Q: Please rate your overall perceived safety of the HWD as compared with current production HUDs.* The EPs rated the perceived safety of the HWD compared to a HUD between 'neutral' and 'extremely safe' with a mean rating of 'safe' (Fig. 68).

  EP Comments: I liked the traffic symbology and extended runway centerline. The off-boresight info was good. Depends on whether the HUD has FLIR. I prefer your HWD with FLIR over just a HUD but prefer a HUD with FLIR over the HWD. Detractors were the cable attachment to the HWD and the lateral field-of-view. Without proper training, inexperienced pilots would be distracted by "the show." However, with proper training and exposure, situational awareness would be higher.

- *R: Based on your overall experience with HUDs, would you consider this HWD equivalent to a HUD.* The EPs rated the HUD equivalence between 'slightly disagree' and 'agree' with a mean rating of 'slightly agree' (Fig. 69).

  EP Comments: I prefer a fixed HUD as a PFD; however, there is value to the HWD for aircraft without HUDs. Improve the latency and fix the (blocked) peripheral vision and this (HWD) would be a great option as a HUD. Information was good. Tough to use in VMC/VFR in turbulence. Remove the jitter from the HWD. Reduce the latency of the HWD.

- *S: What improvements do you feel could be made to the HWD for better comfort, performance, etc.*

  - Improve the comfort and need more vertical field-of-view.
  - Reduce the jitter (latency) of the display.
  - For a certified HWD, the HWD can not obscure the peripheral field-of-view.
  - I would like declutter options especially during flare.

6.2.7 HUD Equivalence

The EPs were asked to perform a post-experiment paired comparison on the equivalency of the HWD to the HUD based on their experience. Pilots rated the HWD as superior (i.e., equivalent to, if not better than) to the HUD for flight guidance.

To supplement the paired comparison, pilots were also asked to provide a rating of agreement to the following statement, "Based on my overall experience with HUDs, I would consider this HWD equivalent to a HUD" using a 1 to 5 Likert scale (1-disagree to 5-agree). The EPs scores ranged from 2 to 5 with a mean score of 3.7 and a standard deviation of 1.15 (Fig. 69). It is important to note that the flight test did not employ a HUD and, therefore, these ratings are based on the pilots' extensive experience with HUD use. Pilots commented that the HWD, in many ways, had characteristics that provide significant advantages to the HUD.

The EPs had many favorable comments with the unlimited-field-of-regard capability of the HWD including that they liked traffic diamonds. These added symbologies (traffic diamonds, extended centerline and the 360° horizon with heading tick marks) increased SA at minimal expense of the lateral field-of-view. Pilots commented that the traffic diamonds displayed in cyan greatly reduced the clutter of the display. As color is not used on a HUD, the color on the HWD permitted the EPs to quickly disregard the cyan traffic diamond if it was not pertinent to the task. In other words, they could ignore or see the traffic diamond intuitively. In addition, the unlimited field-of-regard allowed pilots to quickly locate traffic when identified by ATC. For example, when ATC called traffic at 10 o'clock low, pilots could look to that position and quickly acquire the traffic highlighted by the cyan traffic diamond, something that cannot be done with a fixed HUD.

6.2.8 HWD Ergonomics

Encumbrance of the HWD is a concern for commercial crews. As the goal is to obtain a sunglasses form factor, this HWD system represented the best analog to that form factor to date. In general, pilots commented that the overall comfort level was acceptable; however, if the test lasted any longer, they felt the HWD would quickly become painful. The HWD system consisted of a strap to keep the HWD stable on the pilots head; however, the tightness of the strap created "hot spots" over time, mainly on the bridge of the nose. Therefore, the HWD system weighing 7 oz may be too heavy for wearing for more than an hour duration. Reducing the weight of the HWD system continues to be a goal for NASA research.

Several technical challenges were identified with the HWD tested. The absence of peripheral cues in crosswind conditions is a hindrance. This is due to the tested HWDs specific design and should not be a problem with a design tailored to the aviation domain. System latency must be minimized to reduced imagery jitter. Bore-sight positioning must be optimized and a better technique is desirable. The traffic diamonds on the HWD demonstrated the de-cluttering effect of the display but further research would be needed to determine if other symbology could also benefit. Another unknown is how colors will appear at different contrast levels when flying in various lighting and weather conditions. (The performance of monochrome green is well known.)

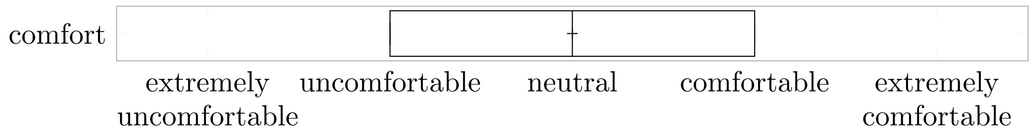

Figure 62. Post-test statement F: EPs ratings of overall comfort of the HWD.

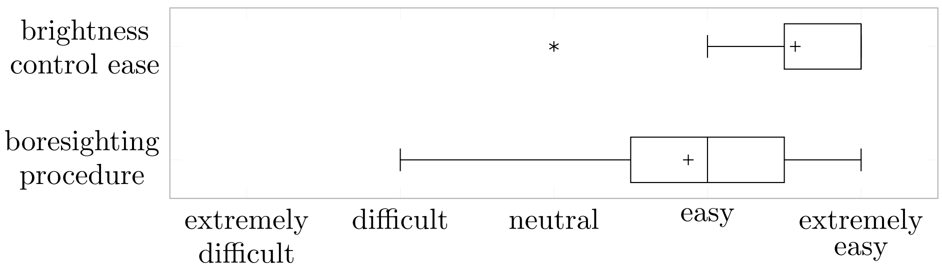

Figure 63. Post-test statements G and H: EPs ratings of boresight effectiveness and the ease of use of the brightness controls.

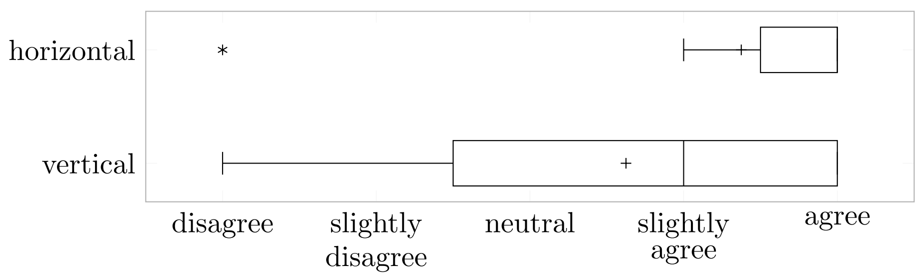

Figure 64. Post-test statement I and L: EPs ratings that the field-of-view of the HWD was sufficient to provide flight guidance information throughout the intended operational envelope.

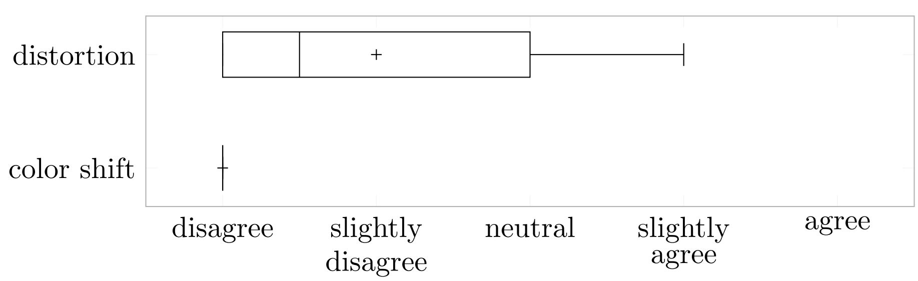

Figure 65. Post-test statement J and K: EPs ratings that the HWD caused a color shift and distortions to the outside world.

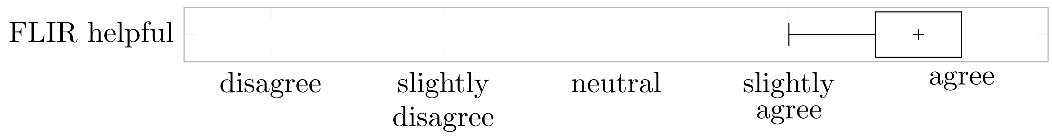

Figure 66. Post-test statement M: EPs ratings of the simulated FLIR was helpful in increasing situation awareness during tasks.

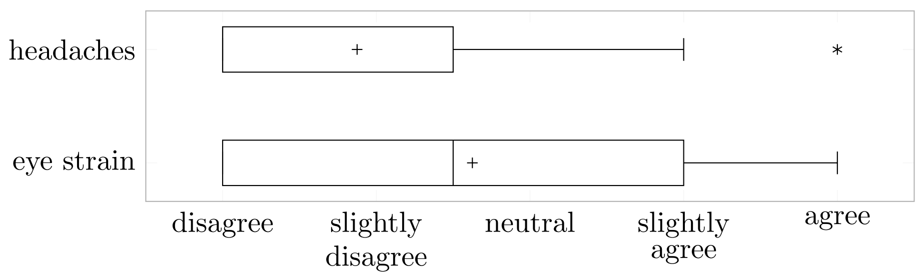

Figure 67. Post-test statement N and O: EPs ratings that the HWD caused eye strain and headaches.

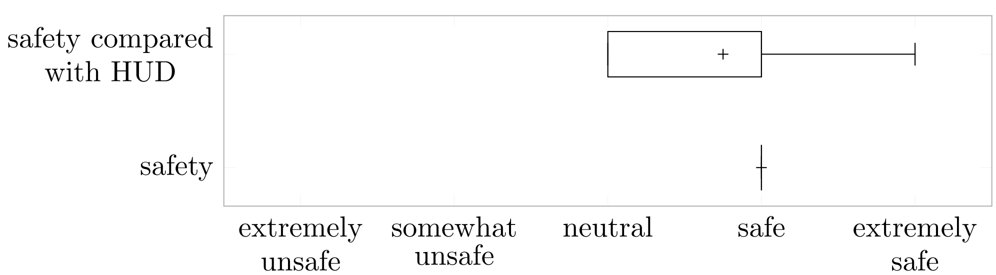

Figure 68. Post-test statement P and Q: EPs ratings of the perceived safety of the HWD and the safety of the HWD as compared with current production HUDs.

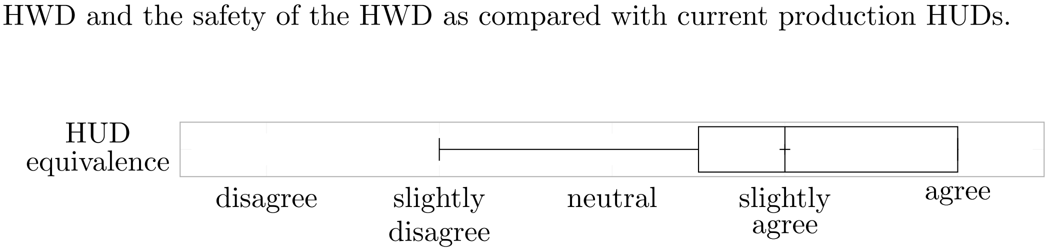

Figure 69. Post-test statement R: EPs ratings that the HWD is equivalent to a HUD.

7 Conclusions

The goal of the research presented in this paper was to investigate operational equivalence between a head-tracked HWD system and a HUD. The results from the simulator and the flight test showed that there were no statistical differences in the crews' performance in terms of approach, surface operations, departure, and off-nominal scenarios. Further, there were no statistically significant differences between the HWD and HUD in pilots' responses to questionnaires for either the simulator or flight test trials. Qualitative results show the evaluation pilots view the prototype HWD as equivalent to a HUD; however, the technology tested had issues that compromised its usability, especially in turbulent conditions. Although these results showed that there were no statistical differences in the crews' performance, there are still technical hurdles to be overcome for complete display equivalence including, most notably, the end-to-end latency of the HWD system. Also, comfort, boresighting, brightness and declutter controls, field-of-view, and HWD weight factors must be considered for future certified HWD systems.

Future Research

NASA will continue to conduct HWD research experiments in two focus areas. The first focus area is exploring the individual characteristics needed to make the HWD viable in commercial operations. Examples include total system latency requirements, performance in turbulent conditions, symbology optimization, ergonomics, and boresighting. The second focus area will be to expand the types of operational environments used in research experiments. The majority of the HWD research has occurred during low visibility surface operations, straight-in approaches and departures. Future operations will include complex curved approaches in low visibility conditions and traffic conflict scenarios in the en route environment.

References

1. R.E. Bailey, K.J. Shelton, and J.J. Arthur III. Head-worn displays for NextGen. In *Enhanced and Synthetic Vision 2011*, volume 8041, Bellingham, WA, April 2011. SPIE.

2. J.J. Arthur III, L.J. Prinzel III, S.P. Williams, and L.J. Kramer. Synthetic vision enhanced surface operations and flight procedures rehearsal tool. In Jacques G. Verly and Jeff J. Guell, editors, *Enhanced and Synthetic Vision Proceedings of SPIE*, volume 6226, page 62260I, Bellingham, WA, 2006. SPIE.

3. J.J. Arthur III, L.J. Prinzel III, K.J. Shelton, R.E. Bailey, L.J. Kramer, S.P. Williams, and R.M. Norman. Design and testing of an unlimited field-of-regard synthetic vision head-worn display for commercial aircraft surface operations. In Jacques G. Verly and Jeff J. Guell, editors, *Enhanced and Synthetic Vision 2007*, volume 6559, Bellingham, WA, April 2007. SPIE.

4. J. J. Arthur III, L.J. Prinzel III, R.E. Bailey, K.J. Shelton, S.P. Williams, L.J. Kramer, and R.M. Norman. Head-worn display concepts for surface operations for commercial aircraft. Technical Report NASA/TP-2008-215321, NASA Langley Research Center, Hampton, VA, June 2008.

5. W.B. Albery. Multisensory cueing for enhancing orientation information during flight. *Aviation, Space, and Environmental Medicine*, 78(5):186–190, 2007.

6. T.W. Frey and H.J. Page. Virtual HUD using an HMD. In Ronald J. Lewandowski, Loran A. Haworth, Henry J. Girolamo, and Clarence E. Rash, editors, *Helmet- and Head-Mounted Displays VI*, volume 4361, Bellingham, WA, April 2001. SPIE.

7. R. E. Bailey. Head-up display (HUD) lessons learned for helmet-mounted display (HMD) development. In R. J. Lewandowski, W. Stephens, and L. A. Haworth, editors, *Helmet- and Head-Mounted Displays and Symbology Design Requirements*, volume 2218, Bellingham, WA, June 1994. SPIE.

8. FAA. Advisory circular: Enhanced flight vision systems. Technical Report 90-106, FAA, June 2010.

9. F. Cupero, B. Valimont, J. Wise, C. Best, and B. De Mers. Head worn display system for equivalent visual operations. Technical Report NASA/CR-2009-215781, NASA, Hampton, VA, Jul 2009.

10. FSF. Head-up guidance system technology - a clear path to increasing flight safety. Technical report, Flight Safety Foundation, Alexandria, Virginia, USA, November 2009.

11. S. Barber, D. Schwab, and K. Zimmerman. Head up and eyes out enabling equivalent visual operations with the head up display. *SAE International Journal of Aerospace*, 6(1):237–246, Sept 2013.

12. FAA. Criteria for Approval of Category III Weather minima for Takeoff, Landing and Rollout. Advisory Circular 120-28D, US Department of Transportation, Federal Aviation Administration, Washington DC, USA, July 1999.

13. FAA. Procedures for the Evaluation and Approval of Facilities for Special Authorization Category I Operations and All Category II and III Operations. Order 8400.13D, US Department of Transportation, Federal Aviation Administration, Washington DC, USA, October 2009.

14. R. E. Bailey, L. J. Kramer, and S. P. Williams. Enhanced vision for all-weather operations under NextGen. In Jeff J. Güell and Kenneth L. Bernier, editors, *Enhanced and Synthetic Vision 2010*, volume 7689, Bellingham, WA, April 2010. SPIE.

15. C.E. Rash, M.B. Russo, T.R. Letowski, and E.T. Schmeisser. *Helmet-Mounted Displays: Sensation, Perception and Cognition Issues*. U.S. Army Aeromedical Research Laboratory, 2009. ISBN 978-0-615-28375-3.

16. J.J. Arthur III, S.P. Williams, L.P. Prinzel III, L.J. Kramer, and R.E. Bailey. Flight simulator evaluation of display media devices for synthetic vision concepts. In Clarence E. Rash and Colin E. Reese, editors, *Helmet- and Head-Mounted Displays IX: Technologies and Applications*, volume 5442, pages 213–224, Bellingham, WA, 2004. SPIE.

17. D. C. Foyle and B. L. Hooey. Improving evaluation and system design through the use of off-nominal testing: A methodology for scenario development. In *12th International Symposium on Aviation Psychology*, pages 397–402, Dayton, OH, 2003.

18. RTCA. Minimum aviation system performance standards (MASPS) for enhanced vision systems synthetic vision systems, combined vision systems and enhanced flight vision systems. Technical Report DO-315, RTCA, Inc., December 2008.

19. R. E. Bailey, J. J. Arthur III, and S.P. Williams. Latency requirements for head-worn display S/EVS applications. In J. G. Verly, editor, *Enhanced and Synthetic Vision 2004*, volume 5424, pages 98–109, Bellingham, WA, April 2004. SPIE.

20. W. T. Nelson, R. S. Bolia, M. M. Roe, and R. M. Morley. Assessing simulator sickness in a see-through HMD: Effects of time delay, time on task, and task complexity. In *IMAGE 2000 Conference*, Scottsdale, AR, July 2000.

21. R. M. Taylor. Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations*, number 478 in AGARD Conference Proceedings, pages 3–1 – 3–37, Aerospace Medical Panel Symposium, Copenhagen, October 1990.

22. Lawrence L. Ames and Edward J. George. Revision and verification of a seven-point workload estimate scale. Technical Report Technical information manual Jan-Jun 92, Air Force Flight Test Center Edwards AFB, July 1993.

23. S. G. Hart and L. E. Staveland. Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human mental workload*, pages 139–183, Amsterdam, The Netherlands, 1988. Elsevier.

24. R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, pages 203–220, 1993.

25. L.J. Kramer, R.E. Bailey, K.E. Ellis, S.P. Williams, J. J. Arthur III, L.J. Prinzel III, and K.J. Shelton. Enhanced flight vision systems and synthetic vision systems for NextGen approach and landing operations. Technical Report NASA/TP-2013-218054, NASA Langley Research Center, Hampton, VA, November 2013.

26. L.J. Kramer, R.E. Bailey, K.E. Ellis, R.M. Norman, S.P. Williams, J.J. Arthur III, K.J. Shelton, and L.J. Prinzel III. Enhanced and synthetic vision for terminal maneuvering area nextgen operations. In Guell and Bernier, editors, *Display Technologies and Applications for Defense, Security, and Avionics V; and Enhanced and Synthetic Vision 2011*, volume 8042, Bellingham, WA, 2011. SPIE.

27. N.K. Link, R.V. Kruk, D. McKay, S. Jennings, and G. Craig. Hybrid enhanced and synthetic vision system architecture for rotorcraft operations. In *Enhanced and Synthetic Vision 2002*, volume 4713, pages 190–201. SPIE, 2002.

28. T. L. Saaty. *The analytical hierarchy process.* McGraw-Hill, New York, 1980.

29. M. A. Vidulich and E. R. Hughes. Testing a subjective metric of situation awareness. In *Human Factors Society 35th Annual Meeting*, pages 1307–1311, Santa Monica, CA, 1991. Human Factors Society.

30. Colonel Randall W Gibb, Dr Rob Gray, and Dr David M Regan. *Aviation Visual Perception: Research, Misperception and Mishaps.* Ashgate Publishing, Ltd., 2012.

31. L.J. Kramer, S.P. Williams, J.J. Arthur III, S.A. Rehfeld, and S.J. Harrison. Motion-base simulator evaluation of an aircraft using an external vision system. In *31st Digital Avionics Systems Conference*, Williamsburg, VA, 2012. DASC.

32. FAA. Advisory Circular: Criteria for Approval of Category I and Category II Weather Minima for Approach. Technical Report 120-29A, FAA, August 2002.

33. FAA. Instrument Rating Practical Test Standards. Flight Standards Service S-8081-4D, US Department of Transportation, Federal Aviation Administration, Washington DC, USA, April 2004.

Appendix A

Eye Tracking Analysis

Eye and head tracking data was collected for the Pilot Flying (PF) and Pilot Monitoring (PM) using the Smart Eye® eye-head tracking system installed in the simulator. Prior to data collection, each pilot had a profile created in order for the Smart Eye system to properly track the eyes and head. This profile was created without donning the HWD. Of particular interest in this experiment was the eye and head tracker comparative results for the PF between the Head-Up Display (HUD) and Head-Worn Display (HWD). There was a concern that the HWD might adversely affect the accuracy and reliability of the oculometer; therefore, to quantify the effects of the HWD on the eye tracking system, a stand-alone test was conducted.

For the stand-alone test, 3 test conditions were conducted: 1) eye/head tracking with no HWD; 2) wearing the HWD with head tracking InfraRed (IR) flashers powered off, and 3) wearing the HWD with head tracking IR flashers powered on. As the eye tracking system also uses IR flashers to illuminate markers for proper head-tracking, it was desired to ascertain the effects of the head tracking IR flashers on the eye tracking system. In all 3 stand-alone test cases, no symbology or imagery was displayed on the HWD.

The stand-alone test procedure was to look out-the-window for 5 seconds, look at the Primary Flight Display (PFD) for 2 seconds, look at the Navigational Display (ND) for 2 seconds and finally the Electronic Flight Bag (EFB) for 2 seconds. This head-up / head-down scan pattern was repeated 10 times for each test condition (no HWD, HWD with head tracking IR flashers powered off, and HWD with head tracking IR flashers powered on). Each scan was approximately 11 seconds which equates to 110 seconds (about 2 minutes) for each test condition.

Gaze quality is a confidence measurement reported by the oculometer system. From Fig. A1, it can be seen that the HWD caused the gaze quality to drop significantly. A gaze quality of 0 equates to poor eye tracking while a gaze quality of 0.7 or greater denotes good eye tracking. The number of bad gaze quality values (values of 0) increase from a count of 250 with the no HWD case to close to a count 5,000 when donning the HWD. This equates to about 70% of the eye tracking data being poor gaze quality when wearing the HWD. Without the HWD, only 3% of the data has poor gaze quality. Also, from Fig. A1, it can be seen that the power state of the HWD IR flashers have little effect on the gaze quality.

For each 60 Hz eye-head tracking measurement cycle, the eye-head tracking system reports a value of 1.0 for good quality head position/rotation values and a value of 0.0 for low confidence values. Figs. A2 and A3 show the effects of donning the HWD on the quality of the head position and the head rotation, respectively. Donning the HWD with the head tracking IR flashers powered on has little effect on the head position and rotation quality.

Fig. A4 shows the PFs head pitch values reported by the eye-head tracking system with and without the HWD. Each plateau in the graph represents the head-up (out the window) portion of the scan pattern conducted. From this time series

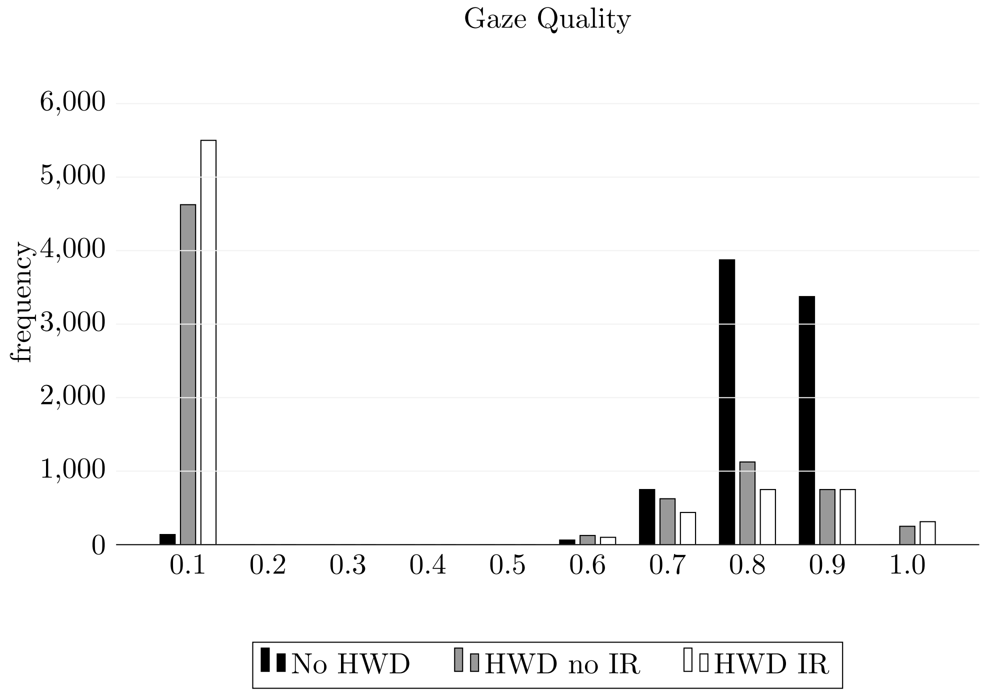

Figure A1. Histogram of Gaze Quality during stand-alone test.

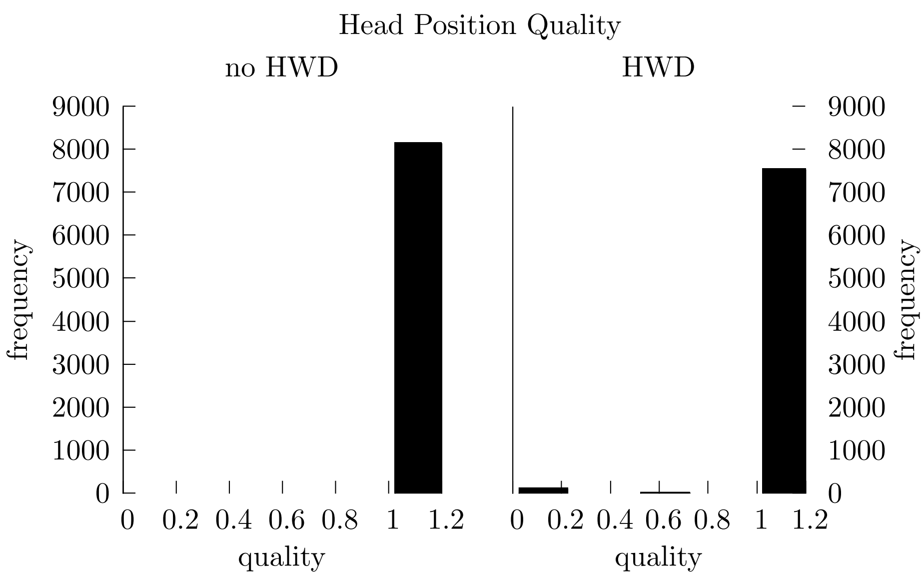

Figure A2. Head position quality with and without the HWD.

plot, it shows the head pitch values for the HWD shifted down from a range of 5° to 9°with an average of 6°. In other words, from this small test, the HWD causes the eye-tracking system to report a head pitch value lower than reported without the HWD.

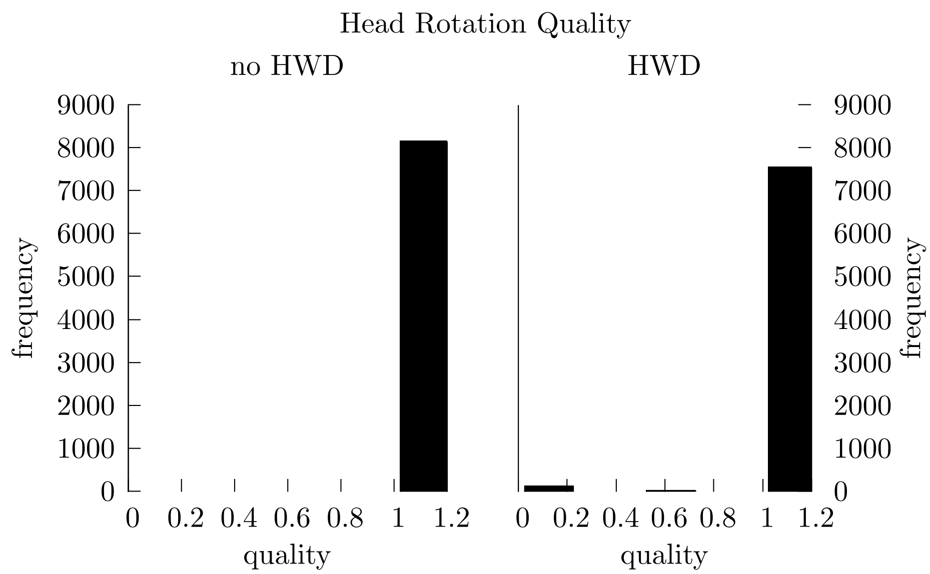

Figure A3. Head rotation quality with and without the HWD.

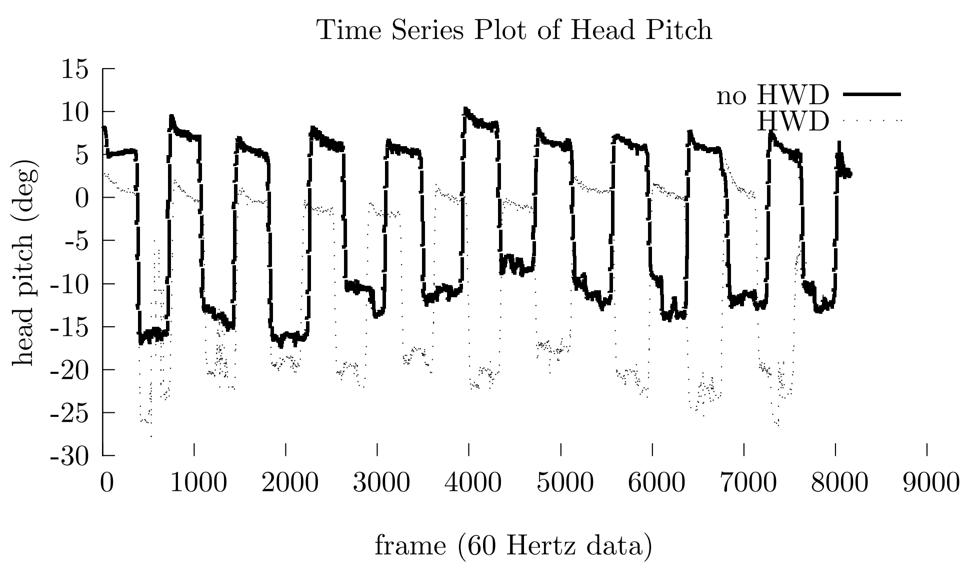

Figure A4. Head pitch values with and without the HWD.

Appendix B

Descriptive Statistics Tables

Table B1. Descriptive statistics for approach FTE measures.

	Concept	**Mean**	**Std. Dev.**	**N**
FTE Glideslope	HUD	0.1157	0.05965	24
	HWD-Virtual	0.1343	0.08993	24
	HWD-Split	0.1299	0.07123	24
	All	0.1266	0.07403	72
FTE Localizer	HUD	0.0221	0.01451	24
	HWD-Virtual	0.0212	0.01634	24
	HWD-Split	0.0256	0.02565	24
	All	0.0230	0.01927	72

Table B2. Descriptive statistics for approach RMSE Dependent Measures.

	Concept	**Mean**	**Std. Dev.**	**N**
RMSE Glideslope	HUD	0.1074	0.09273	28
	HWD-Virtual	0.1147	0.08992	28
	HWD-Split	0.0985	0.05355	28
	All	0.1069	0.08003	84
RMSE Localizer	HUD	0.0217	0.01440	28
	HWD-Virtual	0.0228	0.01681	28
	HWD-Split	0.0281	0.02793	28
	All	0.0242	0.02052	84
RMSE Sink Rate	HUD	1.4127	0.90845	28
	HWD-Virtual	1.7767	1.27035	28
	HWD-Split	1.6708	0.90232	28
	All	1.6201	1.04016	84

Table B3. Descriptive statistics for approach maximum value dependent measures.

	Concept	**Mean**	**Std. Dev.**	**N**
Max Glideslope	HUD	0.2590	0.20134	28
	HWD-Virtual	0.2881	0.24179	28
	HWD-Split	0.2496	0.14877	28
	All	0.2656	0.19919	84
Max Localizer	HUD	0.0384	0.02499	28
	HWD-Virtual	0.0420	0.03039	28
	HWD-Split	0.0501	0.04934	28
	All	0.0435	0.03633	84
Max Sink Rate	HUD	3.5214	2.22342	28
	HWD-Virtual	4.5540	3.29341	28
	HWD-Split	4.4174	2.31027	28
	All	4.1643	2.66259	84

Table B4. Descriptive statistics for equivalent visual RMSE performance.

	Concept	**Mean**	**Std. Dev.**	**N**
RMSE Glideslope	HUD	0.2911	0.14111	24
	HWD-Virtual	0.2405	0.18215	24
	HWD-Split	0.2995	0.23561	24
	All	0.2770	0.18939	72
RMSE Localizer	HUD	0.0178	0.01110	24
	HWD-Virtual	0.0144	0.00786	24
	HWD-Split	0.0205	0.01781	24
	All	0.0176	0.01300	72
RMSE Sink Rate	HUD	1.1143	0.52364	24
	HWD-Virtual	1.0212	0.42199	24
	HWD-Split	1.2089	0.66721	24
	All	1.1148	0.54468	72

Table B5. Descriptive statistics for equivalent visual maximum value dependent measures.

	Concept	**Mean**	**Std. Dev.**	**N**
Max Glideslope	HUD	0.6816	0.37139	24
	HWD-Virtual	0.5162	0.42584	24
	HWD-Split	0.5952	0.51257	24
	All	0.5977	0.43950	72
Max Localizer	HUD	0.0242	0.01409	24
	HWD-Virtual	0.0207	0.01144	24
	HWD-Split	0.0307	0.02491	24
	All	0.0252	0.01804	72
Max Sink Rate	HUD	2.1196	1.06652	24
	HWD-Virtual	1.8904	0.71219	24
	HWD-Split	2.1835	1.23202	24
	All	2.0645	1.02008	72

Table B6. Descriptive statistics for the threshold crossing deviation.

Threshold Crossing Deviation	**Concept**	**Mean**	**Std. Dev.**	**N**
Lateral	HUD	0.5980	3.10816	24
(feet)	HWD-Virtual	1.5294	3.94953	24
	HWD-Split	2.6050	6.95628	24
	All	1.5775	4.95380	72
Vertical	HUD	-20.2726	5.76390	24
(feet)	HWD-Virtual	-20.6392	4.39762	24
	HWD-Split	-20.8361	6.28950	24
	All	-20.5826	5.46779	72
Sink Rate	HUD	9.2585	1.82903	24
(ft/sec)	HWD-Virtual	9.3330	1.47824	24
	HWD-Split	9.0061	1.97851	24
	All	9.1992	1.75485	72

Table B7. Descriptive statistics for centerline tracking on rollout.

	Concept	**Mean**	**Std. Dev.**	**N**
RMSE Lateral Rollout	HUD	4.7559	2.69531	24
(feet)	HWD-Virtual	5.4243	2.36458	24
	HWD-Split	5.6671	3.47177	24
	All	5.2824	2.86700	72
Max Lateral Rollout	HUD	9.1344	4.63294	24
(feet)	HWD-Virtual	11.3425	5.08751	24
	HWD-Split	11.4679	7.23758	24
	All	10.6482	5.78545	72

Appendix C

Simulator Post-Run Statement Ratings

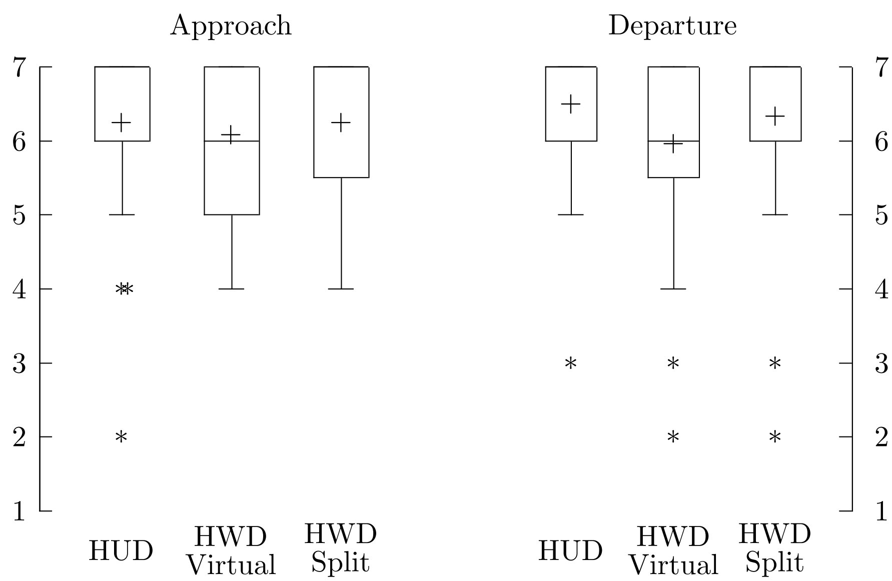

Figure C1. PF ratings for Statement A: I was aware of ownship position.

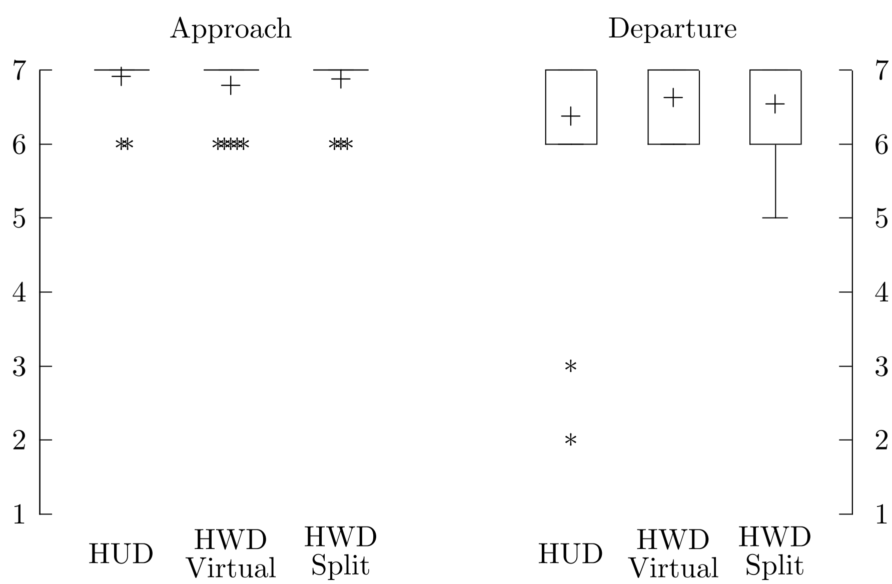

Figure C2. PM ratings for Statement A: I was aware of ownship position.

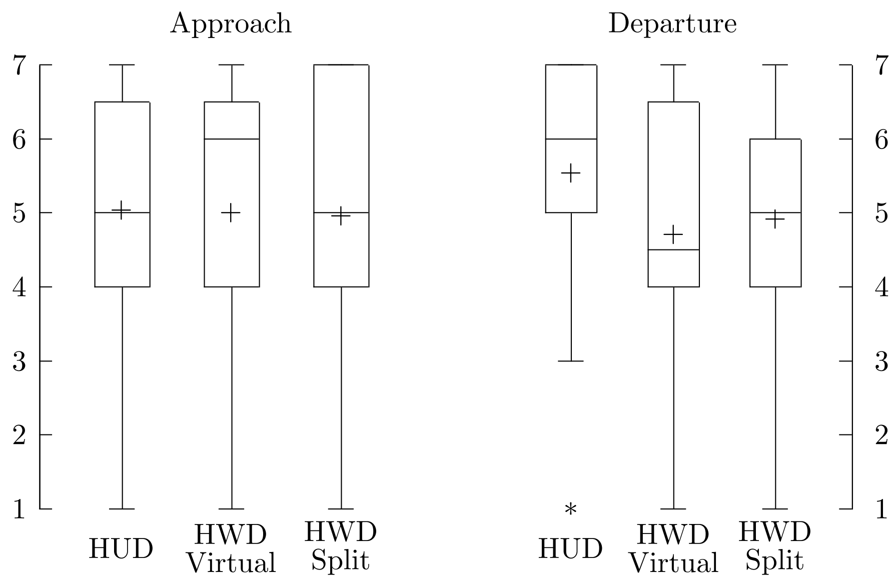

Figure C3. PF ratings for Statement B: I was aware of traffic and other vehicles during operations.

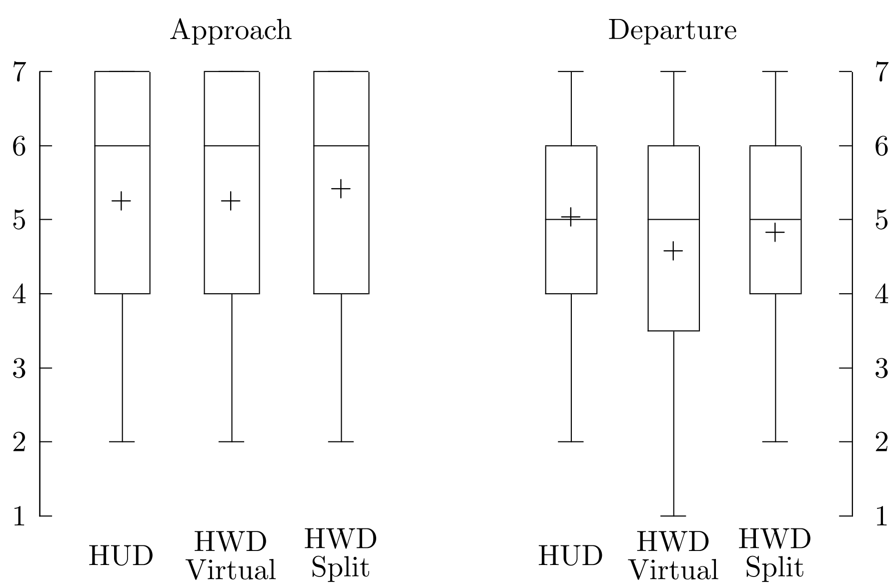

Figure C4. PM ratings for Statement B: I was aware of traffic and other vehicles during operations.

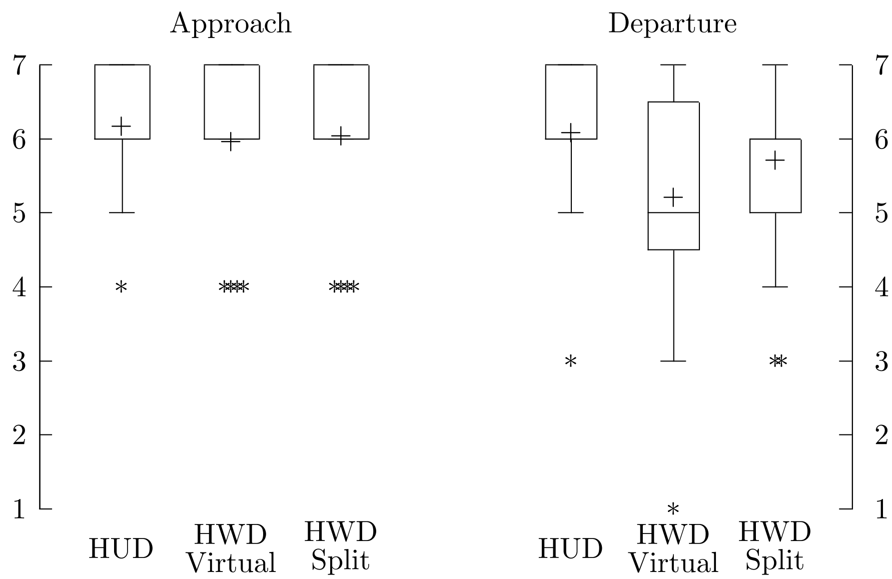

Figure C5. PF ratings for Statement C: The display concepts were effective for maintaining SA.

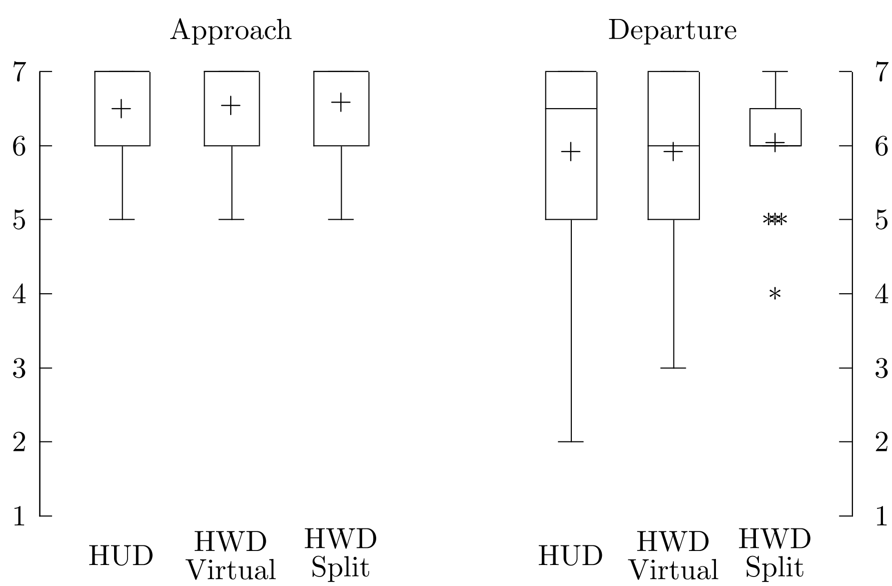

Figure C6. PM ratings for Statement C: The display concepts were effective for maintaining SA.

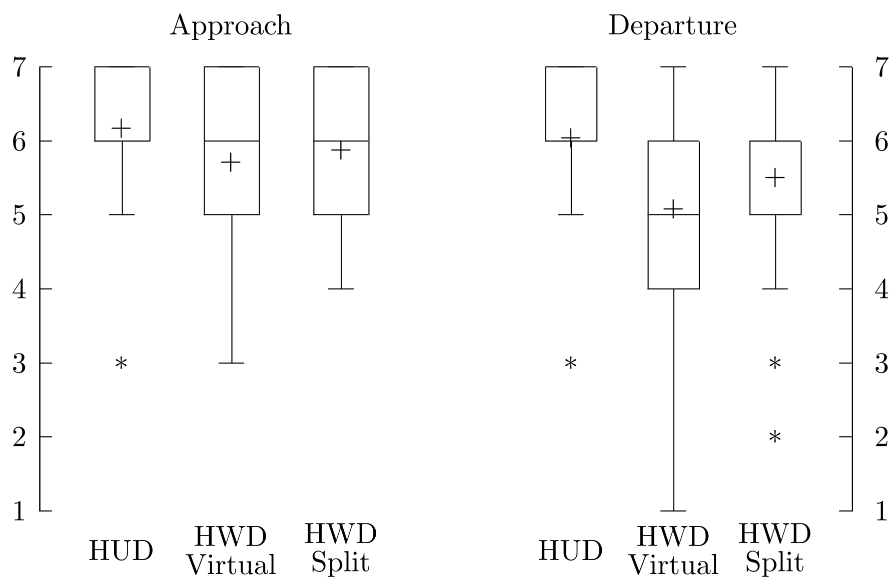

Figure C7. PF ratings for Statement D: The display concepts were effective for management of mental workload.

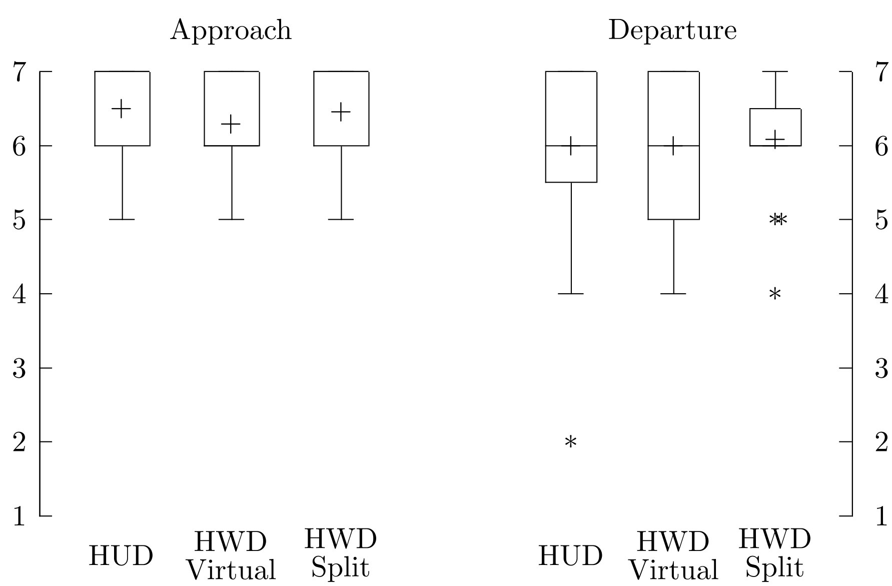

Figure C8. PM ratings for Statement D: The display concepts were effective for management of mental workload.

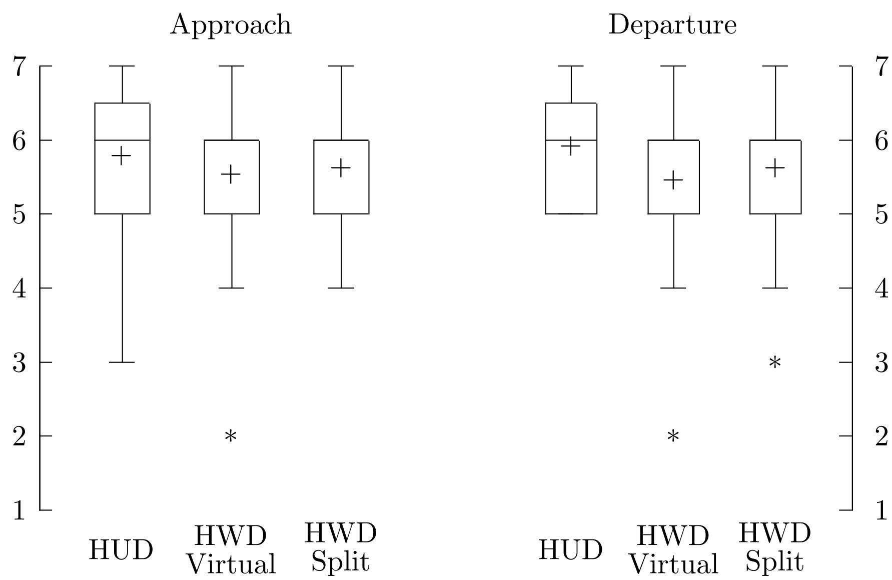

Figure C9. PF ratings for Statement E: The display concepts contributed to communication effectiveness (ATC and crew).

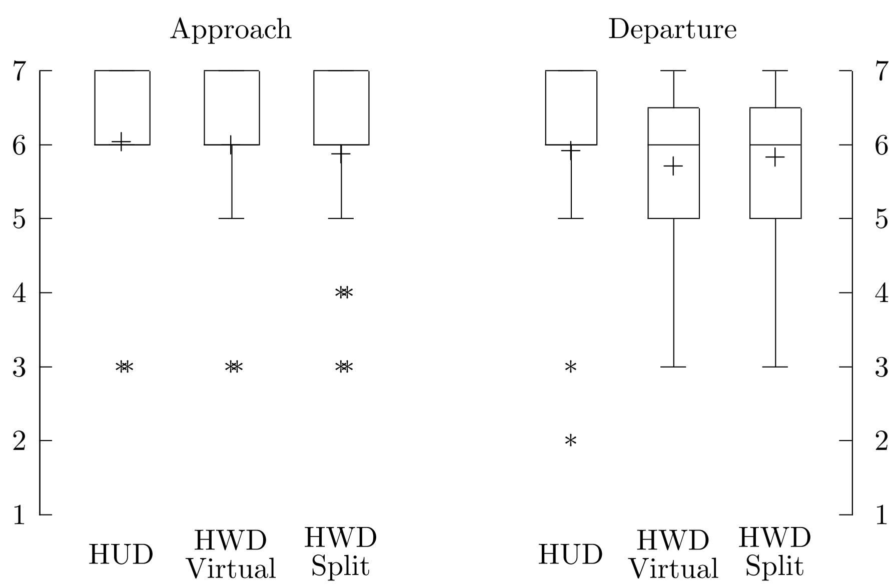

Figure C10. PM ratings for Statement E: The display concepts contributed to communication effectiveness (ATC and crew).

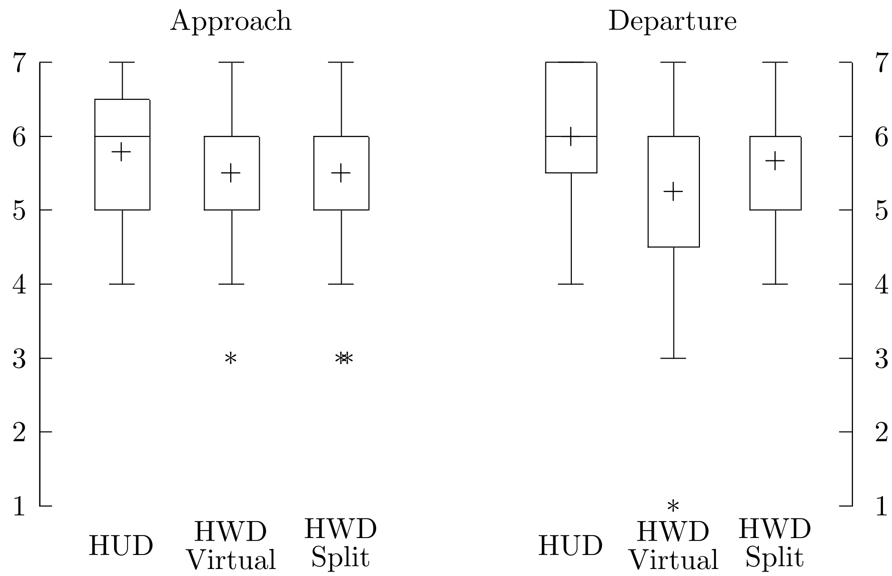

Figure C11. PF ratings for Statement F: The display concepts promoted crew resource management, coordination, and cohesion.

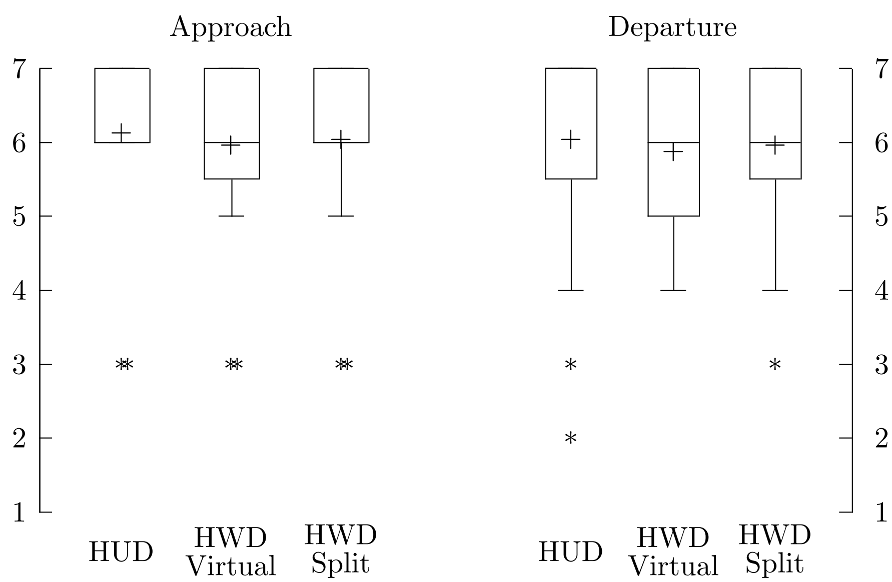

Figure C12. PM ratings for Statement F: The display concepts promoted crew resource management, coordination, and cohesion.

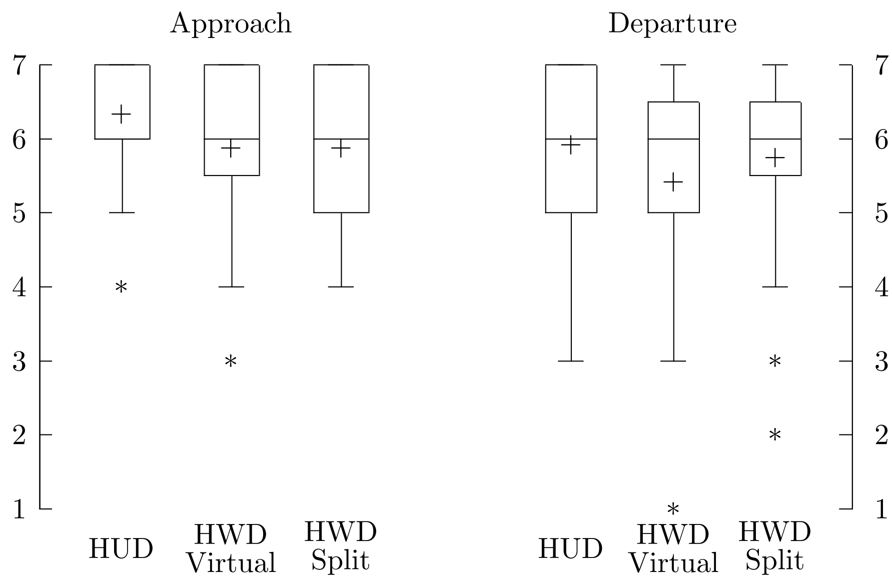

Figure C13. PF ratings for Statement G: The display concepts contributed to perceived safety.

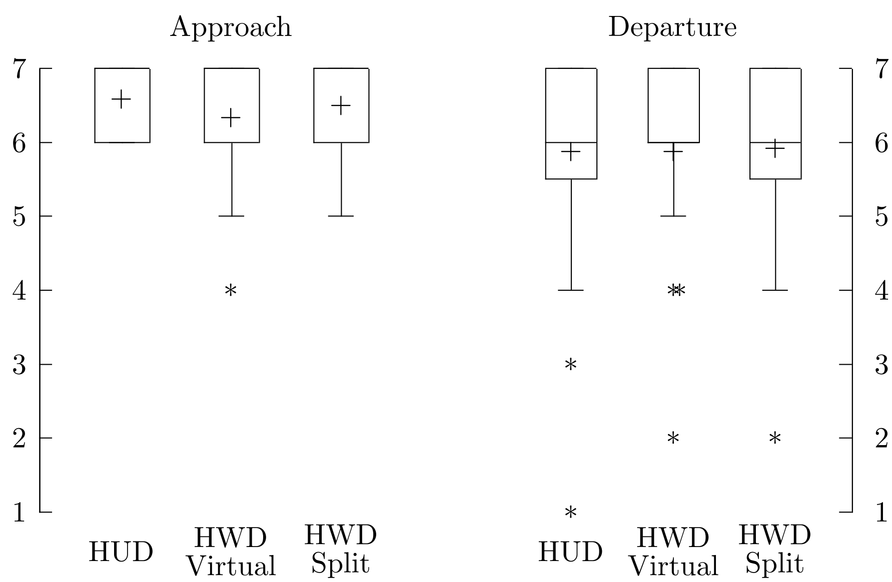

Figure C14. PM ratings for Statement G: The display concepts contributed to perceived safety.

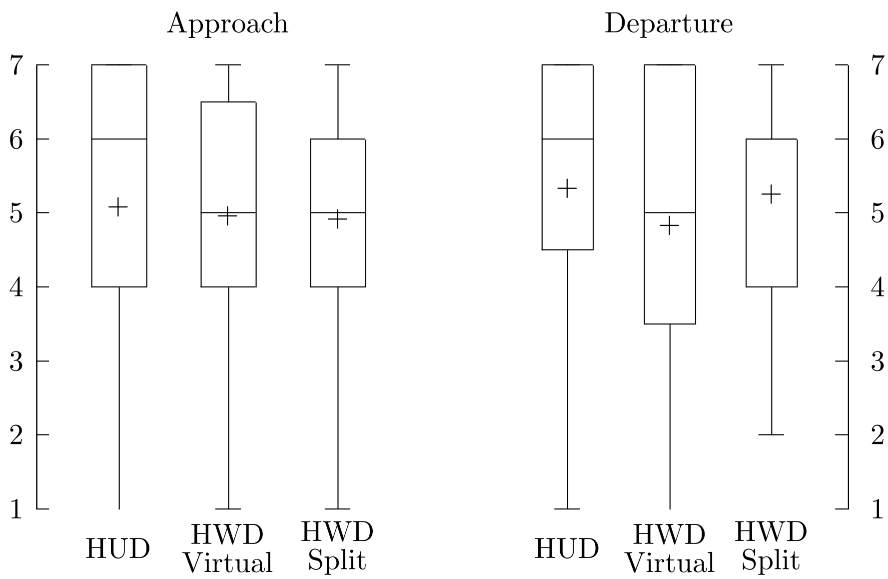

Figure C15. PF ratings for Statement H: The display concepts were effective for the detection of potential surface conflicts.

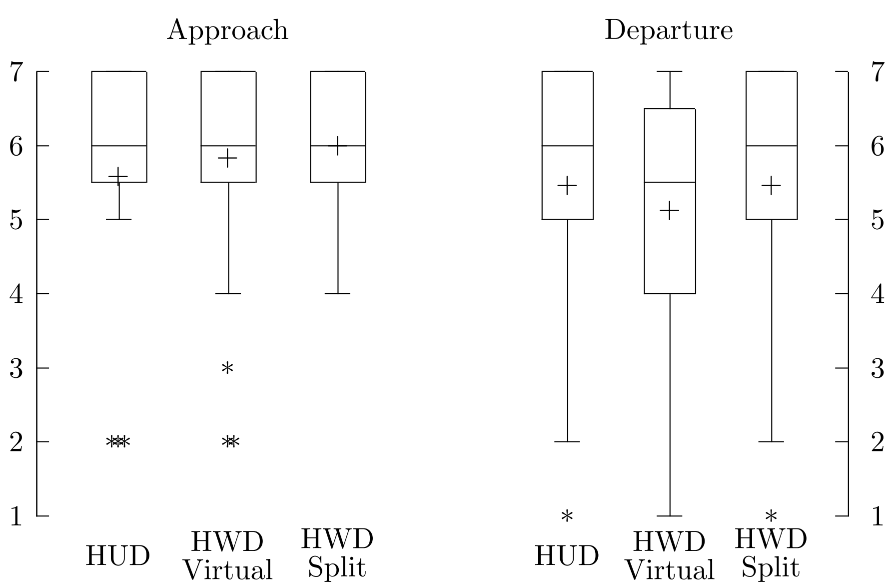

Figure C16. PM ratings for Statement H: The display concepts were effective for the detection of potential surface conflicts.

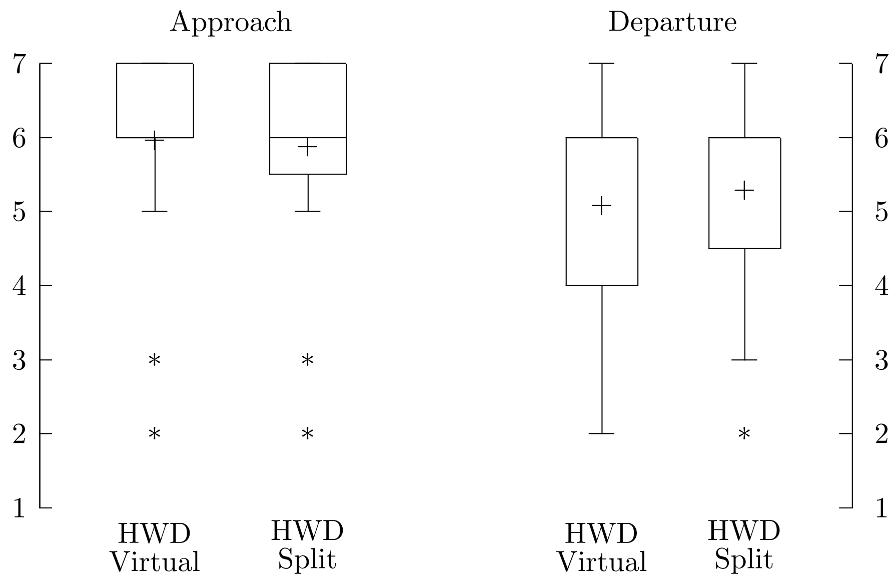

Figure C17. PF ratings for Statement I: If applicable, the display flown was equivalent for use during the approach/departure as the HUD.

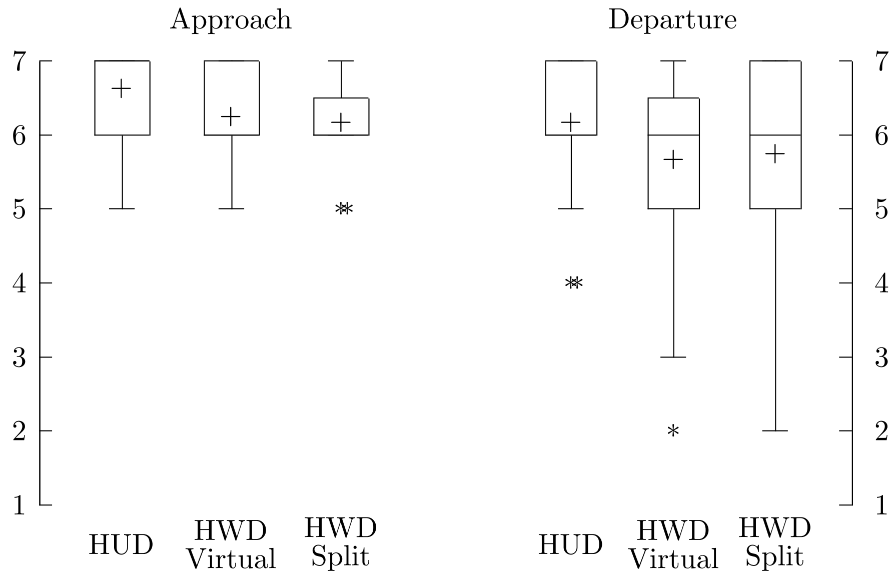

Figure C18. PF ratings for Statement J: The display concepts provided for adequate visual references and awareness.

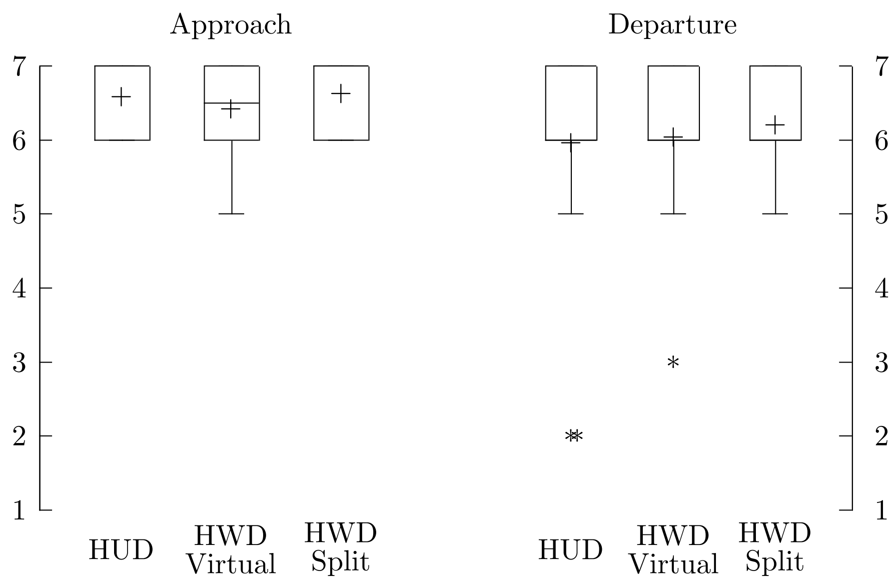

Figure C19. PM ratings for Statement J: The display concepts provided for adequate visual references and awareness.

Appendix D

Simulator Post-Test Question Ratings

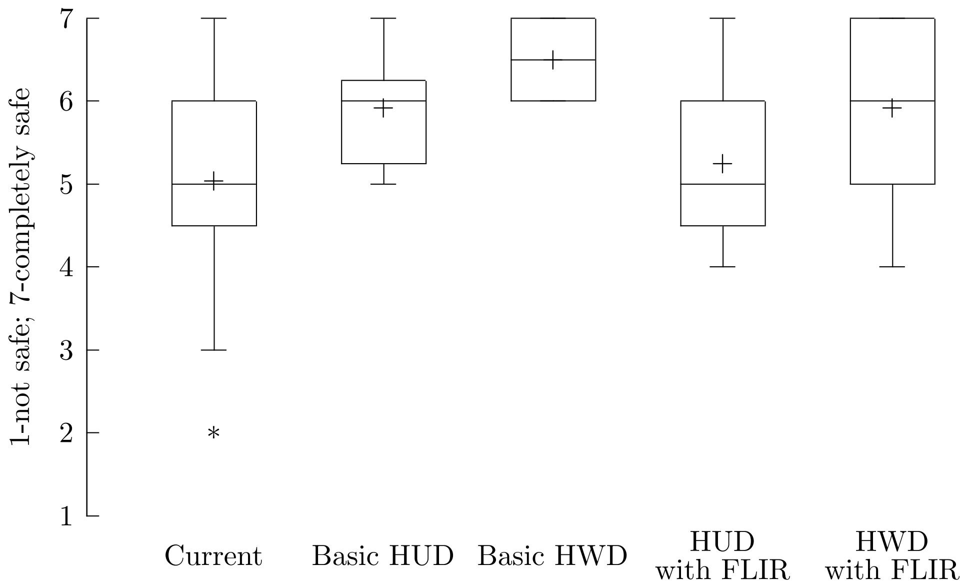

Figure D1. Post-Test Question 1: the PFs ratings of perceived safety for 1000 feet RVR approaches.

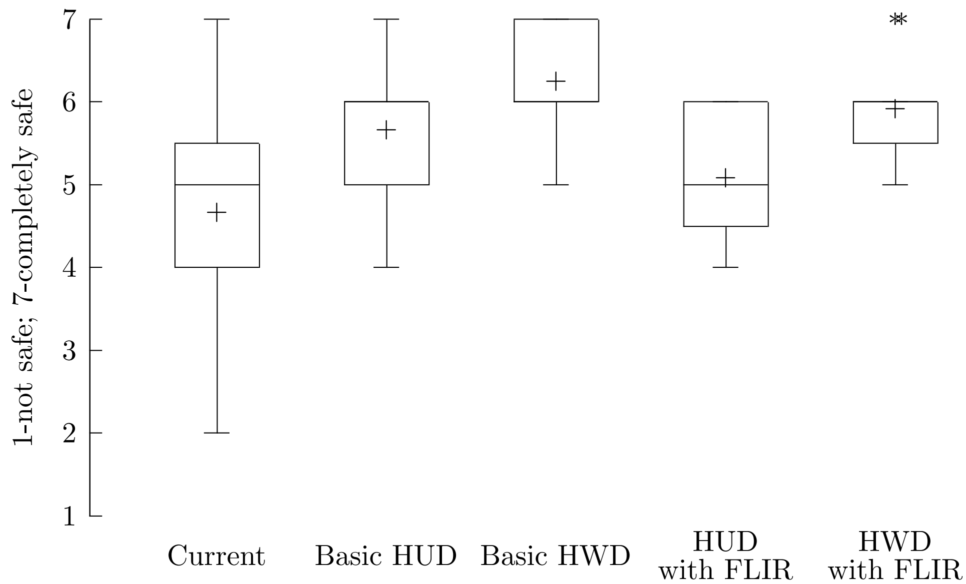

Figure D2. Post-Test Question 2: the PFs ratings of perceived safety for 300 feet RVR taxi-out and departure.

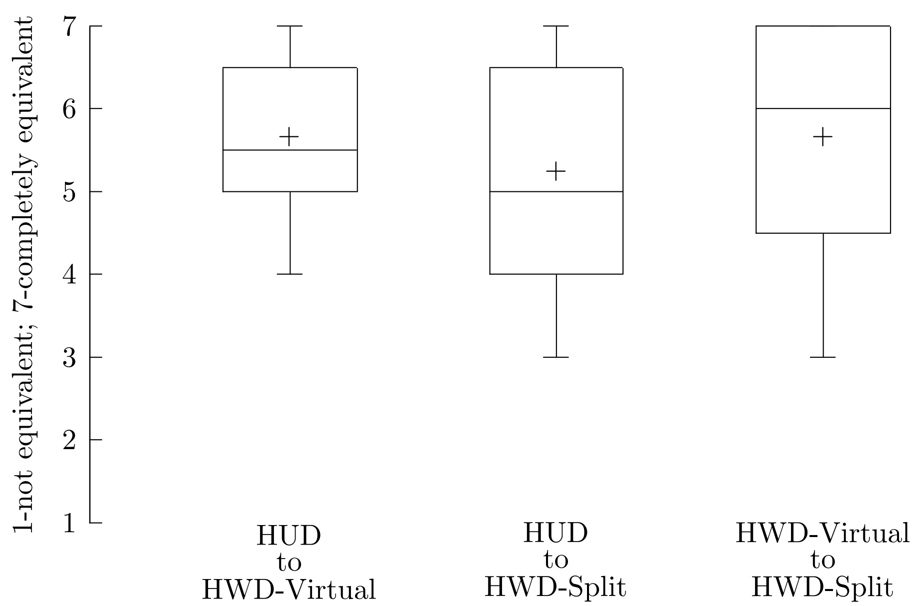

Figure D3. Post-Test Question 3: the PFs ratings comparing Display Concepts on approach with no EV (baseline).

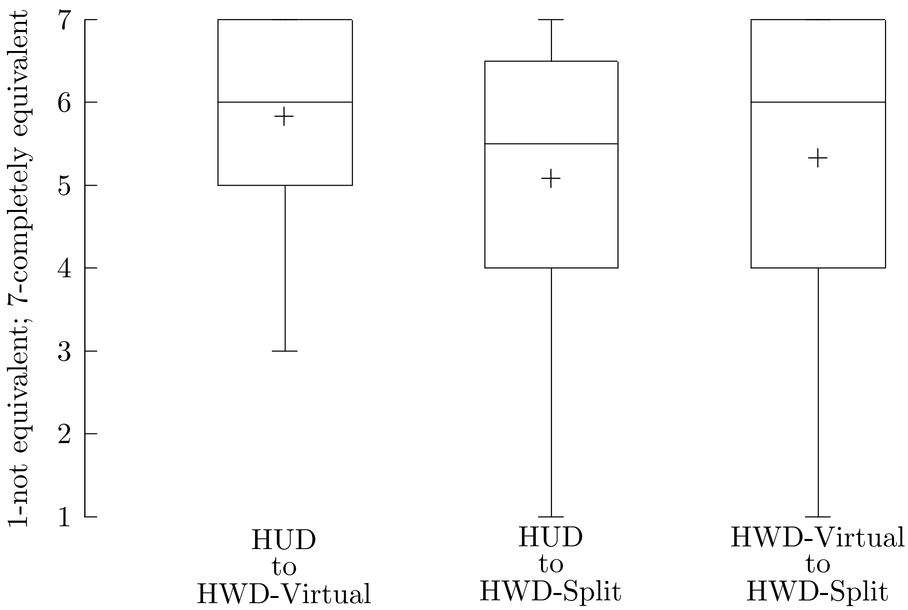

Figure D4. Post-Test Question 4: the PFs ratings comparing Display Concepts of EV presentation on approach.

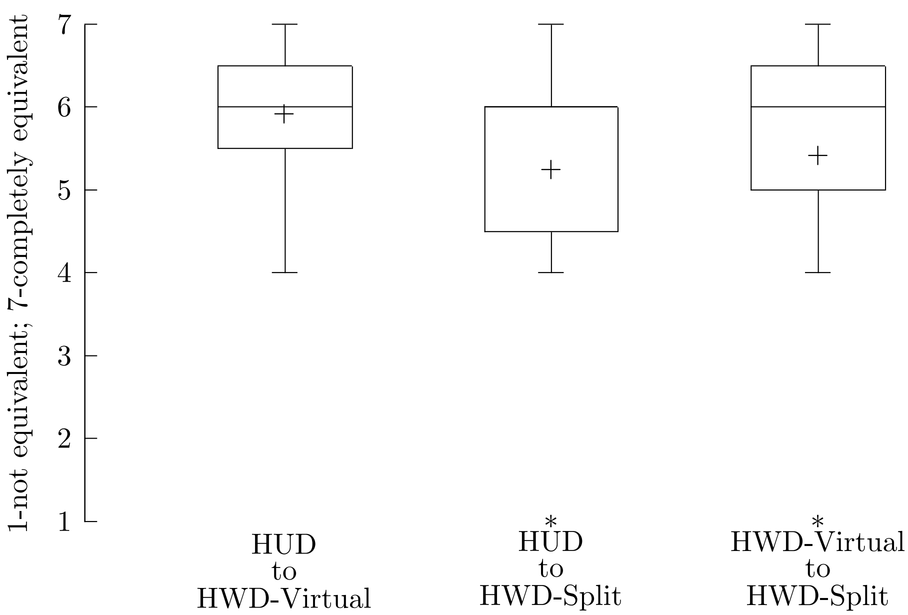

Figure D5. Post-Test Question 5: the PFs ratings comparing Display Concepts for the departure.

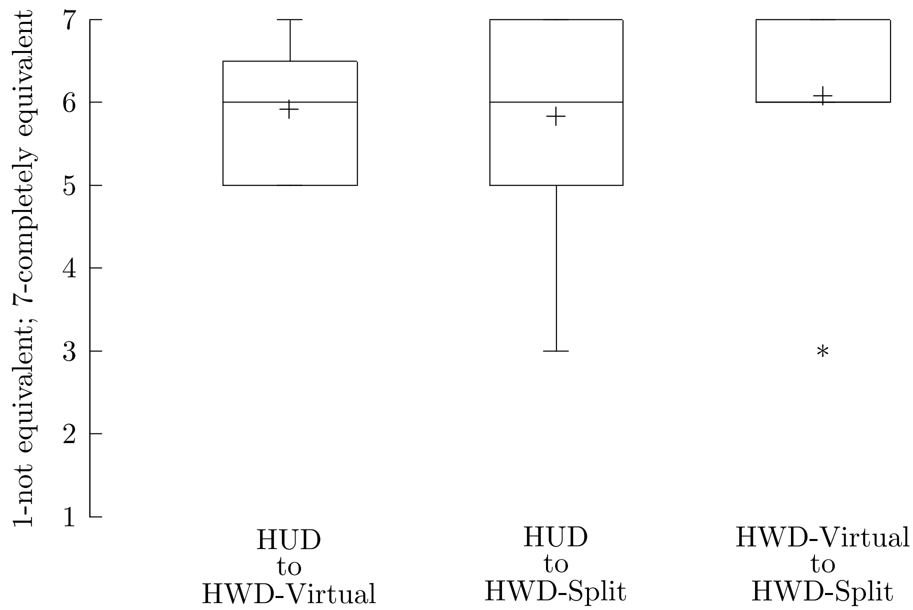

Figure D6. Post-Test Question 8: the PFs ratings of the EV across Display Concepts.

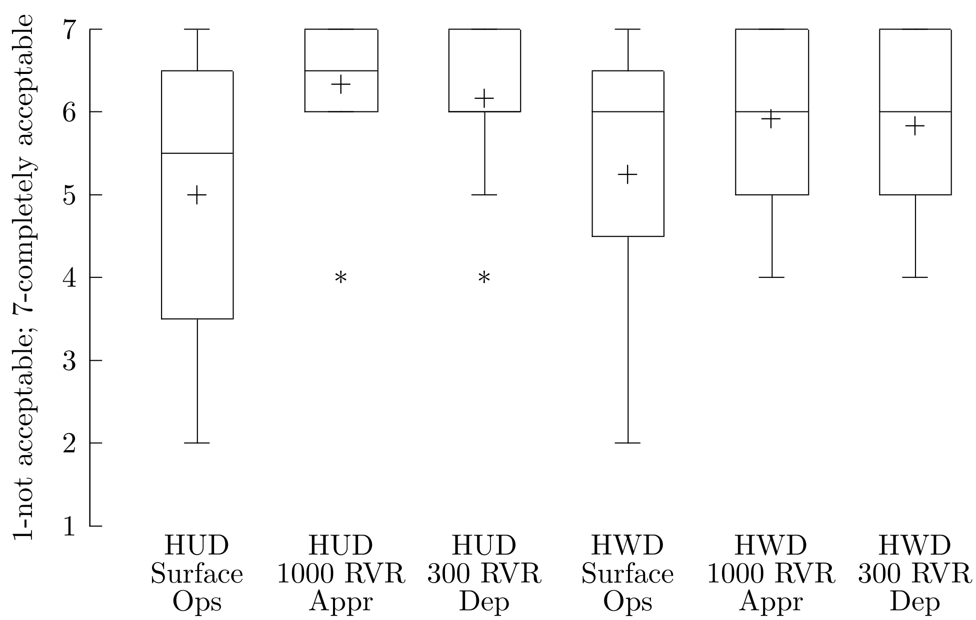

Figure D7. Post-Test Question 9: the PFs ratings of the display efficacy for a given operation.

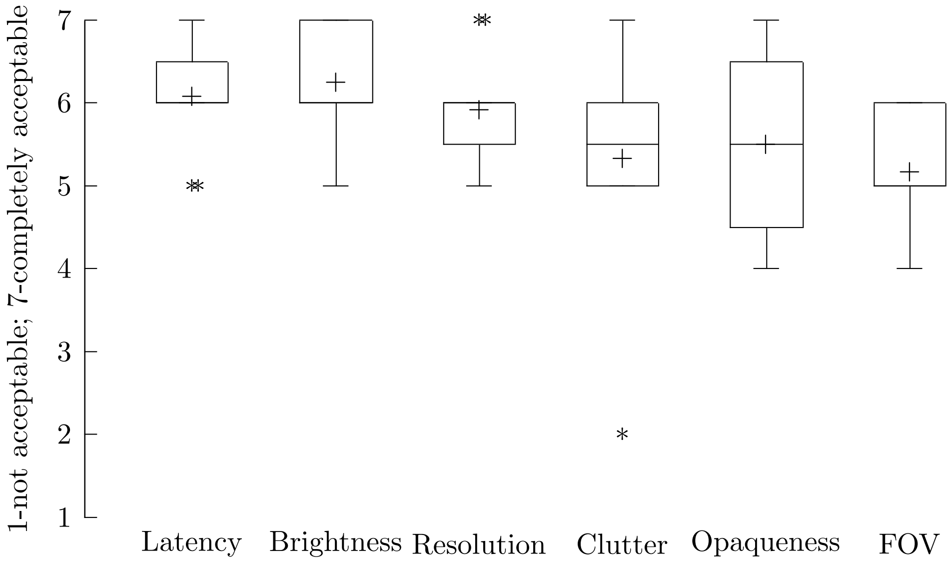

Figure D8. Post-Test Question 10: the PFs ratings of the EV (FLIR) attributes.

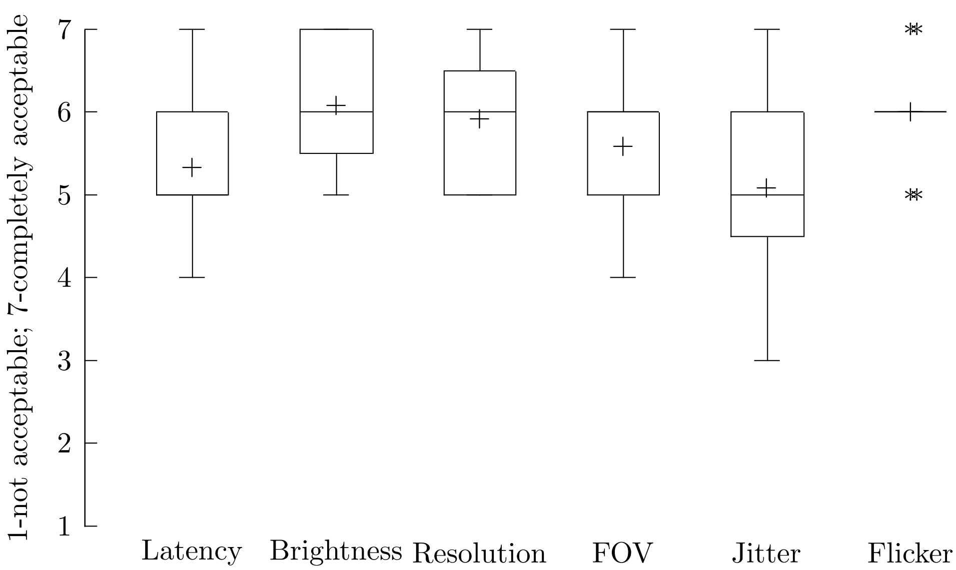

Figure D9. Post-Test Question 11: the PFs ratings of the acceptability of the HWD attributes.

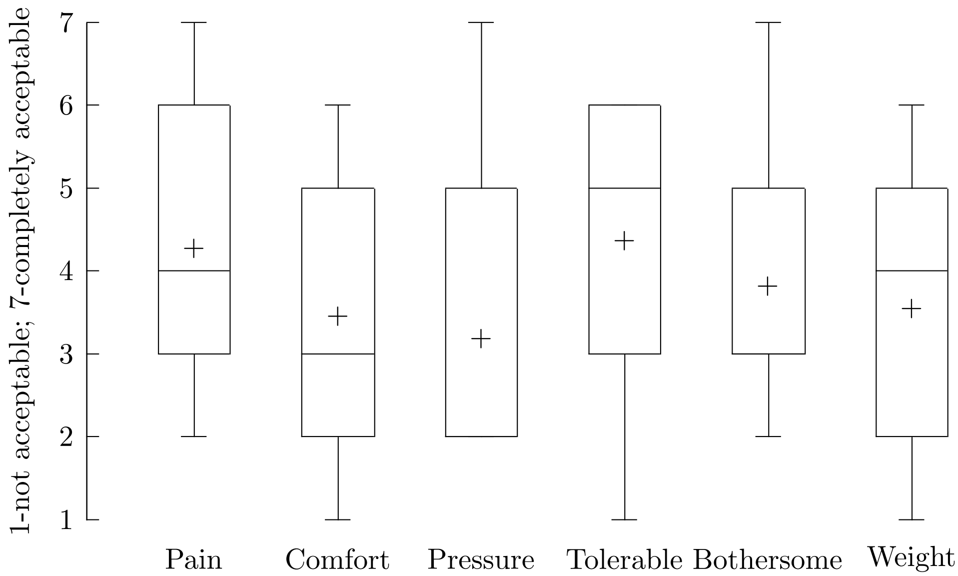

Figure D10. Post-Test Question 12 (part 1): the PFs ratings of the encumbrance of the HWD.

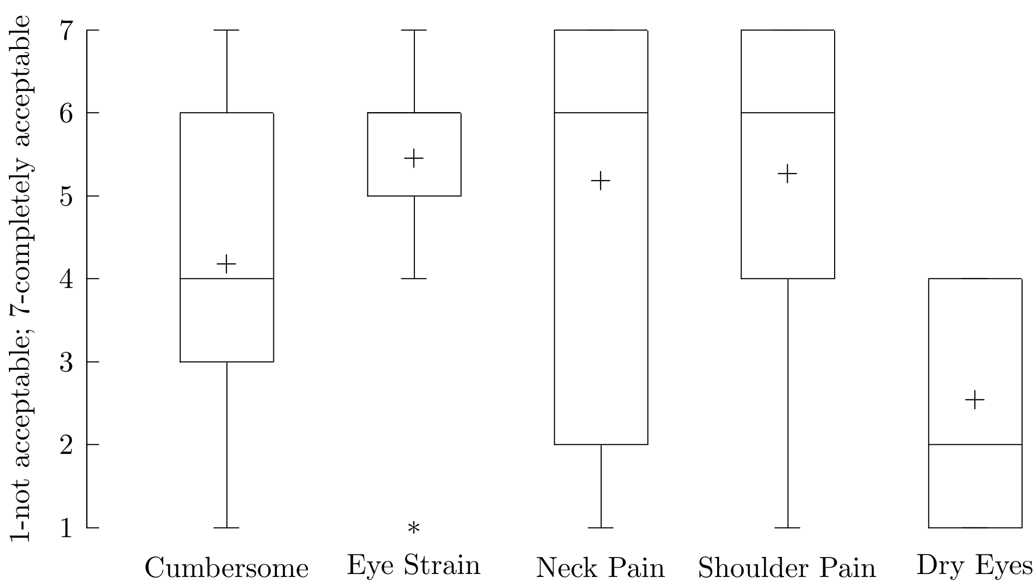

Figure D11. Post-Test Question 12 (part 2): the PFs ratings of the encumbrance of the HWD.

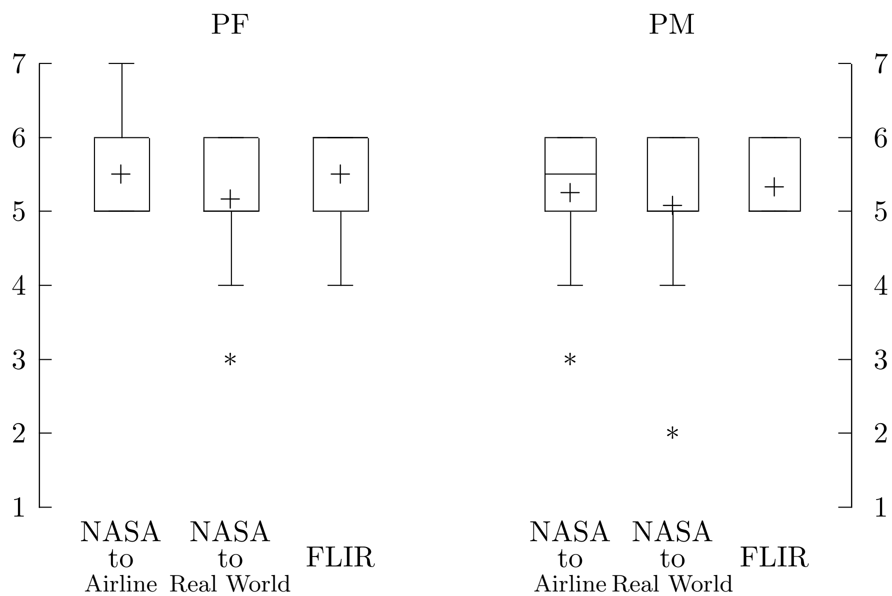

Figure D12. Post-Test Question 13: the PFs ratings of the NASA simulator to the real world and a typical airline simulator.

Appendix E

Flight Test Localizer and Glideslope Calculations

All approaches were either to Runway 26 at Langley Airforce Base (FAA identifier: KLFI) or Runway 25 at Patrick Henry Field (FAA identifer: KPHF). For the research aircraft, the localizer and glideslope were not available on the data bus; thus, the localizer and glideslope indications were generated via the display software using the GPS-based aircraft position data. The approach path was a straight-in, 3° glideslope originating from the runway touchdown point. The runway data used for the flight test is shown in Table E1 and this data was obtained from the FAA Internet Datasheet Viewer (http://webdatasheet.faa.gov).

Table E1. Runway data.

LFI		Chart date: 08/21/2014
	Runway	26
	Position	37.08811944°,-76.34468889°
	Altitude (MSL)	8.0 feet
	Bearing	247.62°
	Touchdown point	1160.0 feet from threshold
PHF		Chart date: 05/31/2012
	Runway	25
	Position	37.13671436°,-76.47815067°
	Altitude (MSL)	41.2 feet
	Bearing	237.81°
	Touchdown point	1050.0 feet from threshold

The localizer error was calculated by

$$loc_{raw} = \arctan\left(\frac{error_{lat}}{loc_d + range}\right)$$

$$loc_{dot} = \frac{loc_{raw}}{0.25 * loc_w}$$

where loc_{dot} is the localizer error in dots, $error_{lat}$ is the lateral path error in feet, loc_d is a constant localizer distance dependent upon the runway (10099.0 feet for Runway 26 at LFI and 8042.0 feet for Runway 25 at PHF), $range$ is the current slant range in feet from the ownship position to the touchdown point and loc_w is a constant localizer width in degrees dependent upon the runway (3.56° for Runway 26 at LFI and 4.42° for Runway 25 at PHF). The value of loc_{dot} was constrained between ±2 dots.

The glideslope error was calculated by

$$angle_{deg} = \frac{180^o}{\pi}[\arctan(\frac{altitude_{thresh}}{range})]$$

$$gs_{dot} = \frac{3.0 - angle_{deg}}{0.35}$$

where gs_{dot} is the glideslope error in dots, $altitude_{thresh}$ is the height difference between ownship altitude and the runway threshold altitude, and $range$ is the current slant range from the ownship position to the touchdown point. The value of gs_{dot} was constrained between ± 2 dots.

Appendix F

Questionnaires

F.1 Situation Awareness Rating Technique (SART)

	Low	*	*	|	*	*	High
Demand on Attentional Resources How much demand was placed on attention due to complexity and variability of the task?	1	2	3	4	5	6	7
Supply of Attentional Resources How much spare attention and mental ability was available during the task?	1	2	3	4	5	6	7
Understanding What was the level of understanding of information and familiarity of the situation?	1	2	3	4	5	6	7

F.2 AFFTC Workload Estimate

Rating	**Workload Estimate**
1	Nothing To Do; No System Demands
2	Light Activity; Minimum Demands
3	Moderate Activity; Easily Managed; Considerable Spare Time
4	Busy; Challenging But Manageable; Adequate Time Available
5	Very Busy; Demanding To Manage; Barely Enough Time
6	Extremely Busy; Very Difficult; Non-Essential Tasks Postponed
7	Overloaded; System Unmanageable; Important Tasks Undone

F.3 NASA Task Load Index

Hart and Staveland's NASA TLX [23] method assesses workload.

Mental Demand How mentally demanding was the task?

0 50 100

0 = Very Low Very High = 100

Physical Demand How physically demanding was the task?

0 50 100

0 = Very Low Very High = 100

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

0 50 100

0 = Very Low Very High = 100

Performance How successful were you in accomplishing the task?

0 50 100

0 = Failure Perfect = 100

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

0 50 100

0 = Very Low Very High = 100

Frustration How irritated and stressed were you?

0 50 100

0 = Very Low Very High = 100

F.4 Simulator Post-Run Questionnaire

Post-Run Ratings	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree
Please rate agreement with statements based on display condition just evaluated	1	2	3	4	5	6	7
A. I was aware of ownship position.							
B. I was aware of traffic and other vehicles during operations.							
C. The display concepts were effective for maintaining SA.							
D. The display concepts were effective for management of mental workload.							
E. The display concepts contributed to communication effectiveness (ATC and crew).							
F. The display concepts promoted effective crew resource management, coordination, and cohesion.							
G. The display concepts contributed to perceived safety.							
H. The display concepts were effective for detection of potential surface conflicts.							
I. If applicable, the display flown was equivalent for use during the approach/departure as the HUD.							
J. The display concepts provided for adequate visual references and awareness (for approach, in terms of flight path, altitude, runway, landing zone; for departure, in terms of maintaining centerline and runway heading).							

F.5 Simulator Sickness Questionnaire

Instructions: Please circle the severity of any symptoms that apply to you right now.

1.	General Discomfort	None	Slight	Moderate	Severe
2.	Fatigue	None	Slight	Moderate	Severe
3.	Headache	None	Slight	Moderate	Severe
4.	Eye Strain	None	Slight	Moderate	Severe
5.	Difficulty Focusing	None	Slight	Moderate	Severe
6.	Increased Salivation	None	Slight	Moderate	Severe
7.	Sweating	None	Slight	Moderate	Severe
8.	Nausea	None	Slight	Moderate	Severe
9.	Difficulty Concentrating	None	Slight	Moderate	Severe
10.	Fullness of Head	None	Slight	Moderate	Severe
11.	Blurred Vision	None	Slight	Moderate	Severe
12.	Dizzy (Eyes Open)	None	Slight	Moderate	Severe
13.	Dizzy (Eyes Closed)	None	Slight	Moderate	Severe
14.	Vertigo*	None	Slight	Moderate	Severe
15.	Stomach Awareness**	None	Slight	Moderate	Severe
16.	Burping	None	Slight	Moderate	Severe

* Vertigo refers to a loss of orientation with respect to upright (i.e., you don't know "which way is up")

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

F.6 Simulator Post-Test Questionnaire

Crew Number: ______________

1. Rate the level of perceived safety you believe you would experience if you were to conduct similar type approaches under 1000' RVR reported under tested conditions.

Current Day Operations (i.e., what you have today)

Not Safe	①	②	③	④	⑤	⑥	⑦	Completely Safe

Head-Up Display (HUD) with NO FLIR + Traffic Icons

Not Safe	①	②	③	④	⑤	⑥	⑦	Completely Safe

Head-Up Display (HUD) with FLIR + Traffic Icons

Not Safe	①	②	③	④	⑤	⑥	⑦	Completely Safe

Head-Worn Display (HWD) with NO FLIR + Traffic Icons

Not Safe	①	②	③	④	⑤	⑥	⑦	Completely Safe

Head-Worn Display (HWD) with FLIR + Traffic Icons

Not Safe	①	②	③	④	⑤	⑥	⑦	Completely Safe

2. Rate the level of perceived safety you believe you would experience if you were to conduct similar type taxi-out and departure operations under 300' RVR reported under tested conditions.

Current Day Operations (i.e., what you have today)

Not Safe	①	②	③	④	⑤	⑥	⑦	Completely Safe

Head-Up Display (HUD) with NO FLIR + Traffic Icons

Not Safe	①	②	③	④	⑤	⑥	⑦	Completely Safe

Head-Up Display (HUD) with FLIR + Traffic Icons

Not Safe	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Completely Safe

Head-Worn Display (HWD) with NO FLIR + Traffic Icons

Not Safe	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Completely Safe

Head-Worn Display (HWD) with FLIR + Traffic Icons

Not Safe	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Completely Safe

3. Please provide your rating of display equivalence in terms of operator use during the approach operations only.

BASELINE

HUD compared to HWD-Virtual

Not Equivalent	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Completely Equivalent

HUD compared to HWD-Split

Not Equivalent	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Completely Equivalent

HWD-Virtual compared to HWD-Split

Not Equivalent	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Completely Equivalent

If not completely equivalent, please provide suggestions for improvements that may potentially increase perceived display equivalence to "completely equivalent."

4. Please provide your rating of display equivalence in terms of operator use during the approach operations only.

ENHANCED VISION (FLIR)

HUD compared to HWD-Virtual

Not Equivalent	1	2	3	4	5	6	7	Completely Equivalent

HUD compared to HWD-Split

Not Equivalent	1	2	3	4	5	6	7	Completely Equivalent

HWD-Virtual compared to HWD-Split

Not Equivalent	1	2	3	4	5	6	7	Completely Equivalent

If not completely equivalent, please provide suggestions for improvements that may potentially increase perceived display equivalence to "completely equivalent."

5. Please provide your rating of display equivalence in terms of operator use during the departure operation only.

HUD compared to HWD-Virtual

Not Equivalent	1	2	3	4	5	6	7	Completely Equivalent

HUD compared to HWD-Split

Not Equivalent	1	2	3	4	5	6	7	Completely Equivalent

HWD-Virtual compared to HWD-Split

Not Equivalent	1	2	3	4	5	6	7	Completely Equivalent

If not completely equivalent, please provide suggestions for improvements that may potentially increase perceived display equivalence to "completely equivalent."

____
____
____
____
____
____

6. In your opinion, could a head-worn display replace a head-up display? Why or why not?

____
____
____
____
____
____

7. What did you like/prefer about the HWD compared to the HUD (what were its advantages, if any)? What did you dislike/not prefer about the HWD compared to the HUD?

____
____
____
____
____
____

8. Please provide your rating of display equivalence for the forward looking infrared as presented on the HUD and HWDs.

HUD compared to HWD-Virtual

Not Equivalent	①	②	③	④	⑤	⑥	⑦	Completely Equivalent

HUD compared to HWD-Split

Not Equivalent	①	②	③	④	⑤	⑥	⑦	Completely Equivalent

HWD-Virtual compared to HWD-Split

Not Equivalent	①	②	③	④	⑤	⑥	⑦	Completely Equivalent

9. Please provide a rating of efficacy (its potential) of the HUD and HWD for each type of operation:

Head-Up Display

Surface Operations

Low Efficacy	1	2	3	4	5	6	7	High Efficacy

1000' RVR EFVS Operational Credit (91.175) Approaches

Low Efficacy	1	2	3	4	5	6	7	High Efficacy

300' RVR Low Visibility Take-Off Departures

Low Efficacy	1	2	3	4	5	6	7	High Efficacy

Head-Worn Display

Surface Operations

Low Efficacy	1	2	3	4	5	6	7	High Efficacy

1000' RVR EFVS Operational Credit (91.175) Approaches

Low Efficacy	1	2	3	4	5	6	7	High Efficacy

300' RVR Low Visibility Take-Off Departures

Low Efficacy	1	2	3	4	5	6	7	High Efficacy

10. Please provide your rating of acceptability for each of the following qualities of the forward looking infrared (FLIR) used today.

Latency

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

Brightness

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

Resolution

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

Display Clutter

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

Opaqueness (see-through)

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

Size (Field-of-View of FLIR)

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

11. Please provide your rating of each of the following qualities of the head-worn display.

Latency

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

Brightness

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

Resolution

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

Instantaneous Field-of-View

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

Jitter

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

Flicker

Not Acceptable	1	2	3	4	5	6	7	Completely Acceptable

12. Please provide your rating of the following qualities of encumbrance of the HWD you wore today including rating of comfort quality. Please use the scale below and substitute the attribute (e.g., Painless = 1; Painful = 7; Intolerable = 1; Tolerable = 7).

painless	1	2	3	4	5	6	7	painful
comfortable	1	2	3	4	5	6	7	uncomfortable
no pressure	1	2	3	4	5	6	7	feel pressure
tolerable	1	2	3	4	5	6	7	intolerable
not bothersome	1	2	3	4	5	6	7	bothersome
heavy	1	2	3	4	5	6	7	light
not cumbersome	1	2	3	4	5	6	7	cumbersome
no pain in neck	1	2	3	4	5	6	7	feel pain in neck
no pain in shoulders	1	2	3	4	5	6	7	feel pain in shoulders
no eye strain	1	2	3	4	5	6	7	eye strain
eyes dry	1	2	3	4	5	6	7	eyes tearing

13. Please provide your rating of the quality of the simulator for each of the following compared to your experiences.

NASA Simulator compared to your airline simulator

Very Poor	1	2	3	4	5	6	7	Excellent

NASA Simulator compared to Real-World

Very Poor	①	②	③	④	⑤	⑥	⑦	Excellent

Forward-Looking Infrared (FLIR)

Very Poor	①	②	③	④	⑤	⑥	⑦	Excellent

F.7 Flight Test Ground Questionnaire

A. Please rate the overall comfort of the HWD:

1	2	3	4	5
Extremely Uncomfortable	Uncomfortable	Neutral	Comfortable	Extremely Comfortable

Comments:

B. Please rate the clarity of the symbology:

1	2	3	4	5
Completely Unreadable	Slightly Unreadable	Neutral	Readable	Extremely Readable

Comments:

C. There was a discernible color shift of the external world due to the HWD. (i.e., change in taxi lights, stop bar lights, or similar lighting when the HWD is moved up and down):

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

D. There was a discernible distortion of the external world due to the HWD. (i.e., when the HWD is moved up and down):

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

E. I experienced significant glare while using the HWD:

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

F.8 Flight Test Post-Run Questionnaire

Post-Run Ratings	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree
Please rate agreement with statements based on display condition just evaluated	1	2	3	4	5	6	7
A. The HWD caused me to experience eye strain.							
B. The HWD caused me to experience headaches.							
C. The HWD was comfortable to wear during the task/operation.							
D. The HWD symbology was easy to read (in terms of clarity).							
E. The HWD video/imagery was easy to read (in terms of clarity).							
F. The HWD concept field-of-view was acceptable to perform the task and operation.							
G. The HWD did NOT obscure or impair my ability to see traffic or other vehicles.							
H. The HWD concept provided usable and sufficient visual cues to safely perform the task/operation.							
I. The HWD symbology and imagery was conformal (i.e., aligned and scaled) to the outside world.							

F.9 Flight Test Post-Test Questionnaire

F. Please rate the overall comfort of the HWD:

1	2	3	4	5
Extremely Uncomfortable	Uncomfortable	Neutral	Comfortable	Extremely Comfortable

Comments:

G. Please rate the effectiveness of the bore sighting procedures:

1	2	3	4	5
Extremely Difficult	Difficult	Neutral	Easy	Extremely Easy

Comments:

H. Please rate the easy of use regarding brightness control:

1	2	3	4	5
Extremely Difficult	Difficult	Neutral	Easy	Extremely Easy

Comments:

I. The HWD's vertical field-of-view was sufficient to provide conformal display of flight guidance information throughout the intended operational envelope:

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

J. There was a discernible color shift of the external world due to the HWD. (i.e., change in taxi lights, stop bar lights, or similar lighting when the HWD is moved up and down):

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

K. There was a discernible distortion of the external world due to the HWD. (i.e., when the HWD is moved up and down):

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

L. The HWD's lateral field-of-view was sufficient to provide conformal display of flight guidance information throughout the intended operational envelope:

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

M. The simulated FLIR was helpful in increasing situation awareness during tasks:

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

N. The HWD system caused me to experience eye strain:

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

O. The HWD system caused me to experience headaches:

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

P. Please rate your overall perceived safety of the HWD:

1	2	3	4	5
Extremely Unsafe	Somewhat Unsafe	Neutral	Safe	Extremely Safe

Comments:

Q. Please rate your overall perceived safety of the HWD as compared with current production HUDs:

1	2	3	4	5
Extremely Unsafe	Somewhat Unsafe	Neither More nor Less Safe	Safe	Extremely Safe

Comments:

R. Based on your overall experience with HUDs, would you consider this HWD equivalent to a HUD:

1	2	3	4	5
Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree

Comments:

S. What improvements do you feel could be made to the HWD for better comfort, performance, etc.?

F.10 Flight Test Paired Comparisons

Paired Comparison Rating Instructions: Each paired comparison will be listed on the left side of the questionnaire. The following example shows how to make the comparisons. Do not take an excessive amount of time on each comparison; your first impression is usually best. However, please feel free to correct any comparisons. Also, the data will be checked for consistency; if the results are inconsistent, you may be asked to clarify your responses.

Example	If **not equal**, how much **more** or how much **less**? Barely ... Substantially							
Display Concept 'X' Is (✓ more)(_ equal)(_ less) better than Display Concept 'Y'							✓	

Flight Guidance	If **not equal**, how much **more** or how much **less**? Barely ... Substantially							
HUD Is (_ more)(_ equal)(_ less) better than NO HUD (Baseline)								
HUD Is (_ more)(_ equal)(_ less) better than HWD								
HWD Is (_ more)(_ equal)(_ less) better than NO HUD (Baseline)								

Situation Awareness	If **not equal**, how much **more** or how much **less**? Barely ... Substantially							
HUD Is (_ more)(_ equal)(_ less) SA than NO HUD (Baseline)								
HUD Is (_ more)(_ equal)(_ less) SA than HWD								
HWD Is (_ more)(_ equal)(_ less) SA than NO HUD (Baseline)								

REPORT DOCUMENTATION PAGE						Form Approved OMB No. 0704–0188

1. REPORT DATE *(DD-MM-YYYY)*
01-05-2017

2. REPORT TYPE
Technical Publication

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Evaluation of a Head-Worn Display System as an Equivalent Head-Up Display for Low Visibility Commercial Operations

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Arthur III, Jarvis (Trey), J.; Shelton, Kevin J.; Prinzel III, Lawrence J.; Nicholas, Stephanie N.; Barnes, James R.; Williams, Steven P.; Ellis, Kyle E.; Jones, Denise R.; Harrison, Stephanie J.; Bailey, Randall E.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
330693.04.80.07.02

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, Virginia 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER
L–19449

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA-TP–2017–219603

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified-Unlimited
Subject Category 03
Availability: NASA STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES
An electronic version can be found at http://ntrs.nasa.gov.

14. ABSTRACT
Research, development, test, and evaluation of flight deck interface technologies is being conducted by the National Aeronautics and Space Administration (NASA) to proactively identify, develop, and mature tools, methods, and technologies for improving overall aircraft safety of new and legacy vehicles operating in the Next Generation Air Transportation System (NextGen). One specific area of research was the use of small HWDs to serve as a possible equivalent to a HUD. A simulation experiment and a flight test were conducted to evaluate if the HWD can provide an equivalent level of performance to a HUD. For the simulation experiment, airline crews conducted simulated approach and landing, taxi, and departure operations during low visibility operations. In a follow-on flight test, highly experienced test pilots evaluated the same HWD during approach and surface operations. The results for both the simulation and flight tests showed that there were no statistical differences in the crews' performance in terms of approach, touchdown and takeoff; but, there are still technical hurdles to be overcome for complete display equivalence including, most notably, the end-to-end latency of the HWD system.

15. SUBJECT TERMS
Enhanced Flight Vision Systems, HUD, HWD, Enhanced Vision, NextGen, Equivalent Visual Operations, Flight Simulation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Information Desk (help@sti.nasa.gov)
U	U	U	UU	133	19b. TELEPHONE NUMBER *(Include area code)* (757) 864-9658

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18