

# Performance comparison between a head-worn display system and a head-up display for low visibility commercial operations

Jarvis (Trey) J. Arthur, III,<sup>a</sup> Lawrence J. Prinzel, III,<sup>a</sup> James R. Barnes,<sup>b</sup> Steven P. Williams,<sup>a</sup> Denise R. Jones,<sup>a</sup> Stephanie J. Harrison,<sup>c</sup> and Randall E. Bailey<sup>a</sup>

<sup>a</sup>NASA Langley Research Center, Hampton, VA

<sup>b</sup>Booz Allen Hamilton Engineering Services, Hampton, VA

<sup>c</sup>Old Dominion University, Norfolk, VA

## ABSTRACT

Research, development, test, and evaluation of flight deck interface technologies is being conducted by NASA 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). Under the Vehicle Systems Safety Technologies (VSST) project in the Aviation Safety Program, one specific area of research is the use of small Head-Worn Displays (HWDs) as an equivalent display to a Head-Up Display (HUD). Title 14 of the US Code of Federal Regulations (CFR) 91.175 describes a possible operational credit which can be obtained with airplane equipment of a HUD or an “equivalent” display combined with Enhanced Vision (EV). If successful, a HWD may provide the same safety and operational benefits as current HUD-equipped aircraft but for significantly more aircraft in which HUD installation is neither practical nor possible. A simulation experiment was conducted to evaluate if the HWD, coupled with a head-tracker, can provide an equivalent display to a HUD. Comparative testing was performed in the Research Flight Deck (RFD) Cockpit Motion Facility (CMF) full mission, motion-based simulator at NASA Langley. Twelve airline crews conducted approach and landing, taxi, and departure operations during low visibility operations (1000’ Runway Visual Range (RVR), 300’ RVR) at Memphis International Airport (Federal Aviation Administration (FAA) identifier: KMEM). The results showed that there were no statistical differences in the crews performance in terms of touchdown and takeoff. Further, there were no statistical differences between the HUD and HWD in pilots’ responses to questionnaires.

**Keywords:** Enhanced Flight Vision Systems, HUD, HWD, Enhanced Vision, NextGen, Equivalent Visual Operations, Flight Simulation

## 1. INTRODUCTION

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). As part of this work, specific research issues associated with the NextGen Terminal Control Area (TCA) are being addressed. 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). This work is part of NASA’s technical challenge under the Vehicle Systems Safety Technologies (VSST) project under the Aviation Safety Program to demonstrate new capabilities that enable pilots to better understand and respond safely to complex situations which might otherwise result in accidents or incidents.

NASA has conducted numerous studies evaluating the potential benefits of using HWDs for surface operations.<sup>1-3</sup> HWDs 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. Further, the HWD may create a “Virtual HUD” concept.<sup>4-6</sup>

---

Further author information: Jarvis (Trey) Arthur, E-mail: Trey.Arthur@nasa.gov, Telephone: 1-757-864-6609

## 1.1 HUD operational credit

The Flight Safety Foundation identified significant safety benefits of head-up/HUD flight operations.<sup>7</sup> In addition to safety benefits, “operational credits” are now being derived from HUD equipage.

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) Advisory Circular (AC) 120-28D);
2. Low visibility takeoff minima of 300’ 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);
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, 1400-foot RVR in lieu of centerline and touchdown zone lighting (as per FAA Order 8400.13).

In this paper, the additional operational credit afforded HUD operations with the simultaneous use of Enhanced Vision (EV) on head-up displays was 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 applications of EV technology for commercial, business, and GA aircraft was energized in January 2004<sup>8</sup> 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 then operate below the published Decision Altitude (DA), DH or Minimum Descent Altitude (MDA) when using an approved Enhanced Flight Vision System (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.

The HUD is the only display currently certified and approved for use as an EFVS. With EFVS, a pilot may descend below the published DA, DH, or 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. However, provisions for the use of an equivalent display were made.

What constitutes an equivalent display is not explicitly defined, but by inference, 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 conformal 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:<sup>9</sup>

- 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 is working toward making the HWD a HUD replacement.<sup>5</sup> However, achieving this capability for business and commercial aircraft is a formidable challenge.

The goal of this research is to evaluate a HWD system as a replacement for a standard flight HUD. If this equivalence can be shown, then the unique capabilities of the HWD - that is, unlimited field-of-regard head-up operations for piloted surface operations<sup>3</sup> - can be capitalized. The design challenge (and certification challenge) is to create this equivalent capability without pilot workload, encumbrance, or obscuration of their normal vision.<sup>10</sup>

## 2. EXPERIMENT

The main objective for this experiment is to determine if a HWD system can provide equivalent performance compared to a HUD. The data from this work should help quantify the characteristics that define an equivalent display. Secondary objectives include the influence of Augmented Reality (AR) symbology on the HWD. Specifically, the AR symbology consisted of a traffic icon which would denote the position of an aircraft based on its ADS-B position. The traffic icon presented head-up is of interest for the prevention of runway incursions.

### 2.1 Simulation facility

This experiment was conducted in the NASA Langley Research Center (LaRC) Research Flight Deck (RFD) motion-based simulator (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 as well as an aircraft moving map were provided on an Electronic Flight Bag (EFB). For this experiment, the EV system was a FLIR camera fixed to the aircraft. The simulated 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.

The test facility had navigation and communication facilities simulation that replicated realistic voice communication during approach, taxi-out, and takeoff scenarios. The communications included pre-recorded voice communications to stage most of the scenarios. A researcher also provided simulated Air Traffic Control (ATC) directives and information to augment the pre-recorded ATC audio, as well as intervention when required (i.e., when a crew queried ATC with questions regarding a clearance).

#### 2.1.1 Head-Worn Display

The HWD used in this experiment is shown in Fig. 2. A prototype head tracker provided head orientation and was mounted on the left side of a pair of Lumus DK-32 glasses. The Lumus glasses specifications are in Table 1 along with the HUD specifications for comparison. The Lumus eye-wear is see-through, full color 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 as to not introduce a confound when comparing to the monochrome HUD.



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

A new prototype head tracker for use in motion simulators and flight test vehicles was developed and consisted of 2 inertial units so that the head movements can be isolated from vehicle movement. With HWD head-tracked systems, latency becomes a major concern.<sup>11</sup> Latency is inherent in HUD systems as well, but latency effects are magnified in a HWD head coupled display system.<sup>11</sup> For the head tracker used in this study, the total “end-to-end” HWD latency was measured to be 85 milliseconds. This total latency measurement was obtained via the “windshield washer” test.<sup>12</sup>

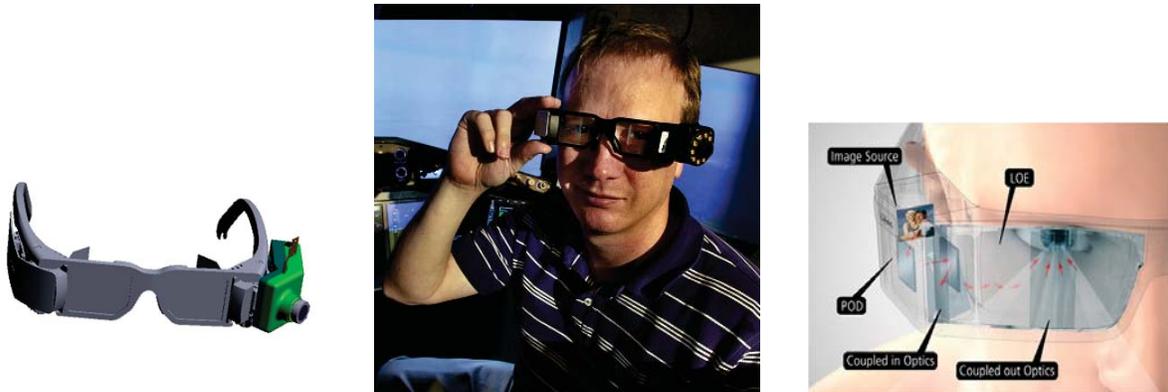


Figure 2. The HWD system used in the experiment. Left figure is a CAD image showing the head tracker mounted on the left temple of the glasses. The middle image shows the device used in testing. The far right image show details of the HWD optics.

Table 1. Display Specifications

	HWD	HUD
<b>Resolution</b>	1280 (H) x 720 (V)	1400 (H) x 1050 (V)
<b>Field-of-View</b>	35 H x 20 V	46 H x 34.5 V
<b>Weight</b>	0.20 kg	14 kg (combiner + overhead)

### 2.1.2 Head-Up Display

The HUD used in this experiment was a Rockwell Collins HGS-6700.

### 2.1.3 HWD/HUD Symbology

During flight, the HWD symbology was designed to replicate the HUD HGS-6700 symbology and functionality (Fig. 3). The flight symbology set was typical HUD symbology for a commercial transport including a flare cue. 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 and flight path angle.

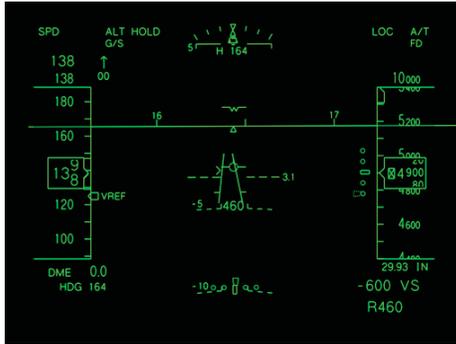


Figure 3. The flight symbology set.

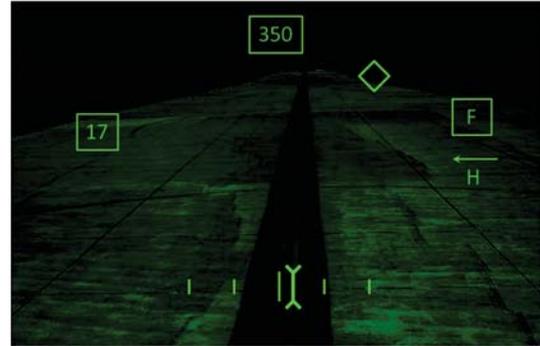


Figure 4. The surface symbology set.

When the nose wheel was on the ground and the ground speed was less than 80 knots, the symbology set (Fig. 3) would automatically transition to the taxi symbology (Fig. 4). The taxi symbology set consisted of ground speed, heading, current taxiway the aircraft is on and the next taxiway on the cleared route. Above the next taxiway text, a left or right arrow was rendered to denote the direction of the next cleared taxiway. Near the bottom of the display was a raw data indicator showing linear deviation from the taxiway centerline. These symbology sets were displayed on both the HUD and HWD.

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

### 2.1.4 Head-Down Displays

The head-down displays (Fig. 5) showed: a) Pilot's left display, including a Primary Flight Display (PFD); b) Pilot's right display including a Navigational Display (ND); c) Co-pilot's left display, including a FLIR display and a ND; and, d) Co-pilot's right display, including a PFD. For data trials where FLIR was displayed on the HUD or HWD, the Pilot Monitoring (PM) would have a "repeater" FLIR head-down. The head-down FLIR repeater did not have overlaying symbology. Two EFBs were utilized (pilot side and co-pilot side) for various functions, including charts, checklists and displaying the airport surface map.

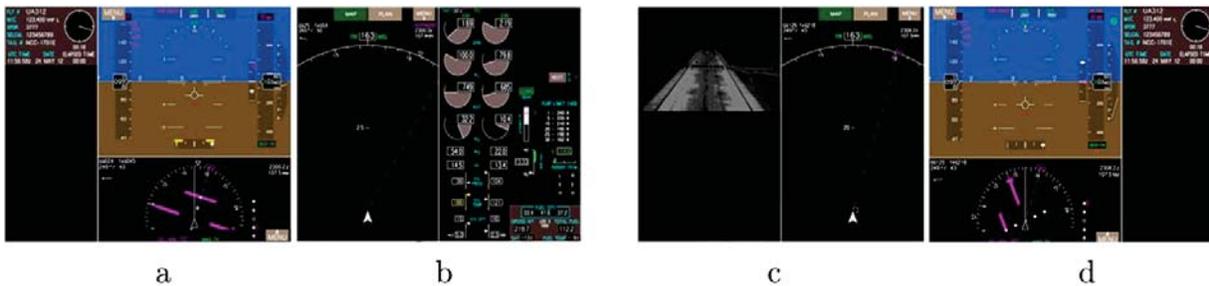


Figure 5. The pilot flying head-down displays (left) and the pilot monitoring head-down displays (right). The FLIR repeater is shown on the pilot monitoring navigation display in the upper left corner.

## 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. The image shift (i.e., error) due to camera parallax for this case was half of the maximum error allowable for an EFVS in accordance with RTCA DO-315,<sup>13</sup> equating to a  $2.5$  mrad 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 Pilot Flying (PF) and sat in the left seat. The First Officer was the PM for the duration of the experiment; thus, crew members did not switch roles for the duration of the experiment. Each pilot 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 Flight crew 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 a  $2$  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 Methodology

Approach and surface operations were conducted at Memphis International Airport (FAA identifier KMEM). The data runs were blocked by display device (HUD/HWD) within an operation block (approach/departure). The experiment was blocked by display to minimize the need for EPs donning and removing the HWD between runs. During data trials using the HWD, the HUD was stowed and EPs donned the HWD for approximately  $45$  minutes.

## 2.6 Evaluation task

The simulated weather conditions were  $1000'$  RVR for the approaches and  $300'$  RVR for the departures. Both approach and departure scenarios were conducted in the daytime with no winds or turbulence. 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 used 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 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.

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 around common practice in current EFVS operations and FAA requirements ( $14$  CFR  $91.175$  (1)).

For approach scenarios, crews were briefed on their starting position ( $1000'$  AGL) 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. Once crews contacted the tower, they were given a landing clearance with a high speed turn-off, if feasible. Once the aircraft was clear of the runway, the approach scenario ended.

For departure scenarios, crews were briefed on their position on the airport. Crews called the ground controller to receive taxi instructions. 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. During the departure trials, crews were instructed by the ground controller to switch to the tower frequency at which time they would receive their departure clearance. Departures scenarios ended at an approximate altitude of 1000' AGL after takeoff.

Post-run questionnaires were given to both EPs after each scenario, and consisted of 1) a 3-point Situational awareness rating technique (SART)<sup>14</sup> form, 2) an Air Force Flight Test Center (AFFTC) 7-point workload scale,<sup>15</sup> 3) a NASA Task Load Index (TLX) workload rating,<sup>16</sup> and 4) 10 questions addressing HWD equivalence, crew interaction, operational effectiveness, and EV usability. At various points during the data collection trials, a Simulation Sickness Questionnaire (SSQ)<sup>17</sup> was administered. These questionnaires were given immediately after the end of each data trial.

### 2.6.1 Approach and landing task

The approach experiment matrix consisted of 1 independent variable, the Display Condition, which consisted of 3 display types: 1) HUD; 2) HWD-Virtual HUD concept; 3) HWD-Split concept. Each display concept was replicated twice for each crew.

The HUD display condition was a typical HUD with FLIR 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 HUD was stowed for all HWD Display Conditions.

The HWD-Split Display Condition consisted of a mix of screen-referenced symbology and conformal symbology. In this condition, non-conformal symbology (such as airspeed and altitude tapes, roll scale, etc.) were drawn in reference to the screen space. In other words, these 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. However, since the velocity vector and the FLIR imagery must remain conformal, these conformal symbology elements were spaced-stabilized as in the HWD "Virtual HUD" condition.

### 2.6.2 Taxi and departure task

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' AGL. The independent variables for the taxi portion of the scenario consisted of Display Condition (HUD, HWD-Virtual, and HWD-Split) and Display Features where the Display Features were: 1) Baseline - a typical HUD with no FLIR image; and, 2) EV+TD - a conformal EV (FLIR) image and traffic "diamond" symbology to denote other traffic (see Fig. 4). The traffic diamond was displayed in a perspective format on the HUD or HWD. In addition, the size of the traffic diamond would grow or shrink depending on the proximity of the traffic to ownship.

During taxi, a surface symbology set was used for both the HUD and the HWD (Fig. 4). Once the aircraft reached the runway, the symbology set would transition from the ground symbology to the takeoff symbology.

## 3. 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) measures were recorded and used in a detailed data analysis to answer the purpose of the research.

### 3.1 Quantitative results

#### 3.1.1 Touchdown performance

Touchdown performance was measured by the crews' ability to land in the touchdown zone of the intended landing runway with acceptable sink rate using normal maneuvering. The touchdown aim point was on the runway centerline and 1000 feet down the runway from the threshold. The landing criteria used was derived from performance standards required by Cat III auto-land systems.<sup>18</sup> 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 performance: 1) "desired", 2) "adequate", and 3) "not adequate". These performance values are defined in Table 2.

Table 2. Touchdown Performance Criteria.

	Desired	Adequate	Not adequate
<b>Lateral</b>	within 27 ft	between 27 and 58 ft	> 58 ft
<b>Longitudinal</b>	750 to 2250 ft	between 200 and 750 ft or between 2250 and 2700 ft	< 200 or > 2700 ft
<b>Sink Rate</b>	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 from the aim point, all landings were "adequate" (Fig. 6) and 88% of the landings were in the "desired" zone.

An Analysis of Variance (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,  $F(2, 69) = 0.105, p = 0.211$ . 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 ( $p = 0.109$ ).

Vertical speed was also captured at the point of touchdown (Fig. 7). The sink rate was adequate for 94% of the landings. Four of the 72 (6%) sink rate values were not adequate and all of those were with the HWD display concept. 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,  $F(2, 69) = 2.678, p = 0.076$ .

The average sink rate was within the "desired" range of sink rate performance at touchdown based on AC 120-28D<sup>18</sup> criteria for all concepts. Examination of maximum sink rate at touchdown evinces that 4 of 72 trials 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), but these were all during trials 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 trials (1.3 ft/sec SD). Considering these data, the HWD performance does not seem anomalous.

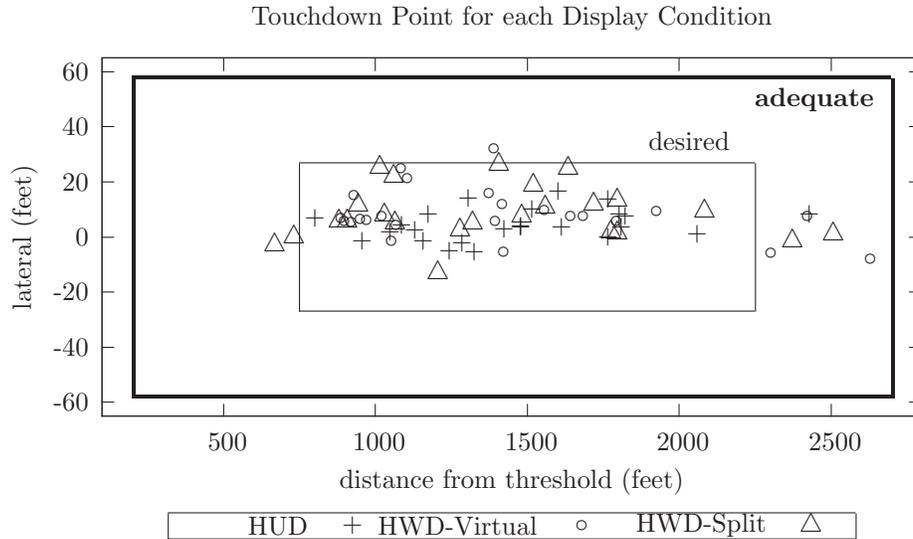


Figure 6. Scatter plot showing touchdown point for all approach runs broken out by display concept.

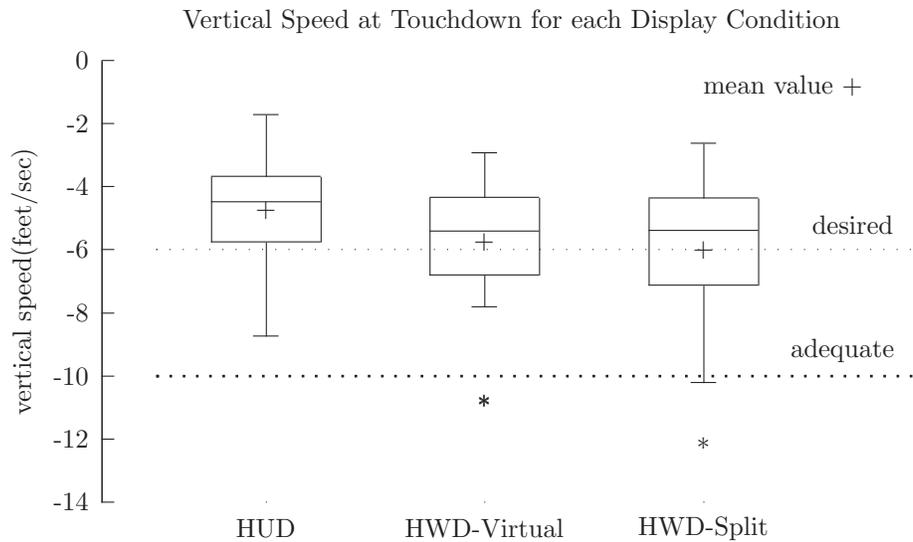


Figure 7. The sink rate at touchdown broken out by Display Condition.

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 8 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.

### 3.1.2 Flight technical error on approach

An ANOVA was conducted on Flight Technical Error (FTE) for localizer dot error (Fig. 9) and glideslope dot error (Fig. 10) tracking performance from an altitude of 1000' AGL to 50 feet. The results found no significant effects for localizer,  $F(2, 69) = 0.341, p = 0.712$ ; or glideslope,  $F(2, 69) = 0.409, p = 0.666$ .

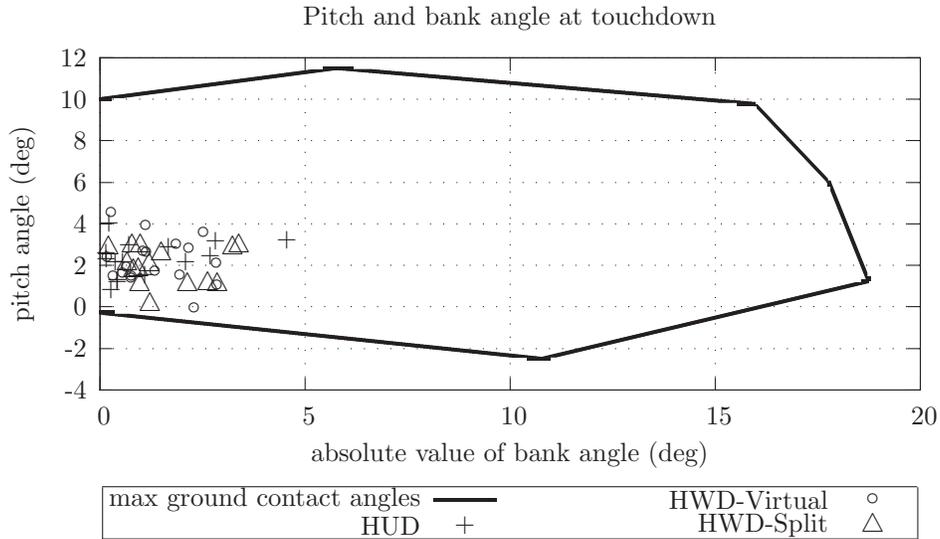


Figure 8. The airplane orientation at touchdown for each Display Condition.

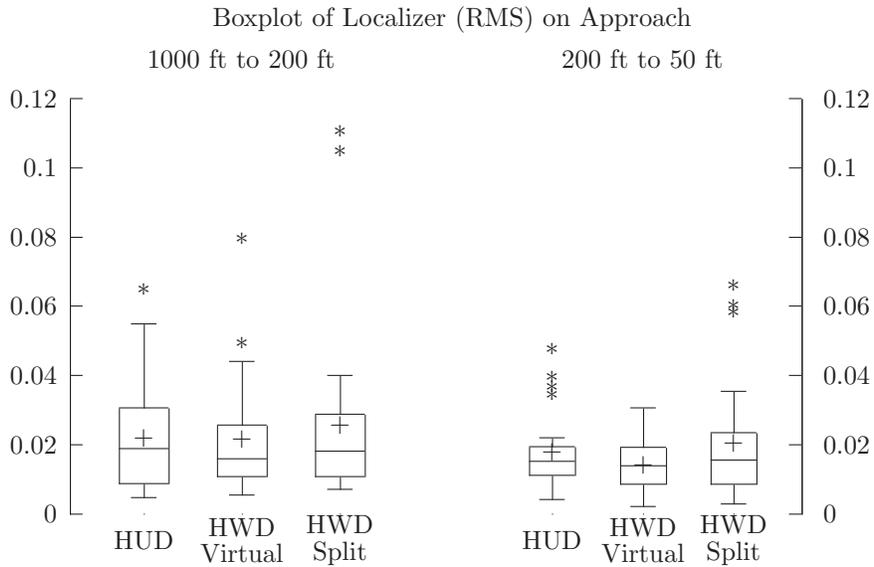


Figure 9. Box plot of the Localizer Dot Error (RMS) on Approach.

### 3.1.3 Average taxi speed to runway

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.

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 11 shows the average taxi speed broken-out by Display Condition and Display Features.

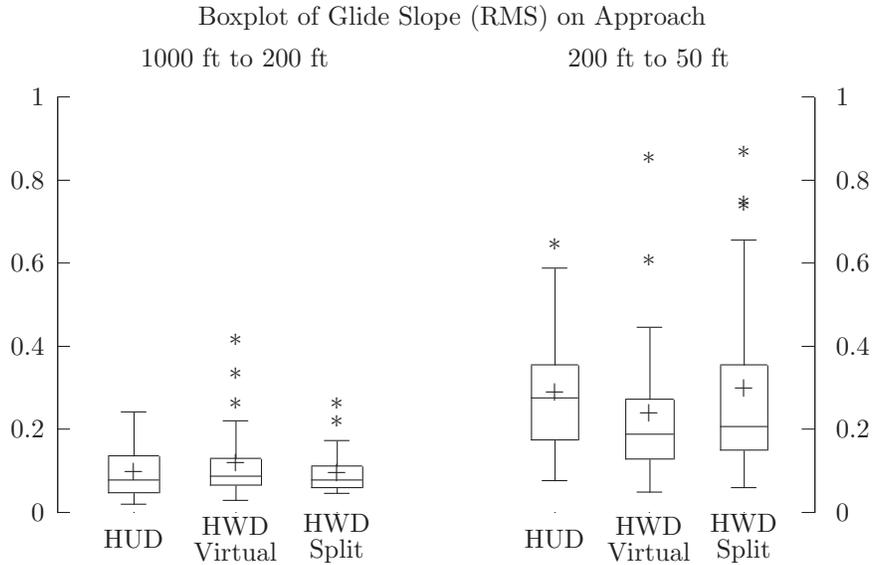


Figure 10. Box plot of the Glide Slope Dot Error (RMS) on Approach.

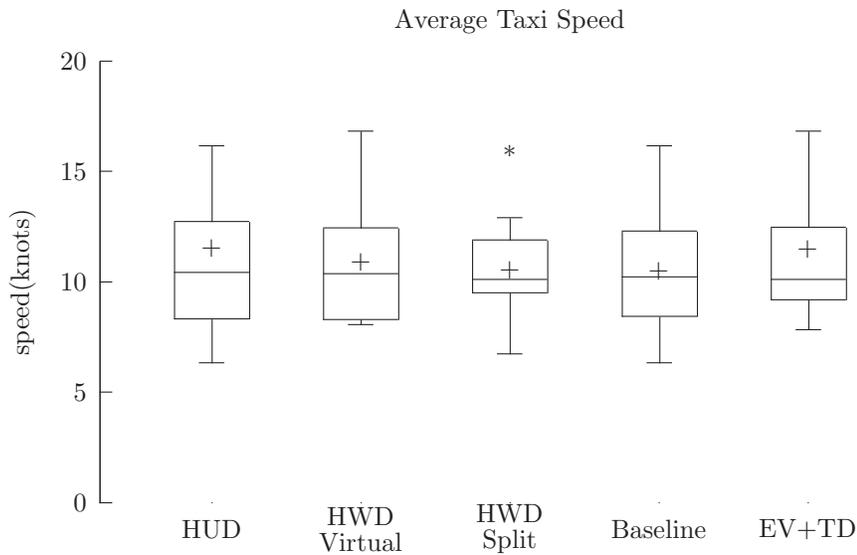


Figure 11. Box plot of the average taxi speed per display and features for departures.

### 3.1.4 Taxi errors

A total of 7 taxi errors occurred during testing. Four of the errors were with the HWD display concept and 3 errors occurred with the HUD condition. The 4 errors with the HWD consisted of one crew going past a hold short line without proper clearance. 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. The remaining 2 errors with the HUD consisted of crews turning onto taxiways which deviated from the cleared taxi route.

### 3.1.5 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 centerline time

during departure roll-out. 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 centerline time during departure roll-out,  $F(2, 69) = 0.709, p = 0.619$ .

### 3.2 Qualitative results

#### 3.2.1 Situation awareness

A 3-point SART was also administered after each run. 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 12 shows the total SART scores broken-out by Display Condition and Approach/Departure.

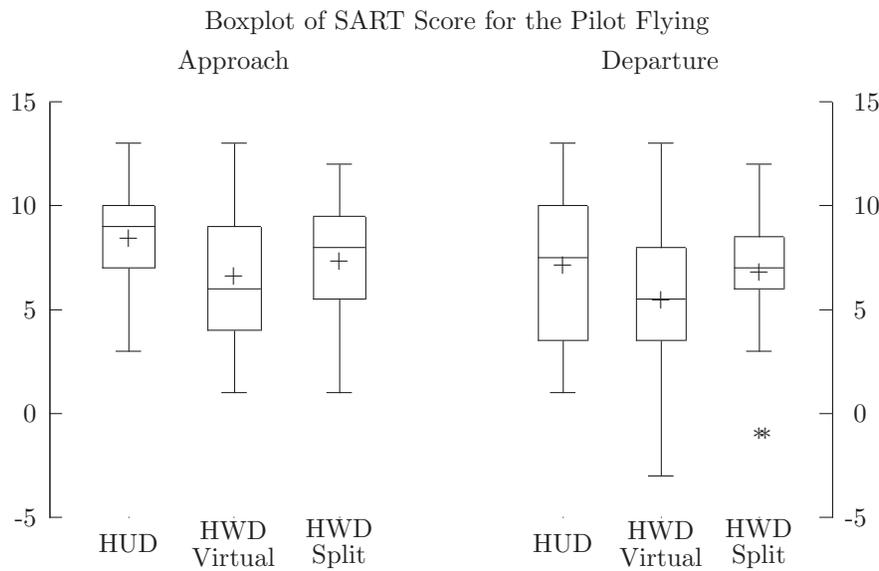


Figure 12. Box plot of the SART scores for the pilot flying (PF).

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

The results for SART demonstrate that pilots did not report any significant differences in situation awareness (based on SART ratings) across display concepts for either the approach ( $F(2, 69) = 0.879, p = 0.420$ ); or the departure ( $F(2, 69) = 0.735, p = 0.483$ ).

#### 3.2.2 Workload

Workload was assessed via the NASA TLX and the AFFTC 7-point subjective workload scale. The TLX scores across the 6 components were averaged to determine workload rating.

No significant differences were found between the PF and the PM for any of the dependent measures ( $p > 0.05$ ) and were collapsed across role. For the approach for mental workload, 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, the ANOVA results evinced no significant differences for NASA TLX,  $F(2, 69) = 0.905, p = 0.195$ ; or AFFTC,  $F(2, 69) = 1.672, p = 0.195$ . Figures 13 and 14 shows the pilots’ responses for the NASA TLX and AFFTC workload.

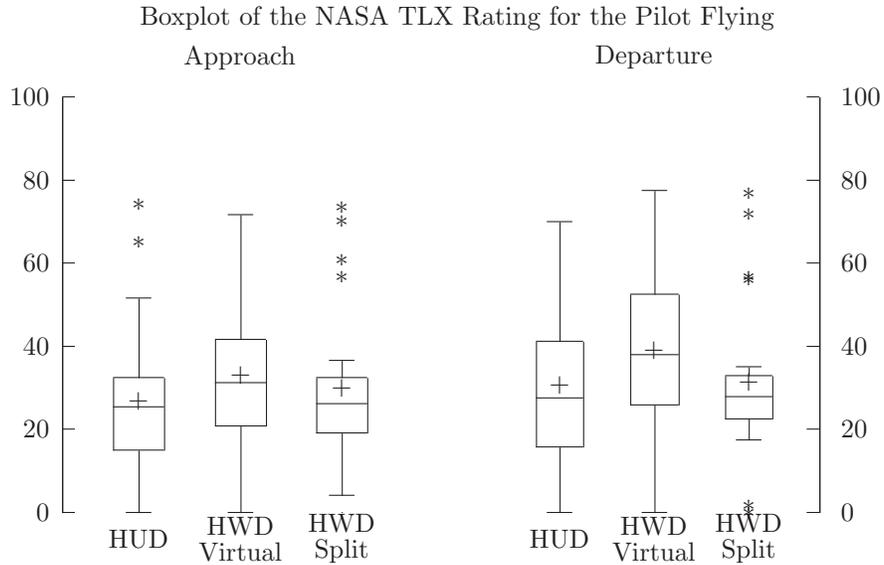


Figure 13. Box plot of the NASA TLX rating for the pilot flying (PF).

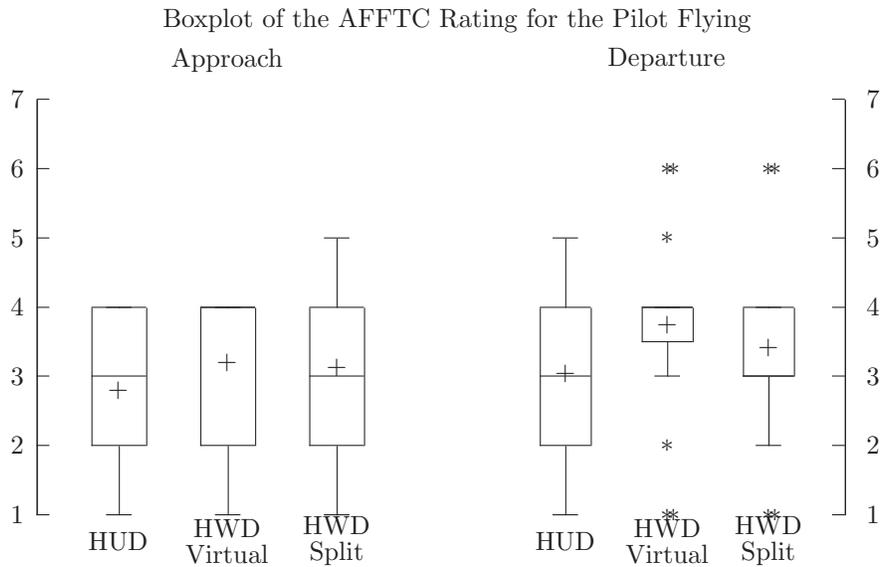


Figure 14. Box plot of the AFFTC rating for the pilot flying (PF).

### 3.3 Simulation sickness

One concern with HWD systems is latency induced sickness. 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. The 6 scores were equally distributed among the 3 display conditions and all were “slight symptoms.”

#### 3.3.1 Post-run questions

Ten questions were presented to crews who were asked for their level of agreement on a 7-point Likert scale where 1 was “strongly disagree” and 7 was “strongly agree.” For the approach, no significant differences were found

for any post-run questionnaire items for display comparisons ( $p > 0.05$ ).

Of particular interest to this paper, the crews were asked on one question if the HWD (after flying the HUD) was equivalent for use during the approach/departure as the HUD. The data, shown in Figure 15 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. However, there were two outliers (ratings) of “disagree” and “slightly disagree” for the approach task and several in the departure task. 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 FLIR image would make it difficult to track the centerline on takeoff.

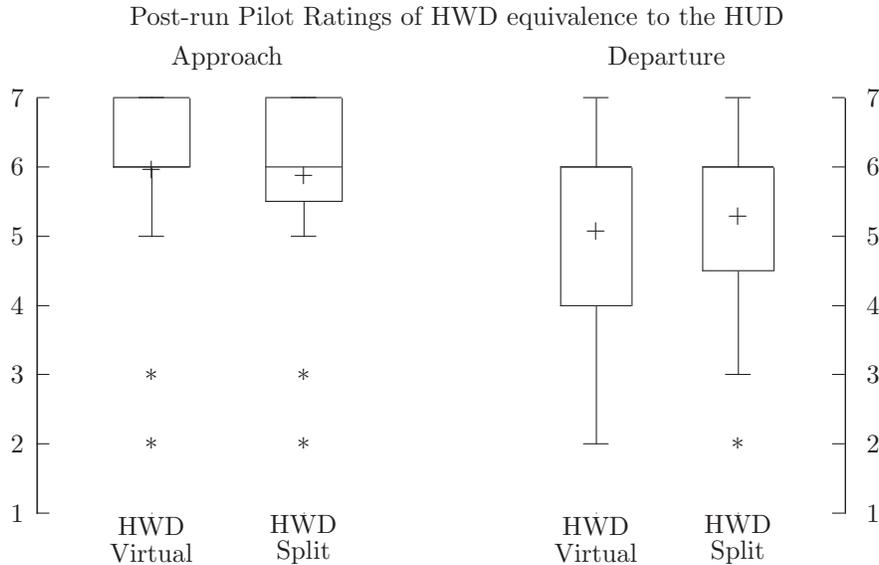


Figure 15. Box plot of pilot ratings of HWD equivalence to the HUD.

One objective of the departure trials was to evaluate the addition of traffic symbology and enhanced vision to determine whether these advanced features would significantly enhance traffic and hazard awareness. The data was demarcated between trials of displays with either (a) no advanced features (Baseline) or (b) enhanced vision and traffic diamonds (EV+TD). The statistical results evince that the presence of traffic diamonds and enhanced vision on any display significantly enhances traffic and other vehicle awareness,  $F(1, 70) = 7.971, p = 0.006$ ; perceived safety,  $F(1, 45) = 4.33, p = 0.043$ ; and detection of potential surface conflicts,  $F(1, 45) = 9.337, p = 0.004$ . Flight crews also rated “hazard awareness” to be significantly greater under EV+TD trials than baseline trials,  $F(1, 70) = 12.456, p = 0.001$ . No other post-run questionnaire items were found to be significant between EV+TD and baseline conditions ( $p > 0.05$ ).

### 3.4 Post-experiment questionnaire

During a semi-structured verbal debrief session, pilots were asked several questions that were designed to elicit flight crew 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. The present paper is focused on display equivalence of the HUD and HWD concepts and, therefore, only relevant questions are presented here.

Flight crews were asked to provide pairwise ratings of “display equivalence” between HUD and HWD concepts. Geo-means were calculated based on the ratings and subsequent parametric statistics<sup>19,20</sup> 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 trials and generalized to real-world operations,  $F(4, 44) = 1.062, p = 0.387$ .

#### 4. DISCUSSION OF RESULTS

The main goal of this study was to determine if a HWD system could provide equivalent performance as a HUD. Two main tasks were performed by EPs in low visibility conditions using either the HWD or the HUD. Symbology was held constant over both displays so that a comparison between the HWD and HUD could be made with little confounds. Both the quantitative and the qualitative results support the hypothesis that a HWD can be equivalent to a HUD.

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’ RVR) may have contributed to the slightly higher sink rates;<sup>21</sup> however, Kramer et al.<sup>22</sup> 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.

There were no significant differences in flying the straight-in approach in 1000’ 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 msec) as any slight head movement would manifest itself into an apparent symbology oscillation/misalignment. 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. Others have concluded that the helmet-mounted display latency requirements are: 50 msec preferred, 100 msec marginal, 150 msec unacceptable.<sup>23</sup> However, for HUD equivalence on a high-resolution, large field-of-view display, the total system latency may need to be on the order of 20 msec for pilot acceptability.<sup>11</sup>

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.

Comparing the HWD-Virtual to the HWD-Split condition, there were no statistical differences between the display concepts 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 FLIR 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 would be less attractive for certification as the clutter/obscuration issue would have to be researched and resolved. 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 FLIR imagery and traffic symbology was. As expected, the addition of FLIR and traffic symbology allowed crews to monitor traffic not visible out-the-window because of visibility conditions.

## 5. 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 showed that there were no statistical differences in the crews performance in terms of touchdown and takeoff. Further, there were no statistically significant differences between the HUD and HWD in pilots’ responses to questionnaires. Future work includes flight testing an HWD system, determining latency requirements, the effects of turbulence and HWD symbology optimization. Though a HWD-Virtual concept seems intuitive for fastest certification and acceptance, the latency requirements will need to be understood before full certification of a HWD system could be realized. NASA will continue to research the advantages of an unlimited field-of-regard HWD. Previous NASA studies have shown the potential advantage of a HWD for surface operations but the commercial flight envelope has not been fully explored. The research in this paper investigated straight-in approaches. Future research will include complex curved approaches in low visibility conditions.

## REFERENCES

1. J. Arthur III, L. Prinzel III, S. Williams, and L. Kramer, “Synthetic vision enhanced surface operations and flight procedures rehearsal tool,” in *Enhanced and Synthetic Vision Proceedings of SPIE*, J. G. Verly and J. J. Guell, eds., **6226**, p. 62260I, SPIE, (Bellingham, WA), 2006.
2. J. Arthur III, L. Prinzel III, K. Shelton, R. Bailey, L. Kramer, S. Williams, and R. Norman, “Design and testing of an unlimited field-of-regard synthetic vision head-worn display for commercial aircraft surface operations,” in *Enhanced and Synthetic Vision 2007*, J. G. Verly and J. J. Guell, eds., **6559**, SPIE, (Bellingham, WA), April 2007.
3. J. J. Arthur III, L. Prinzel III, R. Bailey, K. Shelton, S. Williams, L. Kramer, and R. Norman, “Head-worn display concepts for surface operations for commercial aircraft,” Tech. Rep. NASA/TP-2008-215321, NASA Langley Research Center, Hampton, VA, June 2008.
4. W. Albery, “Multisensory cueing for enhancing orientation information during flight,” *Aviation, Space, and Environmental Medicine* **78(5)**, pp. 186–190, 2007.
5. T. Frey and H. Page, “Virtual HUD using an HMD,” in *Helmet- and Head-Mounted Displays VI*, R. J. Lewandowski, L. A. Haworth, H. J. Girolamo, and C. E. Rash, eds., **4361**, SPIE, (Bellingham, WA), April 2001.
6. R. E. Bailey, “Head-up display (HUD) lessons learned for helmet-mounted display (HMD) development,” in *Helmet- and Head-Mounted Displays and Symbology Design Requirements*, R. J. Lewandowski, W. Stephens, and L. A. Haworth, eds., **2218**, SPIE, (Bellingham, WA), June 1994.
7. FSF, “Head-up guidance system technology - a clear path to increasing flight safety,” tech. rep., Flight Safety Foundation, Alexandria, Virginia, USA, November 2009.
8. FAA, “Enhanced flight vision systems,” Advisory Circular 90-106, US Department of Transportation, Federal Aviation Administration, Washington DC, USA, June 2010.
9. R. E. Bailey, L. J. Kramer, and S. P. Williams, “Enhanced vision for all-weather operations under NextGen,” in *Enhanced and Synthetic Vision 2010*, J. J. Güell and K. L. Bernier, eds., **7689**, SPIE, (Bellingham, WA), April 2010.
10. J. Arthur III, S. Williams, L. Prinzel III, L. Kramer, and R. Bailey, “Flight simulator evaluation of display media devices for synthetic vision concepts,” in *Helmet- and Head-Mounted Displays IX: Technologies and Applications*, C. E. Rash and C. E. Reese, eds., **5442**, pp. 213–224, SPIE, (Bellingham, WA), 2004.

11. R. E. Bailey, J. J. Arthur III, and S. Williams, "Latency requirements for head-worn display s/evs applications," in *Enhanced and Synthetic Vision 2004*, J. G. Verly, ed., **5424**, pp. 98–109, SPIE, (Bellingham, WA), April 2004.
12. R. E. Bailey, J. J. Arthur III, S. P. Williams, and L. J. Kramer, "Latency in visionic systems: Test methods and requirements," in *RTO HFM Workshop on "Toward Recommended Methods for Testing and Evaluation of EV and E/SV Based Visionic Devices"*, **RTO-MP-HFM-125**, NATO, (held in Williamsburg, Virginia), April 2004.
13. RTCA, "Minimum aviation system performance standards (MASPS) for enhanced vision systems synthetic vision systems, combined vision systems and enhanced flight vision systems," Tech. Rep. DO-315, RTCA, Inc., December 2008.
14. R. M. Taylor, "Situational awareness rating technique (SART): The development of a tool for aircrew systems design," in *AGARD Conference Proceedings No 478, Situational Awareness in Aerospace Operations*, pp. 3–1 – 3–37, (Aerospace Medical Panel Symposium, Copenhagen), October 1990.
15. L. L. Ames and E. J. George, "Revision and verification of a seven-point workload estimate scale," Tech. Rep. Technical information manual Jan-Jun 92, Air Force Flight Test Center Edwards AFB, July 1993.
16. S. G. Hart and L. E. Staveland, "Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research," in *Human mental workload*, P. A. Hancock and N. Meshkati, eds., pp. 139–183, Elsevier, (Amsterdam, The Netherlands), 1988.
17. 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*, pp. 203–220, 1993.
18. 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.
19. T. L. Saaty, *The analytical hierarchy process*, McGraw-Hill, New York, 1980.
20. M. A. Vidulich and E. R. Hughes, "Testing a subjective metric of situation awareness," in *Human Factors Society 35th Annual Meeting*, pp. 1307–1311, Human Factors Society, (Santa Monica, CA), 1991.
21. C. R. W. Gibb, D. R. Gray, and D. D. M. Regan, *Aviation Visual Perception: Research, Misperception and Mishaps*, Ashgate Publishing, Ltd., 2012.
22. L. Kramer, S. Williams, J. Arthur III, S. Rehfeld, and S. Harrison, "Motion-base simulator evaluation of an aircraft using an external vision system," in *31st Digital Avionics Systems Conference*, DASC, (Williamsburg, VA), 2012.
23. N. Link, R. Kruk, D. McKay, S. Jennings, and G. Craig, "Hybrid enhanced and synthetic vision system architecture for rotorcraft operations," in *Enhanced and Synthetic Vision 2002*, **4713**, pp. 190–201, SPIE, 2002.