



**Federal Aviation
Administration**

DOT/FAA/AM-20/07
Office of Aerospace Medicine
Washington, DC 20591

Human Factors Considerations in Using HUD Localizer Takeoff Guidance in Lieu of Currently Required Infrastructure

Daniela Kratchounova¹

Larry Miller²

Inchul Choi³

Mark Humphreys⁴

Theodore C. Mofle³

Blake L. Nesmith³

¹ FAA Civil Aerospace Medical Institute

6500 South MacArthur Blvd.

Oklahoma City, OK 73169

² FAA Flight Technologies & Procedures Division, Flight Research & Analysis Group

6500 South MacArthur Blvd.

Oklahoma City, OK 73169

³ Cherokee Support, Services & Solution, LLC

6500 South MacArthur Blvd.

Oklahoma City, OK 73169

⁴ Precision Approaches

7 Blazing Star

Irvine, CA 92604

September 2020

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents thereof.

This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications Web site:
(www.faa.gov/go/oamtechreports)

| | | | |
|--|--|--|---------------------------------|
| 1. Report No. DOT/FAA/AM-20/07 | 2. Government Accession No. | 3. Recipient's Catalog No. | |
| 4. Title and Subtitle Human Factors Considerations in Using HUD Localizer Takeoff Guidance in Lieu of Currently Required Infrastructure | | 5. Report Date September 2020 | 6. Performing Organization Code |
| | | 8. Performing Organization Report No. | |
| 7. Author(s) Kratchounove, D, ¹ Miller, L, ² Choi, I ³ , Humphreys, M ⁴ , Mofle, T ³ , Nesmith, B ³ | | 10. Work Unit No. (TRAIS) | |
| 9. Performing Organization Name and Address ¹ FAA Civil Aerospace Medical Institute ² FAA Flight Technologies & Procedures Division, Flight Research & Analysis Group ³ Cherokee Support, Services & Solution, LLC ⁴ Precision Approaches | | 11. Contract or Grant No. | |
| | | 13. Type of Report and Period Covered | |
| | | 14. Sponsoring Agency Code | |
| 12. Sponsoring Agency name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591 | | 15. Supplemental Notes | |
| 16. Abstract <p>The purpose of this research was to examine the human factors considerations for using HUD with localizer guidance symbology in lieu of currently required infrastructure for lower than standard takeoff minima and within the larger conceptual framework of far domain (runway) and near domain (flight deck) visual cues, HUD guidance symbology, and RVR visibility. To identify the differential contributions of these factors, three baseline conditions without HUD localizer guidance symbology and two conditions with HUD localizer guidance symbology were used. Currently, only about 30% of the CAT I runways in the NAS are equipped with CLL. Therefore, the human factors considerations in using HUD localizer guidance in lieu of CLL in low visibility conditions were of principal interest. The results of this study have the potential to inform operational credit changes that would allow more reduced visibility takeoffs and increase the number of viable airports available for takeoff under low visibility conditions. The research was conducted on a Boeing 737-800NG Level D simulator at the FAA Flight Technologies and Procedures Division facility in Oklahoma City, Oklahoma.</p> | | | |
| 17. Key Words Keywords: Head-up Display (HUD), Aviation Human Factors, Crew Workload; Flight Technical Error | | 18. Distribution Statement Document is available to the public through the Internet: http://www.faa.gov/go/oamtechreports/ | |
| 19. Security Classif. (of this report) Unclassified | 20. Security Classif. (of this page) Unclassified | 21. No. of Pages 48 | 22. Price |

Table of Contents

| | |
|----------------------------------|-----|
| Table of Contents..... | ii |
| Table of Figures..... | iv |
| List of Tables..... | v |
| List of Equations..... | vi |
| List of Acronyms..... | vii |
| Introduction..... | 1 |
| Background..... | 1 |
| HUD Operations Research Gap..... | 7 |
| Method..... | 7 |
| Procedures..... | 10 |
| Results..... | 10 |
| Flight Technical Error..... | 10 |
| <i>Normal Operations</i> | 10 |
| <i>Abnormal Operations</i> | 15 |
| Crew Workload..... | 19 |
| <i>Normal Operations</i> | 19 |
| <i>Abnormal Operations</i> | 21 |
| Discussion..... | 23 |
| Flight Technical Error..... | 25 |
| <i>Normal Operations</i> | 25 |
| <i>Abnormal Operations</i> | 27 |
| Crew Workload..... | 30 |
| Pilot Type..... | 31 |

| | |
|----------------------------------|----|
| Equivalent Level of Safety | 31 |
| Conclusion | 31 |
| Acknowledgements..... | 33 |
| References..... | 34 |

Table of Figures

| | |
|---|----|
| Figure 1. Performance envelope for evaluating takeoff guidance systems (Federal Aviation Administration, 1999)..... | 9 |
| Figure 2. RVR significance within Type of Guidance during Normal Operations. | 13 |
| Figure 3. Type of Guidance significance within RVR during Normal Operations. | 13 |
| Figure 4. Estimated FTE marginal means by Type of Guidance and RVR during Normal Operations. | 14 |
| Figure 5. Estimated FTE marginal means by Type of Guidance during Normal Operations. | 15 |
| Figure 6. Estimated FTE marginal means by RVR during Normal Operations. | 15 |
| Figure 7. Differences observed between RVR within Failure Condition..... | 17 |
| Figure 8. Estimated FTE marginal means by Failure Condition and RVR. | 17 |
| Figure 9. Estimated FTE marginal means by Failure Condition. | 18 |
| Figure 10. Estimated FTE marginal means by RVR during Failure Conditions. | 18 |
| Figure 11. Estimated marginal means of Total NASA TLX Scores for PF by Type of Guidance and RVR..... | 20 |
| Figure 12. Estimated marginal means of Total NASA TLX Scores for PM by Type of Guidance and RVR. | 20 |
| Figure 13. Nested effects of RVR within Failure Conditions..... | 22 |
| Figure 14. Estimated marginal means of Total NASA TLX Scores by Failure Condition and RVR. | 22 |
| Figure 15. Estimated marginal means of Total NASA TLX Scores (Abnormal Operations). | 23 |
| Figure 16. Estimated marginal means of Total NASA TLX Scores by RVR (Abnormal Operations)..... | 23 |

List of Tables

Table 1. Research Matrix 7

Table 2. Results of Normality and Homogeneity Tests 11

List of Equations

| | |
|--|----|
| Equation 1. RMSE Formula for FTE Calculation | 8 |
| Equation 2. Formula for Crosswind Component Calculation..... | 11 |

List of Acronyms

| | |
|----------------|-----------------------------------|
| AIC | Akaike Information Criterion |
| ANCOVA | Analysis of Covariance |
| ANOVA | Analysis of Variance |
| AR | Augmented Reality |
| CAMI | Civil Aerospace Medical Institute |
| CAT I | Category I |
| CAT III | Category III |
| CLL | Centerline Lighting |
| FAA | Federal Aviation Administration |
| FTE | Flight Technical Error |
| GLMM | Generalized Linear Mixed Model |
| HDD | Head-Down Display |
| HGS | Head-Up Guidance System |
| HIRL | High-Intensity Runway Lights |
| HUD | Head-Up Display |
| ILS | Instrument Landing System |
| LOC | Localizer |
| NAS | National Airspace System |
| PF | Pilot Flying |
| PM | Pilot Monitoring |
| RCLM | Runway Centerline Markings |
| RMSE | Root-Mean-Square-Error |
| RVR | Runway Visual Range |

TLX

Task Load Index

Introduction

The purpose of this research was to examine the human factors considerations for using head-up display (HUD) with localizer guidance in lieu of currently required infrastructure for lower than currently authorized visibility minima. In addition, feedback on the equivalent level of safety between using HUD with localizer guidance and using HUD with runway centerline lighting (CLL) was solicited from the pilot participants.

Today, HUDs are standard equipment for many business and commercial aircraft. Their use is encouraged in all phases of flight and mandated by many airlines during critical phases of flight. Research in the past four plus decades suggests that there are a number of advantages in using HUDs. These include:

- Reduced head-down time;
- Reduced time to refocus between flight deck instruments and the external scene;
- Reduced instrument scanning;
- Improved awareness of the outside environment;
- Improved flight performance over traditional head-down displays (HDDs);
- Increased accuracy in flight path tracking;
- Ability to safely perform lower than standard visibility operations;
- Ability to display enhanced, synthetic and combined vision imagery overlaid with conformal symbology, augmenting the external scene in reduced visibility conditions.

Nevertheless, reports of perceptual and cognitive issues with the use of HUD have persisted over the course of several decades. Previous reviews (Crawford & Neal, 2006; Newman, 1995; Weintraub & Ensing, 1992) have provided ample coverage of the longstanding argument about the effects of the perceptual (e.g., misaccommodation) and cognitive (e.g., attention tunneling) issues associated with the use of HUDs reported in the literature. Herein, only a brief background is provided.

Background

HUDs evolved as the original aviation augmented reality (AR) device long before AR was contemplated as a worthy notion in the aesthetic of cool. Looking back, HUDs were developed with

collimated optics to allow the pilot to monitor primary flight information while maintaining visual contact with the outside environment, with the central assumption that the eyes can focus at optical infinity when viewing collimated imagery. Specifically, collimation was intended to position the HUD symbology at the same optical depth as the outside world. In theory, this would assist with eye accommodation and reduce the time necessary to refocus. However, in the 1980s, a repeated experimental observation was reported: when viewing collimated images, instead of automatically focusing at optical infinity, the eyes lapse inward toward their dark focus, or resting accommodation distance (Hull et al., 1982; Iavecchia et al., 1988; Norman & Ehrlich, 1986). Such positive misaccommodation was considered an issue because it could impair pilots' ability to detect targets and judge their distance and size. The supposed perceptual consequence of positive misaccommodation was that the visual scene outside appeared reduced in apparent size. This caused distant objects to appear farther away than they are, and anything below the line of sight (e.g., terrain, runway), to appear higher relative to the horizon (Roscoe, 1985).

After HUDs entered operational use, research studies continued to report that not only did collimation fail to pull the pilot's focus outward to optical infinity, but possibly, it also deepened misaccommodation (Hull et al., 1982; Iavecchia et al., 1988; Norman & Ehrlich, 1986). When using a HUD, about one third of the pilots reported experiencing an increased tendency toward spatial disorientation (Barnette, 1976; Jarvi, 1981; Newman, 1980). Fischer, Haines, and Price (1980) noted instances when pilots failed to attend simultaneously to both the HUD symbology information and the outside world. Pilots also reported a tendency to focus at the near distance, on the HUD combiner glass, instead of on the external scene (Jarvi, 1981). Furthermore, pilots found it necessary to use volitional accommodation when shifting attention between HUD symbology and the outside environment (Morey & Simon, 1993). The HUD myopia seemed to act as a special case of a phenomena identified as instrument myopia (Hennessy, 1975).

In the late 1980s and early 1990s, Newman (1987) and Weintraub and Ensing (1992) contended that there was a substantial amount of evidence supporting the notion of collimated HUDs actually pulling pilot's focus outward, even if it was not always to optical infinity. They argued that earlier research findings were an artifact of the display technology available at that time, and might be attributable to the poor optical quality of HUD and external scenes imagery used in those prior studies. They maintained that if the external visual cues were of poor quality

because of reduced visibility, for example, high-quality HUD images would actually pull the pilots' focus outward, and consequently, partially offset the tendency for the point of accommodation to be closer than the objects in the external environment. In addition, as reported by Foyle et al. (1993), the misaccommodation argument had fallen short of explaining the failure to process simultaneously the outside world information and HUD-like symbology information when using non-collimated displays. They suggested that these results could be attributed to attentional rather than to visual factors. That is, pilots actually switch attention back and forth between the HUD and the external scene instead trying to perceive both at the same time.

Nonetheless, results from a number of studies suggested that there was a potential issue with switching attention between the internal and external scene when using HUD, as well. More specifically, that HUD symbology could capture pilots' attention and impair their ability to detect events in the external environment. This effect has been referred to as cognitive tunneling or cognitive capture (Foyle et al., 1993; Prinzel III & Risser, 2004). Problems associated with cognitive tunneling seemed to revolve around pilots' ability to be effective in switching attention between the HUD and other elements in the same visual scene. As opposed to switching attention between the HDDs and the outside world, switching attention between the HUDs and the outside world involved mentally shifting attention between stimuli within the same visual space. Namely, the latter did not involve a change in visual accommodation (Haines et al., 1980; Weintraub et al., 1984, 1985). Rather, cognitive capture occurred when pilots failed to switch attention and instead fixated on one element of HUD symbology at the expense of other information, either on the HUD or in the external scene.

Previous research observed an altitude/path performance trade-off and proposed two models of visual/spatial attention as potentially viable explanations for it: object-based and location-based (Foyle et al., 1991). The object-based attentional model predicted that the altitude/path performance trade-off should not be affected by a reduction in the distance between altitude and path information as presented on the display. The location-based attentional model suggested that efficient concurrent processing of two separate information sources was only possible when both sources were within the same attentive field. Later, to study the differences between the object-based and location-based attentional models of the altitude/path performance trade-off, Foyle et al. (1993) conducted two experiments. The results indicated that synchronized

processing of both the HUD and the outside world information occurred only in conditions that required visual/attentional scanning. The authors argued that due to its volitional nature, visual scanning in a pattern following known locations of information might be more efficient in alleviating the altitude/path performance trade-off. Particularly, when gathering specific information was required and participants were prepared to process it; attentional tunneling was eliminated by scanning for information in a particular predefined order.

McCann, Lynch, Foyle, and Johnston (1993) and McCann, Foyle, and Johnston (1993) examined the potential cost of attention switching. The results showed that differential motion between symbology with fixed position on the HUD and the optical flow of the world outside was a possible source of attentional tunneling. They found that the differential motion led to increased attention-switching times. Shelden, Foyle, and McCann (1997) and Levy, Foyle, and McCann (1998) proposed that a set of symbology that appears to be a part of the outside world, i.e., to be conformal to it (Wickens & Long, 1994), would alleviate the tunneling. Their results supported the notion that scene-linked symbology yielded improved performance and that the performance advantage is attentional. That is, attention can be divided between scene-linked symbology and the outside world (Levy et al., 1998). Furthermore, the authors made a distinction between two separate methods that could make the symbology look as if it were an integral part of the outside world: scene enhancement and scene augmentation. Scene enhancements would add more saliency to features of already existing objects in the outside world, where scene augmentations would involve the addition of virtual objects that are otherwise non-existent in the real world. In both cases, the HUD symbology would be moving in unison with the external scene. The results further indicated that making the HUD symbology conformal to its outside counterpart to form a single perceptual object, along with flight task integration, supported synchronized processing which in turn yielded performance advantages and had the potential to mitigate the problem of attentional tunneling (Shelden et al., 1997).

Wickens and Long (1994) examined the impact of positioning conformal and non-conformal symbology in head-up or head-down locations to determine the comparative influences of three factors: space-based cost of clutter, space-based cost of scanning, and object-based benefit of conformality. Their results not only aligned with prior research findings about HUDs' significant contribution to safety, but also reinforced the notion that the negative impact of clutter

was very real. The authors strongly cautioned against adding too much non-conformal symbology to HUDs.

We could not find a comprehensive review of the literature on the direct effects of clutter, and more specifically, the effects of the amount of information when presented on HUDs. Nevertheless, a large body of research has addressed the subject of display clutter in general. In summary, preserving the most relevant and unambiguous visual cues pilots use is an art form. It could be successfully accomplished through enhancement, augmentation, task integration, and synchronization of those visual cues in the near and far domain. However, if overdone, the intended benefits might very well be nullified by the resulting clutter (Boston & Braun, 1996; Kaber et al., 2008, 2013; Ververs & Wickens, 1998).

So far, the results from applied research literature on the subject of why it might be difficult for pilots to monitor primary flight information while maintaining visual contact with the outside environment are remarkably inconclusive. Interestingly, findings from prior basic research by Neisser and Becklen (1975) foreshadowed the results of Foyle et al. (1993) mentioned earlier. The authors asserted that selective attention was a direct consequence of skilled perception and did not require any special mechanisms to reject unwanted information. They reported that without any prior practice, it was easy to attend to one dynamically changing scene/episode and ignore another, even when the two were presented in transparent visual overlap. Moreover, they concluded that what outlined one episode and differentiated it from the other was not its location or clarity, but its inherent relevance, structure, and the continuous and coherent sequence of motion. In other words, in order to follow one dynamically changing event rather than another, one must have already made a decision to attend the former rather than the latter. Neisser and Becklen (1975) did not address the question of whether selective attention would have improved had the participants been given the opportunity to practice. Later, Spelke, Hirst, and Neisser (1976) conducted research on complex unimodal (e.g., visual) dual-task performance and found that better performance could be achieved if participants were allowed to practice the tasks in combination. However, two questions remained unanswered: how to combine tasks successfully, and specifically, what aspects of the tasks would make their combination successful in terms of performance. The case of transparent visual overlap of dynamically changing events is a rare and unnatural occurrence for which the human visual system is “unprepared.” Training to follow two simultaneous dynamic episodes that

are visually overlapped, even if they are related, should therefore encompass strategies that allow them to be combined more easily. Littman and Becklen (1976) in effect replicated Neisser and Becklen (1975) with one exception: participants were required to fixate on a spot in the center of the visual-field during 50% of the trials. The results demonstrated that selective looking could be accomplished without eye movements, and that it was not more difficult to follow either a single event, or one of two superimposed events, when such fixation was required. A similar strategy could help with minimizing the incompatibilities (e.g., movement in the periphery) in complex, dual visual tasks. The authors concluded that eye movements do not initiate a perceptual act. Rather, they are dependent upon perceptual expectations.

Research on flight performance has shown that without a direct presentation of the flight path guidance, the pilot was required to visually scan the flight deck instruments and mentally transform the information to determine the path of the aircraft (Fischer et al., 1980). HUD symbology typically includes active flight path information (e.g., flight path vector). This allows for a more natural, intuitive control. Numerous early studies have shown that using HUD with flight path symbology produced superior flight path maintenance and landing precision relative to traditional flight director instrumentation (Boucek et al., 1983; Bray, 1980; Lauber et al., 1982). More recent, high fidelity HUD research conducted in applied settings, and in the context of multi-crew commercial flight operations, is a research area that could benefit from more study.

Training is often suggested as a solution to human factors issues. Nevertheless, it is almost never the sole remedy for inadequate design. Therefore, it is plausible to expect that to complement the continuous improvements of HUDs design over the years; robust training curricula would help pilots learn how to monitor primary flight information on a HUD while maintaining visual contact with the outside environment, in an efficient and effective fashion. In theory, such training would also improve their awareness of the potential for attention capture and ultimately, help them counter it at its onset. It is unknown however, whether it is possible to train pilots to overcome the effects of cognitive tunneling when using HUD, or how much training would be needed if it were possible to do so.

Personal correspondence containing observations from a pilot subject matter expert indicated that almost all pilots new to HUD focus on the flight path vector only for the first 2-3

hours of HUD exposure. In the next 4-6 hours, they start to open their scan to the other information displayed on the HUD. From 7 hours on, they have the ability to move their gaze back and forth between the HUD symbology and the outside world. It is unknown if this gaze shifting can be taught early in a training program. Nevertheless, it seems to be a function of time and exposure to the HUD, before the brain can “gaze shift” (M. Humphreys, personal communication, January 6, 2020).

HUD Operations Research Gap

Presently, Instrument Flight Rules for lower-than-standard takeoff minimums allow commercial aircraft takeoffs as low as 300ft Runway Visual Range (RVR) using High-Intensity Runway Lights (HIRL), CLL, and HUD takeoff guidance symbology. This type of commercial operation is only approved for runways with the lowest Category III (CAT III) Instrument Landing System (ILS) minima and the associated runway infrastructure, such as additional RVR sensors, touchdown zone lighting, and lead on/off lights. This limits 300ft RVR takeoff operations to only a few airports and runways in the National Airspace System (NAS). To expand this capability to additional airports and potentially increase the NAS throughput, the FAA Civil Aerospace Medical Institute (CAMI) [Aerospace Human Factors Research](#) Division assessed HUD localizer-guided low-visibility takeoffs using a Category I (CAT I) localizer and reduced airport infrastructure.

Method

Twenty-four pilot crews participated in this research: 12 airline crews and 12 business jet crews, who were deemed proficient in using a HUD. For normal operations scenarios, five levels of Type of Guidance, three levels of RVR, and two levels of Lighting conditions were examined (Table 1).

Table 1. Research Matrix

| Type of Guidance | RVR (ft) | | |
|------------------|----------|-----|-----|
| | 300 | 500 | 700 |
| | | | |

| | | | | | | |
|--|-----|-------|-----|-------|-----|-------|
| Baseline 1: HUD, No LOC ¹ guidance, RCLM ² only | Day | Night | Day | Night | Day | Night |
| Baseline 2: HUD, No LOC guidance, CLL ³ | Day | Night | Day | Night | Day | Night |
| Baseline 3: No HUD, CLL | Day | Night | Day | Night | Day | Night |
| Condition 1: HUD, LOC guidance, RCLM only | Day | Night | Day | Night | Day | Night |
| Condition 2: HUD, LOC guidance, No RCLM, No CLL | Day | Night | Day | Night | Day | Night |

For normal operations, wind speeds ranging between 3 knots (calm) and 22 knots⁴ and varying directions were randomly assigned to scenarios. For abnormal operations (failure cases), winds between 3 knots (calm) and 15 knots were applied. All tailwinds were limited to 10 knots (Boeing 737-800NG Airplane Flight Manual Limitation). Six abnormal operational conditions were included in this research, as well. They encompassed engine malfunctions, failure of the localizer transmitter, slewing of the localizer azimuth, and loss of HUD. All experimental trials were conducted on a CAT I runway at Memphis International Airport.

For the purposes of this study, deviation from runway centerline was sampled at 5 Hz. To calculate the Flight Technical Error (FTE) scores, a root-mean-square-error (RMSE) of deviations from centerline was calculated (*see* Equation 1).

Equation 1. RMSE Formula for FTE Calculation

$$FTE_{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{Deviation from Centerline})^2}{n}}$$

¹ LOC = localizer

² RCLM = Runway Centerline Markings

³ All CLL conditions assume existing RCLM

⁴ AC 120-28D - Criteria for Approval of Category III Weather Minima for Takeoff, Landing, and Rollout

where n = Number of samples recorded for each scenario

FAA AC 120-28D (Federal Aviation Administration, 1999) requires demonstration of specific performance when using a takeoff guidance system. As it pertained to this research, the RMSE of deviation from runway centerline was not to exceed 7m (~23ft) or 14m (~46ft), depending on the phase of the takeoff roll. For example, for an all engines takeoff, the maximum deviation from centerline should not exceed 7m (~23ft) from the beginning of the takeoff roll to liftoff (V_{LOF}), and 14m (~46ft) from V_{LOF} to 35ft Above Ground Level (Figure 1).

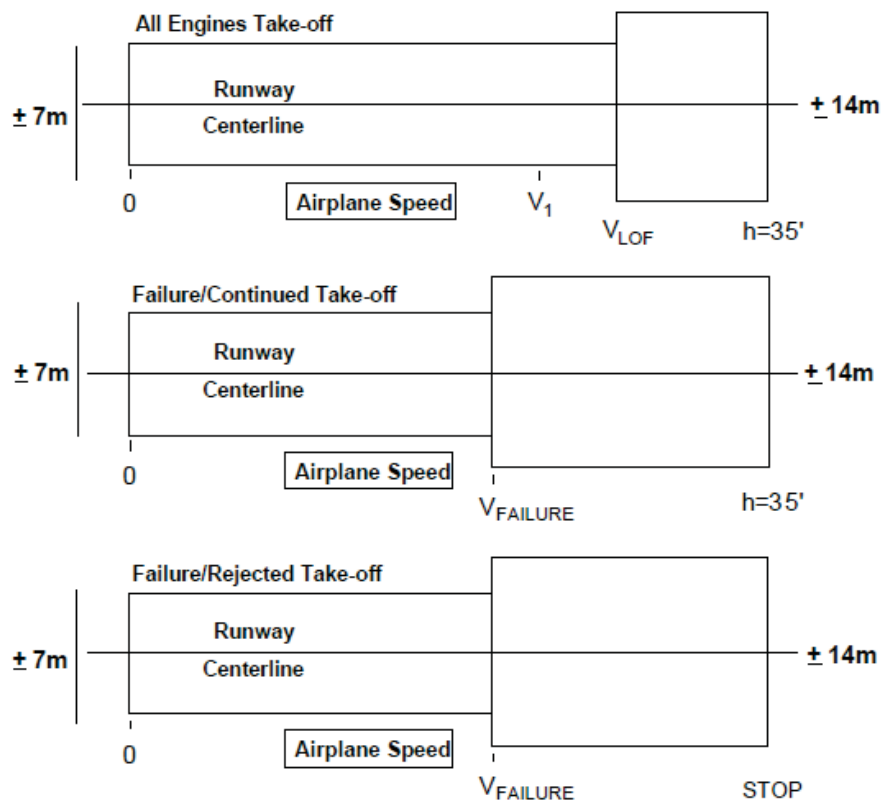


Figure 1. Performance envelope for evaluating takeoff guidance systems (Federal Aviation Administration, 1999).

NASA Task Load Index (TLX) rating scale was used to assess crew workload (Hart, 1986). Specifically, both Pilot Flying (PF) and Pilot Monitoring (PM) completed the pen-and-paper version of the NASA TLX questionnaire after each experimental run. In addition, after each run, pilots were

encouraged to provide additional verbal feedback regarding the level of workload they experienced, the personal techniques they applied when using the HUD symbology, as well as, how they dynamically allocated their attention to monitor primary flight information while maintaining visual contact with the outside environment.

Procedures

Each pilot crew was briefed on the purpose of the research, viewed a video introducing the HUD localizer takeoff guidance symbology implemented on the Rockwell Collins Head-up Guidance System (HGS) Model 6700 as installed on the FAA's Boeing 737-800NG Level D simulator, and received a safety briefing. A 3-hour simulator familiarization session followed. The purpose of the session was to make the pilots familiar and comfortable with the HUD symbology set, flight deck controls, and simulator handling qualities. For the data collection portion of this research, each crew completed 96 takeoff scenarios including 60 normal and 36 abnormal conditions over the course of two days. Each pilot flew 48 scenarios as a PF and 48 as a PM. During the debriefing session, each crew completed the NASA TLX pairwise comparison, and provided feedback regarding the equivalent level of safety of using HUD localizer guided takeoff symbology in lieu of CLL. Specifically, based on their overall experience, the crews reflected on whether it was equally safe to use HUD localizer guidance symbology in lieu of CLL and to what RVR levels. In addition, they made suggestions relating to other conditions/factors (crosswind limitation, additional training, etc.) to be considered in the decision-making process for future operational approval of using HUD with localizer guidance in lieu of currently required infrastructure for lower than standard takeoff minima.

Results

Flight Technical Error

Normal Operations

Crosswind Component

Crosswind Component was considered a covariate in the analysis of FTE scores and NASA TLX with Type of Guidance and RVR as main factors. To calculate a single value for the Crosswind Component for each scenario, the magnitude of wind speed in knots was multiplied by the absolute value of sine of the relative wind direction in degrees from magnetic North ($^{\circ}$) (*see* Equation 2).

Equation 2. Formula for Crosswind Component Calculation

$$XWC = |V \times \sin(\theta)|$$

Where: XWC = Crosswind
Component V = Wind speed (kt)
 Θ = Wind direction ($^{\circ}$)

Analysis Model 1 - ANCOVA

Assumption Tests

To examine the effects of main factors including Type of Guidance and RVR with Crosswind Component as a covariate on FTE scores, a two-way Analysis of Covariance (ANCOVA) was performed. The FTE scores displayed issues with normality and equal variance, violating the assumptions of parametric statistics. To adjust the skew of residuals, a natural log, log10, and square root transformation methods were applied to the FTE scores. However, none of them satisfied the assumptions as shown in Table 2.

Table 2. Results of Normality and Homogeneity Tests

| Model | Levene's test | Kolmogorov-Smirnov test |
|-----------------|-----------------------------------|--------------------------------|
| Raw FTE | $F(719, 720) = 1.393, p < 0.0001$ | $D(1440) = 0.127, p < 0.0001$ |
| Natural Log FTE | $F(719, 720) = 1.172, p = 0.017$ | $D(1440) = 0.300, p = 0.005$ |
| Log10 FTE | $F(719, 720) = 1.172, p = 0.017$ | $D(1440) = 0.300, p = 0.005$ |
| Square root FTE | $F(719, 720) = 1.355, p < 0.0001$ | $D(1440) = 0.068, p < 0.0001$ |

Analysis Model 2 - Generalized Linear Mixed Model

Although a one-way Analysis of Variance (ANOVA) method is generally considered as robust to violations of homogeneity of variance and normality assumptions (Lix et al., 1996; Mena et al., 2017; Pearson, 1931), some researchers have argued that error rates of two-way ANOVA are sensitive to unequal variances, especially with non-normal distributions (Erceg-Hurn & Mirosevich, 2008; Harwell et al., 1992). Notionally, the use of nonparametric statistical techniques could resolve such issues. However, such methods have their own limitations as well. For example, Kruskal-Wallis

tests require the same (e.g., normal or gamma) distributions across all cells in the analysis, while rank transformation is not robust enough for a factorial design (Judd et al., 1995).

Lo and Andrews (2015) contended that the application of transformations can considerably affect results, as well as the theoretical and applied implications of such results. Rather than applying transformations to extremely non-normal data, that still did not meet assumptions; the authors suggested the use of Generalized Linear Mixed Model (GLMM) as a more robust solution to analyze non-normal and heterogeneous data often seen in psychological research. Accordingly, GLMM analysis of repeated measurements was performed to identify the effects from Type of Guidance and RVR on FTE scores. The model used a Gamma probability distribution with a power-link (-0.1) function due to the right skew associated with the data and the lowest Akaike Information Criterion (AIC) value (-4738.898; Dunn & Smyth, 2005; Iyit, 2018; Temple, 2018). Two fixed factors (Type of Guidance and RVR), one random factor (Crew*Pilot), and a covariate (Crosswind Component) were fitted in the FTE model.

The GLMM results indicated a significant interaction between Type of Guidance and RVR ($F(8, 1424) = 16.639, p = < 0.0001$). The covariate, Crosswind Component, was also significant in the model ($F(1, 1424) = 1605.503, p = < 0.0001$) and was assessed at 9.40kt to calculate estimated marginal means for normal operations. Therefore, the nested effects of RVR within Type of Guidance are reported on Figure 2 and those of Type of Guidance within RVR on Figure 3. The line style (outlined in the legend on Figure 2 and Figure 3) indicates significance at different p values and the “greater/less than” sign indicates the directionality of significance of FTE scores between RVR conditions.

For example, in Baseline 1 condition (the first triangle on Figure 2), 300ft RVR and 500ft RVR are connected with a thick red line, and the “larger than” sign indicating that the simple effect for 300ft RVR is significantly larger than for 500ft RVR ($p < 0.001$). The contrast estimate between 300ft RVR and 500ft RVR is 0.864. Similarly, 500ft RVR and 700ft RVR are connected with a black dotted line without a “larger than” sign indicating that there is no significant difference between 500ft RVR and 700ft RVR.

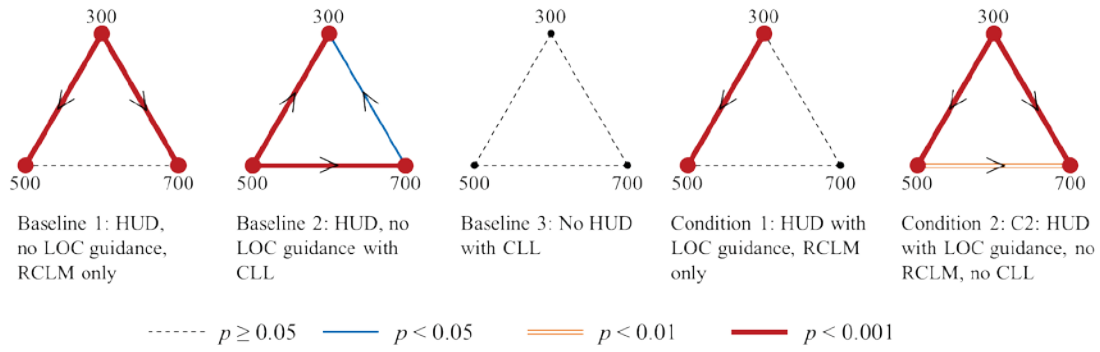


Figure 2. RVR significance within Type of Guidance during normal operations.

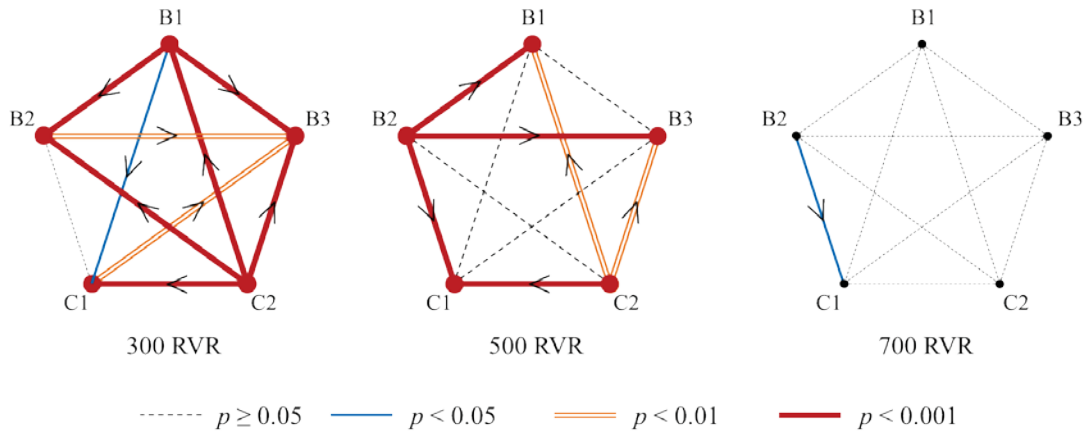


Figure 3. Type of Guidance significance within RVR during normal operations.

The interaction between Type of Guidance and RVR as well as the general trends in the results are shown on Figure 4.

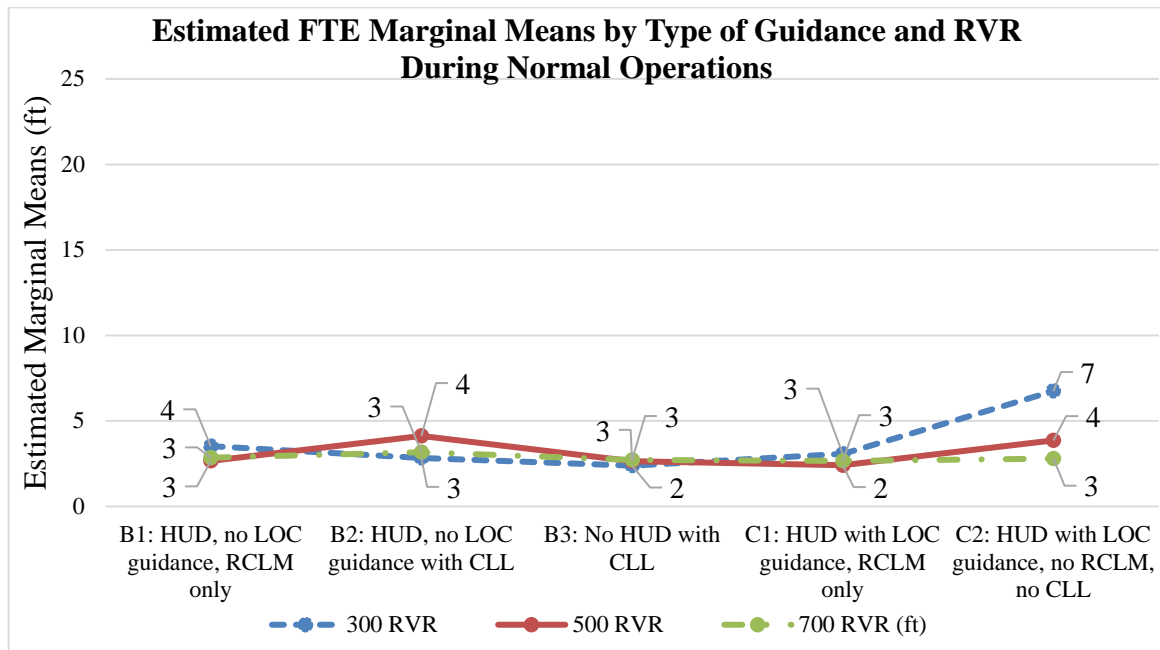


Figure 4. Estimated FTE marginal means by Type of Guidance and RVR during normal operations.

Notably, the overall estimated FTE marginal means were well below the pass/fail criteria established for this research (Figure 1). General trends of Type of Guidance effects are shown on Figure 5. Although an interaction effect was significant in the model, it was worth examining the overall trends of Type of Guidance and RVR. The trends (averaged across RVR conditions) of Types of Guidance showed that Baseline 2 and Condition 2 generally had higher FTE scores compared to other guidance types. Furthermore, trends of the RVR effects are shown on Figure 6. Specifically, there is an inverse relationship between RVR and FTE scores, as FTE scores became larger with lower RVR.

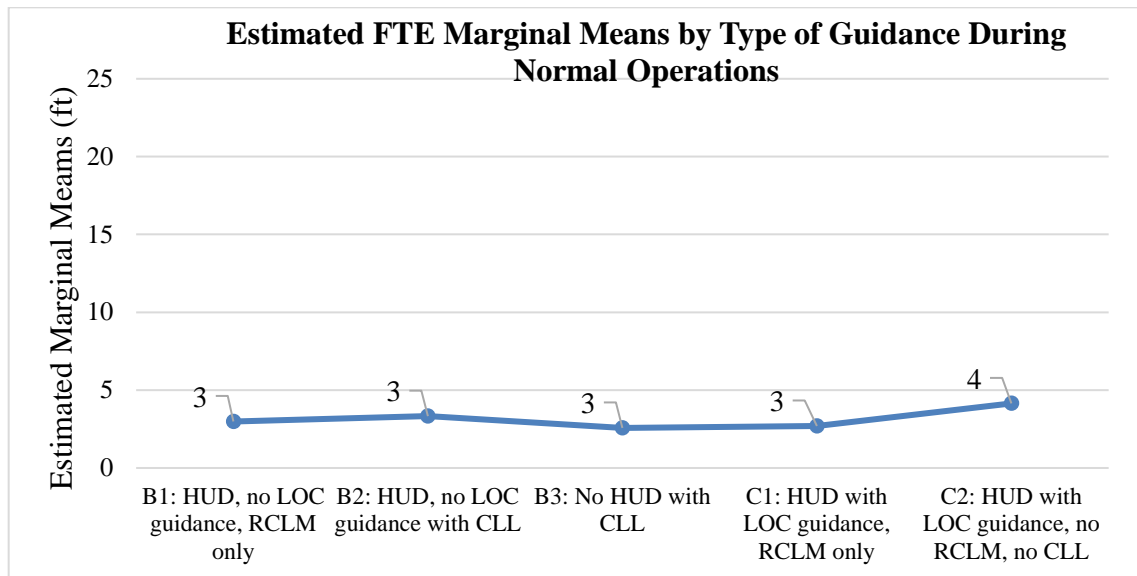


Figure 5. Estimated FTE marginal means by Type of Guidance during normal operations.

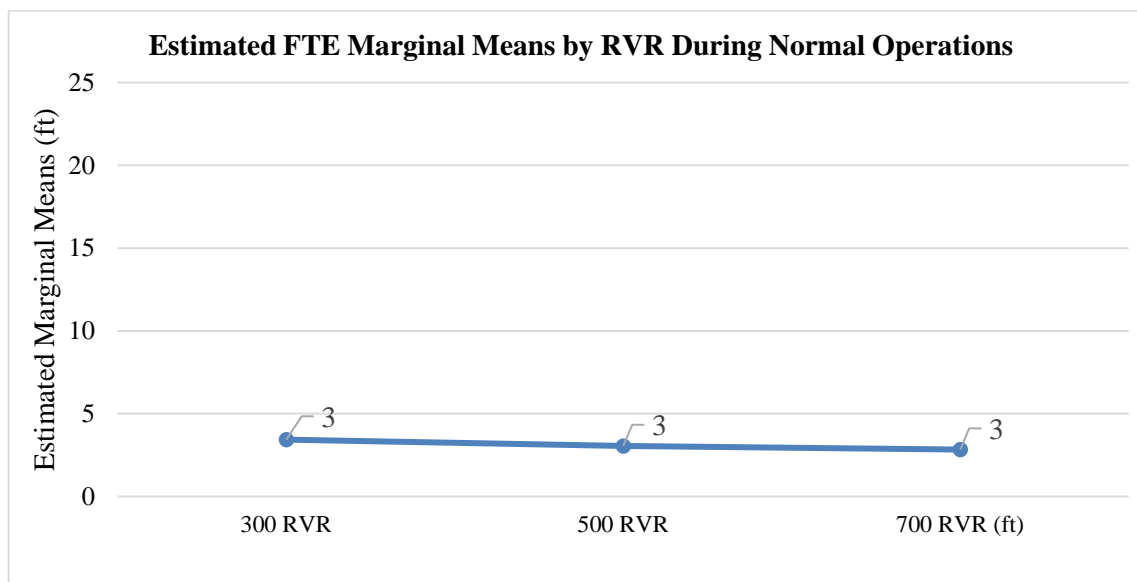


Figure 6. Estimated FTE marginal means by RVR during normal operations.

Abnormal Operations

The following failure conditions were examined for this research:

- Above V_1 Continue: engine malfunction at an IAS above V_1 (122kt), with an expectation to continue the takeoff
- Below V_1 Reject: engine malfunction at an IAS below V_1 (105kt), with an expectation to reject the takeoff
- Below V_{mcg} Reject: engine malfunction at an IAS below V_{mcg} (95kt), with an expectation to reject the takeoff
- LOC Fail: failure of the localizer, removing localizer guidance and the localizer line and scale from the HUD at 70kt
- LOC Bend: slewing of the localizer azimuth at 50 knots, causing a drift of the HGS Guidance Cue
- Loss of HUD at V_R (132kt)

FTE scores in the abnormal conditions had significant issues with normality due to each failure scenario consisting of a unique failure (e.g., engine failure, HUD localizer failure). Similar to the analyses conducted on the normal operations data, and unbound by the assumption of normal distribution, a GLMM model was applied to the abnormal operations data. This model was applied to a Gamma probability distribution with a power-link (-0.1) function and resulted in the lowest AIC value (-2316.632). Two fixed factors (failure scenario and RVR), one random factor (Crew*Pilot), and a covariate (Crosswind Component) were included as well. The results showed that a significant interaction was present between failure scenario and RVR $F(10, 844) = 7.826, p < 0.0001$. The nested effects of RVR for each failure case are reported in Figure 7. The covariate factor, Crosswind Component, was also significant in the model ($F(1, 844) = 19.328, p < 0.0001$), and was assessed at 4.07kt to calculate estimated marginal means for abnormal operations. General trends and the interactions are shown on Figure 8.

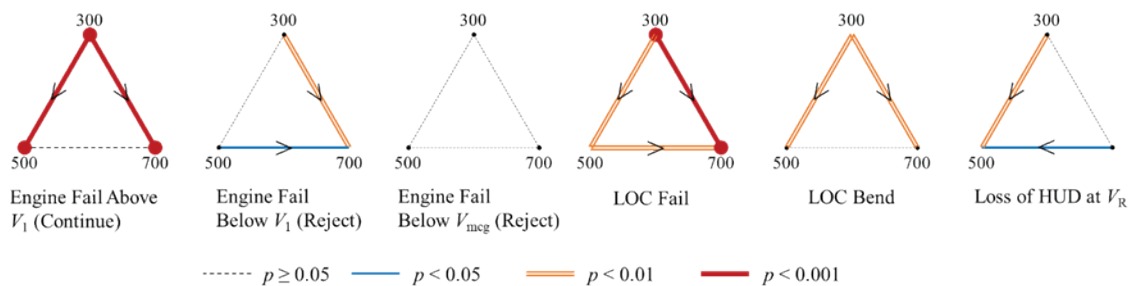


Figure 7. Differences observed between RVR within Failure Condition.

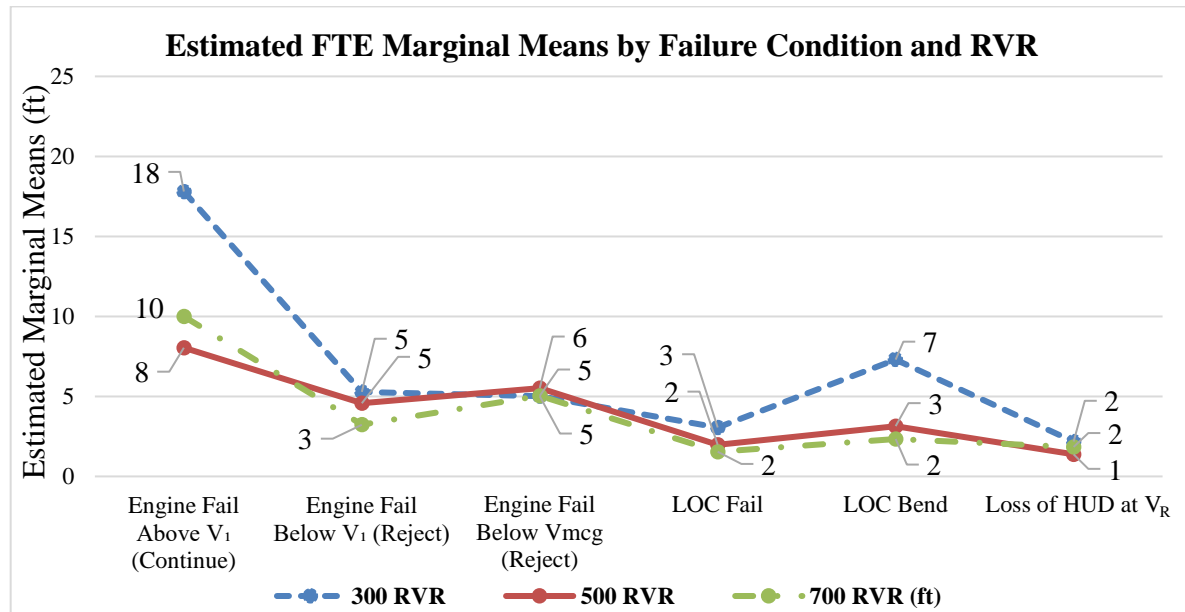


Figure 8. Estimated FTE marginal means by Failure Condition and RVR.

In examining the general trends, it is evident that while the failure conditions are not being statistically compared, the marginal means indicate lower FTE in the loss of HUD, LOC Failure, and LOC Bend conditions (Figure 9).

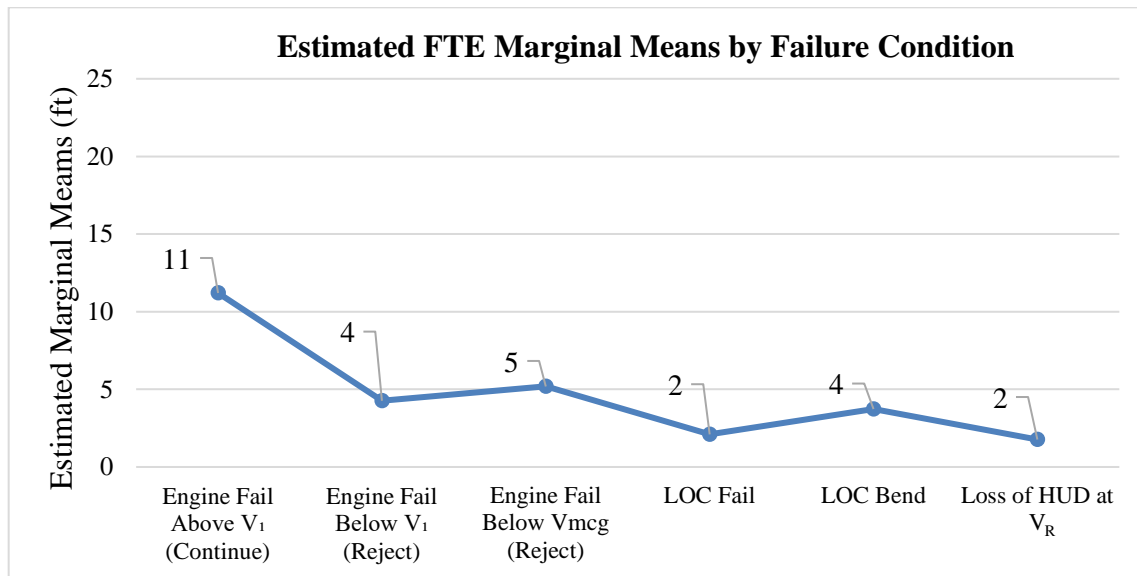


Figure 9. Estimated FTE marginal means by Failure Condition.

Similar general patterns emerged in the results from abnormal operations scenarios as those observed in normal operations with regards to RVR. While almost flat, the FTE scores did increase slightly with lower RVRs (Figure 10).

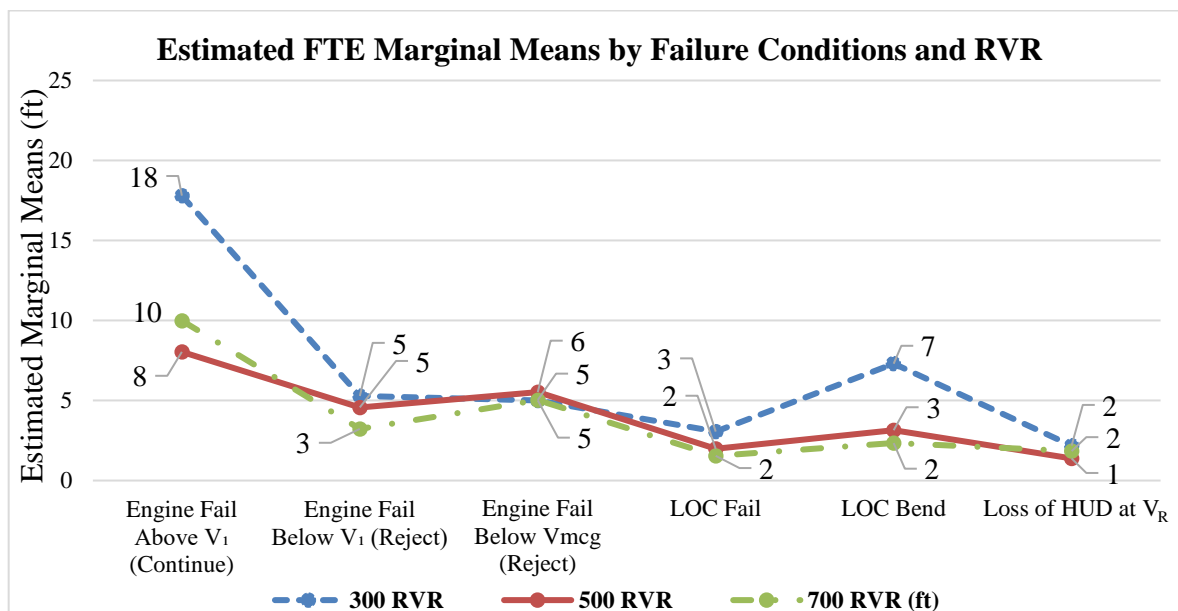


Figure 10. Estimated FTE marginal means by RVR during Failure Conditions.

Crew Workload

Normal Operations

The total NASA TLX scores displayed issues with normality, violating the assumptions needed for parametric statistics. To address these issues and satisfy the assumption of normality, a natural log transformation was applied to the total workload scores. The equal variance assumption was met according to Hartley's F_{\max} test ($F_{\max}(15, 191) = 1.44$) against the critical value of 2.0964 (Kirk, 1995; O'Brien, 1981). Accordingly, statistical results are reported with the natural log transformed data.

An ANCOVA was conducted to evaluate the main effects of Type of Guidance and RVR on crew workload with Crosswind Component as a covariate. The covariate, Crosswind Component, was found to have a significant effect on the total NASA TLX scores ($F(1, 2839) = 93.80, p < 0.0001, \eta^2 = 0.032$), and the coefficient of Crosswind Component was 0.445. Crosswind Component was assessed at 9.40kt to calculate estimated marginal means for normal operations. Analyses of total workload scores revealed no significant interaction effects. Therefore, all factors were evaluated as main effects. Type of Guidance had a significant effect on total workload scores ($F(4, 2839) = 24.876, p < 0.0001, \eta^2 = 0.034$). Bonferroni correction for multiple testing was applied. The results further indicated that in Baseline 1 condition, the crews reported experiencing significantly higher workload than Baseline 2 ($p < 0.0001$) or Condition 1 ($p < 0.0001$). In Condition 2, the crews experienced significantly higher workload than in any other types of guidance condition.

The results showed no significant differences in crew workload between Baseline 2 as compared to Condition 1 indicating levels of crew workload in the conditions of principal concern in this research were not different. Furthermore, results did not show a significant difference in crew workload levels between Baseline 3 compared to Baseline 2 or compared to Condition 1.

In addition, RVR had a significant effect on the total NASA TLX scores ($F(2, 2839) = 71.743, p < 0.0001, \eta^2 = 0.048$). Bonferroni correction for multiple testing was applied, and the results indicated that in the 300ft RVR condition, the crew experienced significantly higher workload than in the 500ft RVR ($p < 0.0001$) and 700ft RVR ($p < 0.0001$) conditions, respectively. In the 500ft

RVR condition the pilots experienced significantly higher workload than in the 700ft RVR ($p = 0.013$) condition (Figure 11 and Figure 12).

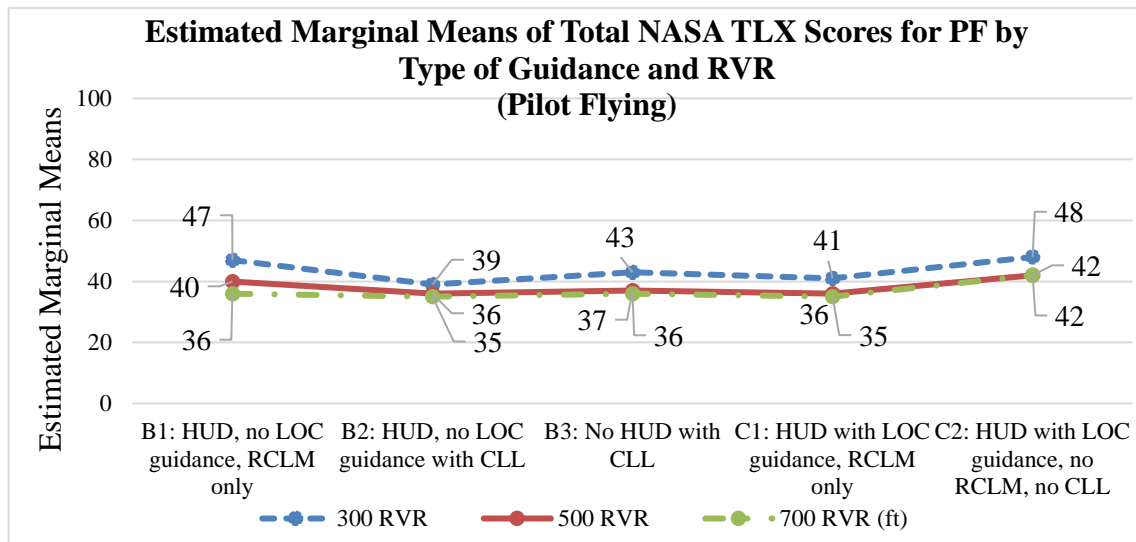


Figure 11. Estimated marginal means of total NASA TLX scores for PF by Type of Guidance and RVR.

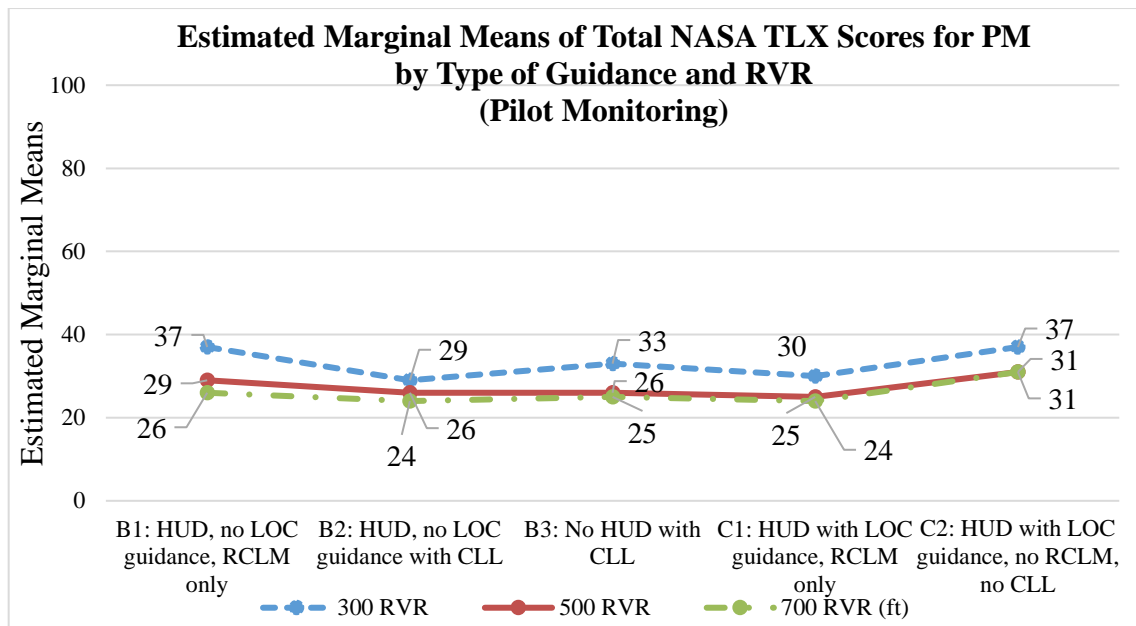


Figure 12. Estimated marginal means of total NASA TLX scores for PM by Type of Guidance and RVR.

In the current model, individual crews and Pilot Type (airline or business jet pilot) were considered mutually inclusive. Therefore, to refine the model, mitigate the subjectivity of the TLX measure, and be able to generalize the results of the total NASA TLX measure to the population. Crew was treated as a random factor. The results showed that Crew had a significant random effect ($F(23, 2839) = 72.063, p < 0.0001, \eta^2 = 0.369$). As expected, Pilot Role (PF/PM) was a significant factor in the model, as well ($F(1, 2839) = 512.206, p < 0.0001, \eta^2 = 0.153$). Parameter estimation indicated a 10.69-point difference between PF and PM, with PM reporting lower levels of workload across all conditions. Lastly, Lighting condition did not have a significant effect on crew workload ($F(1, 2839) = 0.260, p = 0.610$).

Abnormal Operations

Similar to the scores for normal operations conditions, the total NASA TLX scores for abnormal operations violated the normality and equal variance assumptions. Therefore, GLMM was employed here as well. A Gamma distribution with a power-link (-0.1) function produced the smallest AIC value (-7659.699; Dunn & Smyth, 2005; Iyit, 2018; Temple, 2018). Two fixed factors (failure scenario and RVR), two within factors (Lighting condition and Pilot Role [PM/PF]), one random factor (Crew*Pilot), and a covariate (Crosswind Component) were applied. The results of the model indicated a significant interaction between failure condition and RVR ($F(10, 1707) = 2.187, p = 0.016$). The nested effects of RVR within failure conditions are shown on Figure 13. Pilot Role (PM/PF) was a significant within-subjects factor ($F(1, 1707) = 591.964, p < 0.0001$), and total NASA TLX score of PF was significantly higher than those of PM (contrast estimate for PF - PM = 13.007, $p < 0.0001$). Though, Lighting condition ($F(1, 1707) = 0.608, p = 0.436$) and the Crosswind Component ($F(1, 1707) = 0.957, p = 0.328$) were not significant for the total NASA TLX scores and it was assessed at 4.07kt to calculate estimated marginal means for abnormal operations.

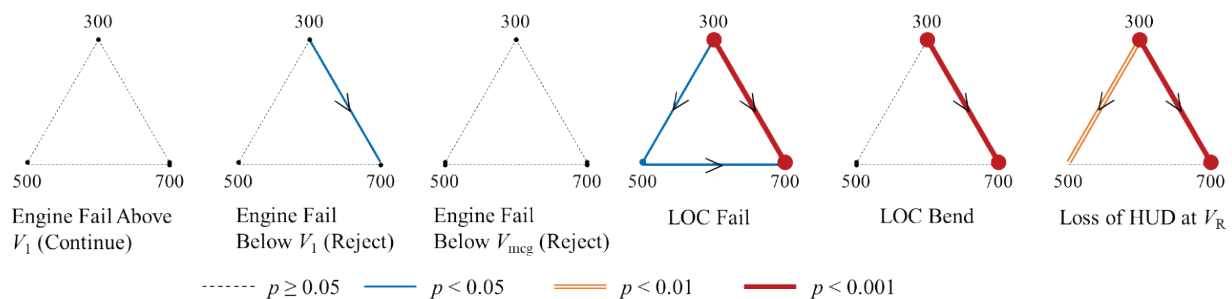


Figure 13. Nested effects of RVR within Failure Conditions.

The overall RVR trends showed identical inverse relationship between RVR and total NASA TLX scores (Figure 14).

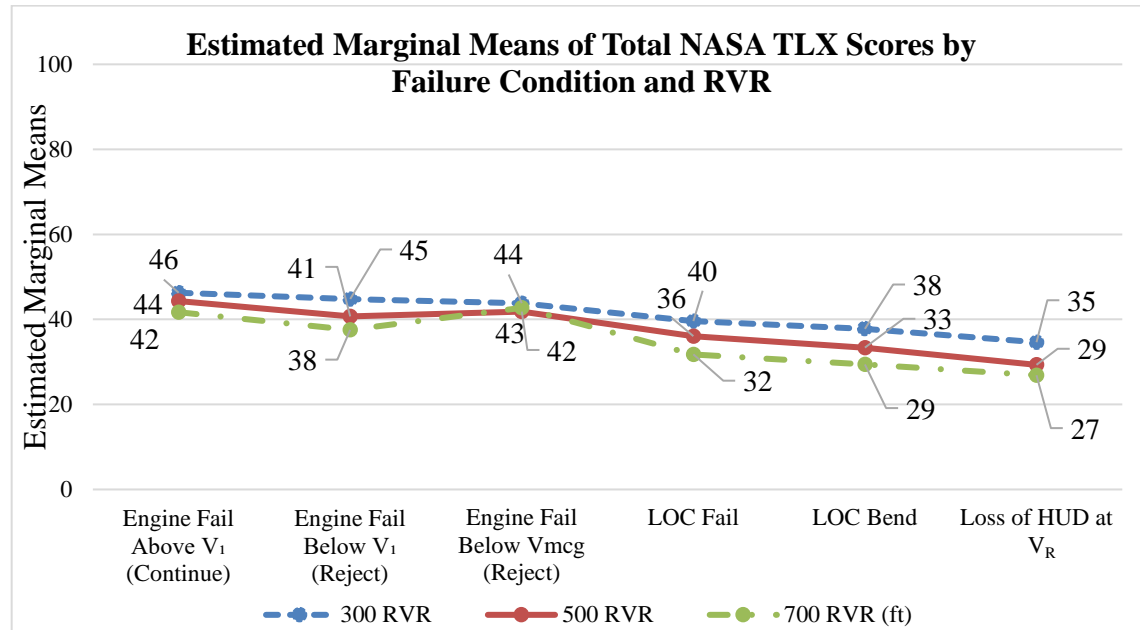


Figure 14. Estimated marginal means of total NASA TLX scores by Failure Condition and RVR.

Comparable trends to those seen in the FTE results emerged in the NASA TLX results as well. That is, during engine malfunctions, the crews experienced higher workload than during the localizer and HUD failures (Figure 15). As expected, there was an inverse relationship between RVR and crew workload present as well (Figure 16).

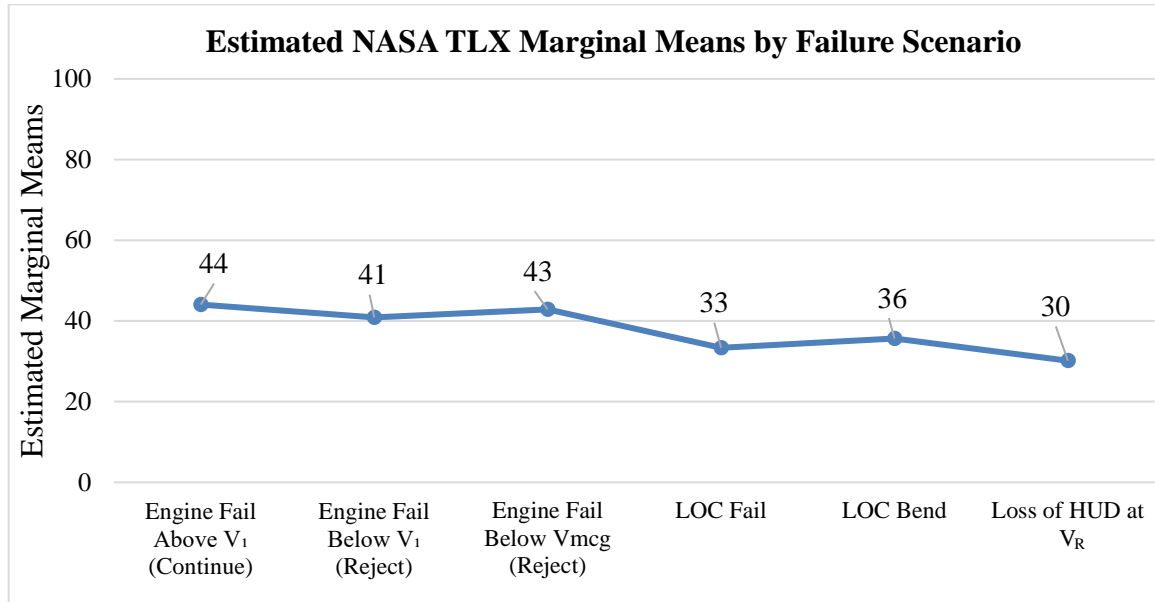


Figure 15. Estimated marginal means of total NASA TLX scores (abnormal operations).

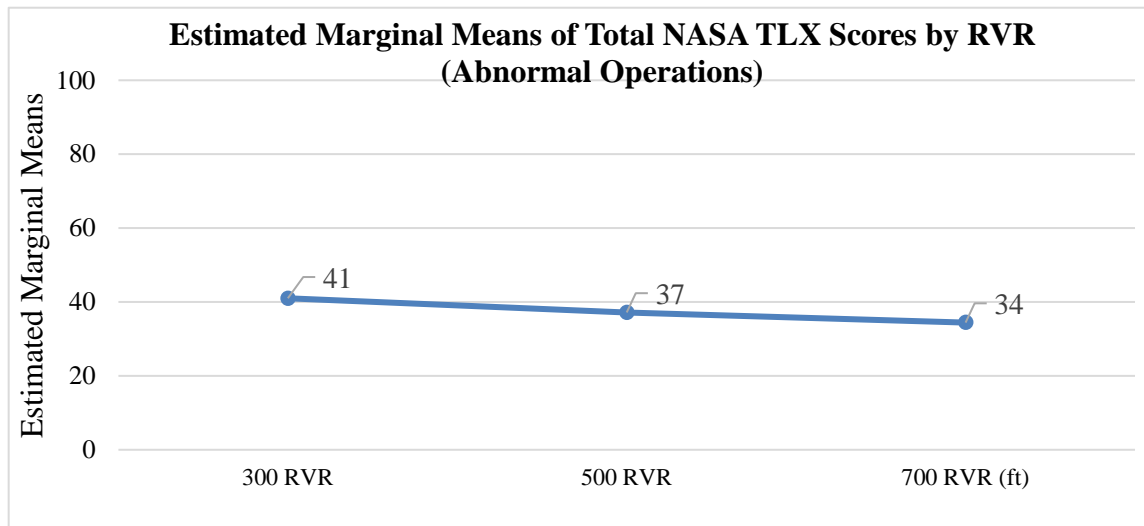


Figure 16. Estimated marginal means of total NASA TLX scores by RVR (abnormal operations).

Discussion

The purpose of this research was to examine the human factors considerations for using HUD with localizer guidance symbology in lieu of currently required infrastructure (i.e., CLL) for lower-than-standard takeoff minima. The results were interpreted within the larger conceptual framework

of attention switching established by previous research (Foyle et al., 1993; Haines et al., 1980; Weintraub et al., 1985, 1984). Specifically, we looked at the interplay between flight deck (near domain) and runway environment (far domain) factors, and their role in supporting pilots' ability to monitor primary flight information efficiently and effectively while maintaining visual contact with the outside environment. To identify the differential contributions of these factors, three baseline conditions without HUD localizer guidance symbology and two conditions with HUD localizer takeoff guidance symbology were used (Table 1). The baseline conditions included currently existing runway infrastructure and markings (e.g., CLL, RCLM) where Baseline 1 and 2 were flown with the HUD and Baseline 3 without the HUD. The two experimental conditions included HUD localizer guidance symbology. Specifically, Condition 1 included RCLM and excluded CLL. Condition 2 excluded both CLL and RCLM.

For this research, RVR conditions within the continuum between 300ft and 1000ft were selected. The research team selected this continuum based on FAA Policy Order 8400.13F (Federal Aviation Administration, 2019) and OpSpec C078/079 regarding takeoffs as follows:

- a) HUD with an approved takeoff guidance system, CLL, HIRL, and front course guidance from a localizer that provides CAT III rollout guidance as indicated by a III/E/4 facility classification (with a published landing minima of 300ft RVR) at or above 300ft RVR but less than 500ft RVR;
- b) HIRL and CLL at or above 500ft RVR but less than 1000ft RVR; and
- c) RCLM and HIRL, or CLL at or above 1000ft RVR but less than 1200ft RVR.

Currently, only about 30% of the CAT I runways in the NAS are equipped with CLL. Therefore, the human factors considerations in using HUD localizer guidance in lieu of CLL, within the RVR continuum defined above were paramount for this research. The experimental conditions of principal interest in that regard were Baseline 2, Baseline 3, and Condition 1 (Table 1).

Inspired by the many years' worth of previous HUD research into the way pilots monitor primary flight information while maintaining visual contact with the outside environment, we sought to obtain more insight into it within the specific context of this study. To accomplish that, we encouraged pilots to share feedback about the techniques they used in this regard for each of the experimental conditions. Pilots reported that they dynamically switch attention between the HUD

symbology (near domain) and the outside environment (far domain). The amount of time they reported looking at the symbology versus looking outside changed with the amount of available outside visual cues. When their attention was focused on the HUD symbology, they reported remaining aware of the outside visual cues and used them as a secondary means for confirming they were continuing on centerline. In conditions with enough external cues visible, the pilots reported using the task-relevant HUD symbology (e.g., Ground Roll Reference Symbol and HGS Guidance Cue) in combination and aligning it with the available outside visual cues (e.g., centerline markings, lighting structures) also while remaining aware of the optical flow in their peripheral vision. This feedback was consistent with findings from previous research (Foyle et al., 1993; Haines et al., 1980; Weintraub et al., 1985, 1984). More specifically, the pilots' verbal feedback supported the notion that due to its volitional nature, visual scanning in a known pattern may mitigate attentional tunneling. On multiple occasions, pilots did report "tunneling" on the Ground Roll Reference Symbol and the HGS Guidance Cue. However, they reiterated the volitional nature of the "tunneling" as a conscious attempt to exclusively focus on that part of the HUD symbology while remaining fully aware of the rest of inside and outside visual cues. The pilot feedback collected during the study was remarkably consistent in emphasizing attention-switching behaviors while monitoring primary flight information and maintaining visual contact with the outside environment. Therefore, we discuss the results from the FTE and crew workload analyses conducted on the data within the attention-switching paradigm.

Flight Technical Error

Normal Operations

Baseline 1 and Baseline 2 conditions contained identical far domain visual cues except for the lack of CLL in Baseline 1 (Table 1). No localizer guidance was provided in either condition. As a result, in Baseline 1, pilots frequently described the Ground Roll Reference Symbol (near domain) as the primary reference in all three RVR conditions. In the background, the RCLM tracking down the middle of the Ground Roll Reference Symbol was used to maintain centerline during the takeoff roll. As expected, at 500ft and 700ft RVR, without the need for switching between the near and far domains, the FTE scores were closely grouped. The increase in FTE at 300ft RVR was due to the reduced amount of relevant information available in the far domain to maintain centerline tracking.

For Baseline 2 and Baseline 3 conditions, the far domain visual cues remained the same (Table 1). In Baseline 2, many pilots reported that at 700ft RVR they primarily followed the CLL (far domain), while simply verifying that the Ground Roll Reference Symbol (near domain) was tracking over the top of the CLL. Nevertheless, at 300ft RVR, pilots often described a different behavior; namely, a concentration on the Ground Roll Reference Symbol (near domain) with a verification that the CLL were tracking through the middle of it. This subtle change in self-described behavior appears to provide further support for the notion of dynamic attention switching between the HUD symbology (near domain) and the outside environment (far domain) depending upon the amount and saliency of visual information available.

Notably, in Baseline 2, the FTE scores at 500ft RVR were higher than at either 300ft RVR or 700ft RVR. One plausible explanation for this phenomenon is that 500ft RVR represents a boundary visibility condition where the CLL (far domain) and the HUD symbology (near domain) were both equally compelling for the pilots, causing them to switch their attention back and forth between the two data sources more frequently, and therefore resulting in higher FTE scores.

In Baseline 3, there was a fundamentally different near domain compared to Baseline 2. That is, no HUD was used and the only information available to the pilot in a head-up position was contained exclusively in the outside environmental cues (e.g., RCLM, CLL, and HIRL). Even though all three visual references were available in this condition, pilots routinely described their actions relative to the CLL. For example, pilots shared personal techniques including: “I lined the centerline lights with my right knee” or “I lined the CLL up with the bolt” (referring to the windshield wiper bolt), indicating that the CLL were more compelling source of information than either the RCLM or HIRL.

Furthermore, Condition 1 built upon Baseline 1 through the addition of the HGS Guidance Cue on the HUD in Condition 1, without any modification to the outside runway visual cues. Similar to Baseline 1, in Condition 1, pilots reported keeping the HGS Guidance Cue centered inside the Ground Roll Reference Symbol (near domain), with RCLM (far domain) tracking down the middle of the Ground Roll Reference Symbol at all RVRs as a confirmation that they were on centerline. There was little need for switching between the near and far domains, resulting in FTE scores similar to those in Baseline 1. While almost identical to Baselines 1, 2 and 3, the FTE scores in Condition 1

were tightly grouped together indicating performance that was more predictable across the different RVR conditions. We attribute these results to the added benefit of the HGS Guidance Cue as part of the HUD symbology provided in this condition.

In Condition 2, all runway visual cues including the RCLM were removed. While Condition 2 would be well outside any approved operational concept, the condition examined the unique utility of the HUD localizer symbology. In the near domain, the pilots were presented with the full set of HUD symbology and only the HIRLs were available as a reference in the far domain. Nevertheless, even in this most challenging experimental condition, where the FTE scores were the highest, those scores were still well below the pass/fail criteria established by FAA AC 120-28D (Figure 1); thus highlighting the robust utility of the HUD localizer guidance symbology. In Condition 2, the inverse relationship between FTE and RVR was most evident when compared to all other guidance types (Figure 4).

Theoretically, the edge lights (HIRLs) could provide an indirect reference to the runway centerline. However, here again, the pilots were remarkably consistent in their feedback with respect to the role HIRLs played in Condition 2. Specifically, the majority of pilots reported that they “did not see them at all” and the few that did, noted that they “knew the lights were there” but did not use them in any capacity as a reference to stay on centerline. This could be attributed to the generally low RVRs used for this research where the number of HIRLs visible in all RVR conditions was very limited before and during the takeoff rollout.

Abnormal Operations

This research also included conditions involving specific aircraft failures occurring within Condition 1 (Table 1). Consequently, all of the preceding discussion regarding pilot use of near domain and far domain visual cues in Condition 1 remains applicable. Nevertheless, during the failure conditions, pilots faced a complex challenge of balancing the two competing tasks of resolving the failures versus maintaining runway centerline. Thus, the FTE scores under those conditions tended to be higher than in the normal operations scenarios. Malfunctions included engine failures at three fixed airspeeds: just below Minimum Control Ground Speed (V_{mcg}) resulting in a rejected takeoff (reject), just below Decision Speed (V_1) resulting in a high-speed rejected takeoff (reject), and just above V_1 resulting in a continued takeoff. When an engine failure occurred, the aircraft experienced an immediate yawing motion around the vertical axis. This yawing motion

produced an immediate steering command via the HGS Guidance Cue to provide the pilot with steering guidance to return to the center of the localizer signal. However, the HGS Guidance Cue was capable of moving rather abruptly depending upon the yaw rate, occasionally causing the pilots to fuse this guidance cue information with information from the external runway visual cues to moderate the magnitude of the steering correction and avoid over-controlling the rudder input while attempting to correct back to centerline. While the HUD also contained a Ground Localizer Line and Scale showing the current aircraft position relative to the center of the localizer signal, pilots often reported minimal use of that information in the near domain, instead focusing primarily on the HGS Guidance Cue and the outside visual cues. Additionally, during the engine failure scenarios that progressed to takeoff rotation, pilots occasionally chose to accept some off-centerline position that was otherwise stable so they could focus on getting the aircraft safely airborne in a controlled manner rather than force a return to centerline that could potentially destabilize the aircraft at takeoff rotation. The V_{mcg} scenario data showed nearly identical FTE at all RVR values (Figure 8). This indicates that initial runway centerline tracking error was mostly a function of the delay from initial pilot reaction to the engine failure prior to making rudder inputs. At the low velocity in this scenario resulting in reduced rudder authority, and with the rapid deceleration via auto-braking, there was little opportunity to measure any meaningful difference in pilot performance at the different RVR values. However, the V_1 reject scenarios produced a larger FTE spread, with the lowest FTE at 700ft RVR and the largest FTE at 300ft RVR. These results are consistent with the previously described fusion of the near and far domains visual cues. The best performance occurred in conditions with higher RVRs, and worsening performance at lower RVRs. With an engine failure at a higher speed for the V_1 reject scenarios compared to the V_{mcg} reject scenarios, the V_1 reject scenarios provided greater rudder authority and a longer rollout distance for steering corrections to take effect and be measured. This resulted in larger differences in performance across different RVRs in the V_1 scenarios. The engine failure just after V_1 resulting in a continued takeoff scenario produced the largest spread of FTE at the different RVR values, with the lowest FTE scores at 700ft RVR and the highest FTE scores at 300ft RVR. These results are consistent with the pilot-described behavior of fusing the near domain steering guidance with the far domain environmental cues, and with performance worsening as far domain visual cues diminished. Particularly in the 300ft RVR scenarios, with the associated difficulty of maintaining sight of the RCLM as the aircraft started to deviate from the runway

centerline, some pilots were content to prioritize the task of making coordinated rudder inputs to counteract the yawing moment and simply getting safely airborne over the task of maintaining exact runway centerline.

Abnormal operations scenarios also included failures of the localizer ground transmitter and an electronic bending of the localizer signal. The localizer transmitter failure was initiated at 70 knots, producing an annunciation on the HUD, and the removal of both the HGS Guidance Cue and the Ground Localizer Line and Scale symbology from the HUD. Absent the HUD localizer guidance symbology after the failure, the pilots were required to switch solely to the RCLM for runway alignment, producing an FTE spread as expected with worsening performance with limited external runway visual cues. For scenarios where the electronic localizer signal was “bent” at 50 knots, pilots were forced to resolve the erroneous indications of the HGS Guidance Cue and Ground Localizer Line and Scale in the near domain that were increasingly divergent from the RCLM in the far domain. This scenario particularly reinforces the concept of volitional attention switching previously described. As the HUD localizer guidance cue diverged from the runway centerline, at both 700ft RVR and 500ft RVR pilots were able quickly to detect that the RCLM were no longer tracking down the center of the flight path vector. They were quick to abandon the erroneous near domain localizer guidance and switch to the far domain RCLM environmental cues. However, at 300ft RVR, the reduced visibility of RCLM made the HUD symbology more compelling to the pilots, thus delaying their switch from the near domain HUD symbology to the far domain RCLM visual cues, resulting in a significantly larger FTE at 300ft RVR. The initiation speed and rate of localizer divergence was identical at all three RVR values. Therefore, the amount of FTE lateral deviation from runway centerline could be directly attributed to the different points where the pilots switched away from the erroneous cues in the near domain to the accurate cues in the far domain. This scenario supports the notion that pilots can switch between near domain and far domain cues volitionally, and do so at a rate dependent upon the amount of outside visual cues available. The last failure condition was a loss of HUD just prior to rotation. Because the pilots were near the takeoff rotation point, and all aircraft control inputs were already stable, this scenario was of little consequence to the pilots; they simply switched to their traditional head-down displays and rotated normally. There was little ability to measure any meaningful difference in pilot performance at the different RVR values, and many pilots called it a “nonevent.”

Crew Workload

No significant differences were found in the total NASA TLX scores when comparing Baseline 2 and Baseline 3, Baseline 3 and Condition 1; and Baseline 2 and Condition 1. We attribute these results to the availability of salient runway centerline visual cues in the near and/or far domains across all conditions. Namely, HUD localizer guidance symbology (near domain) and RCLM (far domain) in Condition 1, and CLL (far domain) in Baseline 2 and Baseline 3. In particular, all three conditions included one of these combinations providing the necessary information about the runway centerline position in reduced visibility conditions. While not significantly higher than in Baseline 2 or Condition 1, pilots reported slightly elevated workload levels in the Baseline 3 condition. This finding could be attributed to the general discontent pilots expressed when asked to fly scenarios without the HUD.

In contrast, the total NASA TLX scores in Baseline 1 and Condition 2 were significantly higher than the scores in all of the other conditions. One plausible explanation of these results could be the lack of salient visual cues either in the near or far domains. For example, in Baseline 1 condition, the only direct runway centerline visual cue was the RCLM. In the RVR conditions used for this research and without HUD localizer guidance, the RCLM were not salient enough to support crew workload levels comparable to those conditions where more salient (stand alone or in combination) visual cues were available. In Condition 2, the only direct runway centerline visual cue was the HUD localizer guidance (near domain) and no CLL or RCLM were available. In this condition, crews reported significantly higher workload levels than in any other condition included in this research. However, even in this most challenging experimental condition, the average reported crew workload levels stayed below the mid-point on the NASA TLX scale.

In summary, the overall crew workload levels reported by the crews across the baseline and experimental conditions were low to moderate in the normal operations scenarios included in this research. The total workload scores during abnormal operations scenarios did not exceed moderate levels on the NASA TLX scale. In line with the results from the statistical analyses and considering the remarkably consistent feedback from the pilot crews in favor of the use of HUD localizer guidance for takeoff, we attribute these results also to the high levels of information redundancy and number of safeguards typical for modern flight decks and multi-crew operations.

Pilot Type

The pilots who participated in this study represented two populations: airline crews and business jet crews. The analyses of total NASA TLX scores across all experimental conditions showed that the business jet pilots experienced slightly elevated levels of workload compared to the airline pilots. For example, the mean differences of total NASA TLX scores between airline and business jet pilots were 6.5 (for PF) and 7.5 (for PM) during the Engine Fail above V_1 (Continue) condition across RVR levels. These results are not surprising due to the natural comfort levels associated with different levels of experience flying a particular platform. The FTE scores of the business jet crews were comparable to the airline pilots where the mean differences of FTE scores were only 0.95 between the different pilot types.

Equivalent Level of Safety

During the debriefing sessions, and in addition to the empirical data collected during the experimental trials, we requested that pilot participants give their personal recommendation for the lowest RVR they considered as equally safe for using HUD localizer guidance symbology in lieu of CLL. Notably, some pilots expressed difficulty to directly compare the HUD localizer guidance symbology and centerline lighting (e.g., “While I don’t believe HUD localizer guided to symbology is ‘equivalently safe’ to operations with centerline lighting, I believe it is safe within certain RVR limitations [I estimate this to be 500 RVR].”).

Conclusion

The results of this research have the potential to inform operational credit changes that would allow more reduced visibility takeoffs and increase the number of viable airports (with reduced infrastructure characteristics) in the NAS that are available for takeoff under low visibility conditions. From a human factors perspective, the analyses indicated that FTE scores and crew workload levels (for both the pilot flying and pilot monitoring) using HUD localizer takeoff guidance without CLL were equal to or less than their corresponding scores with CLL infrastructure at the RVRs used for this research. Additionally, 40.7 percent of study participants indicated that using HUD with localizer guidance was equivalently safe to using CLL with RVR values down to 300ft. Another 31.5 percent stated equivalent levels of safety down to 500ft RVR, meaning that a combined total of 72.2 percent of study participants subjectively assessed that HUD localizer takeoff guidance without CLL was equivalently safe to CLL infrastructure at or above 500ft RVR.

Additional research is needed to investigate the relationship between HUD training, level of proficiency, performance, and workload in both normal and abnormal operations.

Acknowledgements

This project was funded by the FAA NextGen Human Factors Division (ANG-C1) in support of the FAA Office of Aviation Safety, Low Visibility Operations Unit, Flight Technologies & Procedures Division, Flight Operations Group (AFS-410).

References

- Barnette, J. F. (1976). *Role of head-up display in instrument flight* (IFC-LR-76-2). Air Force Instrument Flight Center.
- Boston, B. N., & Braun, C. C. (1996). Clutter and display conformality: Changes in cognitive capture. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 40(2), 57–61.
- Boucek, G. P., Pfaff, T. A., & Smith, W. D. (1983). *The use of holographic head-up display of flight path symbology in varying weather conditions* (SAE Technical Paper Series 831445). Warrendale, PA: Society of Automotive Engineers. Inc.
- Bray, R. S. (1980). *A head-up display format for application to transport aircraft approach and landing* (NASA-TM-81199). NASA.
<https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19800020794.pdf>
- Crawford, J., & Neal, A. (2006). A review of the perceptual and cognitive issues associated with the use of head-up displays in commercial aviation. *The International Journal of Aviation Psychology*, 16(1), 1–19.
- Dunn, P. K., & Smyth, G. K. (2005). Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, 15(4), 267–280. <https://doi.org/10.1007/s11222-005-4070-y>
- Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *American Psychologist*, 63(7), 591.
- Federal Aviation Administration. (1999). *120-28D Criteria for Approval of Category III Weather Minima for Takeoff, Landing and Rollout* (AC 120-28D). U.S. Department of Transportation.
- Federal Aviation Administration. (2019). *Order 8400.13F*. U.S. Department of Transportation.
- Fischer, E., Haines, R. F., & Price, T. A. (1980). *Cognitive issues in head-up displays* (No. 1711). NASA Scientific and Technical Information Branch.
<https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19810005125.pdf>

- Foyle, D. C., McCann, R. S., Sanford, B. D., & Schwirzke, M. F. J. (1993). Attentional effects with superimposed symbology: Implications for head-up displays (HUD). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37(19), 1340–1344.
- Foyle, D. C., Sanford, B. D., & McCann, R. S. (1991). Attentional issues in superimposed flight symbology. *6th International Symposium on Aviation Psychology*, 577–582.
- Haines, R. F., Fischer, E., & Price, T. A. (1980). *Head-up transition behavior of pilots with and without head-up display in simulated low-visibility approaches* (No. 1720). NASA Scientific and Technical Information Branch.
- Hart, S. G. (1986). *NASA Task load Index (TLX). Volume 1.0; Paper and pencil package*. NASA Ames Research Center.
- Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one-and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics*, 17(4), 315–339.
- Hennessy, R. T. (1975). Instrument myopia. *JOSA*, 65(10), 1114–1120.
- Hull, J. C., Gill, R. T., & Roscoe, S. N. (1982). Locus of the stimulus to visual accommodation: Where in the world, or where in the eye? *Human Factors*, 24(3), 311–319.
- Iavecchia, J. H., Iavecchia, H. P., & Roscoe Illiana, S. N. (1988). Eye accommodation to head-up virtual images. *Human Factors*, 30(6), 689–702.
- Iyit, O. O. N. (2018). Modelling the US Diabetes Mortality Rates via Generalized Linear Model with the Tweedie Distribution. *International Journal of Science and Research (IJSR)*, 7(2), 1326–1334. <https://doi.org/10.21275/ART2018368>
- Jarvi, D. W. (1981). *Investigation of spatial disorientation of F-15 Eagle pilots* (ASD-TR-81- 5016). Aeronautical Systems Division.
- Judd, C. M., McClelland, G. H., & Culhane, S. E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology*, 46(1), 433–465.
- Kaber, D. B., Alexander, A. L., Stelzer, E. M., Kim, S.-H., Kaufmann, K., & Hsiang, S. (2008).

- Perceived clutter in advanced cockpit displays: measurement and modeling with experienced pilots. *Aviation, Space, and Environmental Medicine*, 79(11), 1007–1018.
- Kaber, D. B., Kaufmann, K., Alexander, A. L., Kim, S.-H., Naylor, J. T., Prinzel III, L. J., Pankok Jr, C., & Gil, G.-H. (2013). Testing and Validation of a Psychophysically Defined Metric of Display Clutter. *Journal of Aerospace Information Systems*, 10(8), 359–368.
- Kirk, R. E. (1995). Experimental design: procedures for the behavioral sciences (Brooks/Cole, Pacific Grove, CA). *Experimental Design: Procedures for the Behavioral Sciences*. 3rd Ed. Brooks/Cole, Pacific Grove, CA.
- Lauber, J. K., Bray, R. S., Harrison, R. L., Hemingway, J. C., & Scott, B. C. (1982). *An Operational Evaluation of Head-up Displays for Civil Transport Operations-NASA/FAA Phase III Final Report* (No. 1815). NASA Scientific and Technical Information Branch. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19820025505.pdf>
- Levy, J. L., Foyle, D. C., & McCann, R. S. (1998). Performance benefits with scene-linked HUD symbology: an attentional phenomenon? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42(1), 11–15.
- Littman, D., & Becklen, R. (1976). Selective looking with minimal eye movements. *Perception & Psychophysics*, 20(1), 77–79.
- Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research*, 66(4), 579–619.
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6, 1171.
- McCann, R. S., Foyle, D. C., & Johnston, J. C. (1993). Attentional limitations with head-up displays. *Proceedings of the 7th International Symposium on Aviation Psychology*, 70–75.
- McCann, R. S., Lynch, J., Foyle, D. C., & Johnston, J. C. (1993). Modelling attentional effects with head-up displays. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37(19), 1345–1349.
- Mena, B., José, M., Alarcón, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, 2017, Vol. 29, Num. 4, p. 552-557.

- Morey, J. C., & Simon, R. (1993). *Attention Factors Associated With Head-Up Display and Helmet-Mounted Display Systems* (E-19298U). Dynamics Research Corp.
- Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, 7(4), 480–494.
- Newman, R. L. (1980). *Operational Problems Associated with Head-Up Displays during Instrument Flight*. (AFAMRL-TR-80-116). Crew Systems Consultants.
- Newman, R. L. (1987). *Improvement of Head-Up Display Standards. Volume 1. Head-Up Display Design Guide. Appendix* (AFWAL TR-87-3055). Air Force Aeromedical Research Laboratory.
- Newman, R. L. (1995). *Helmet-mounted display symbology and stabilization concepts* (A-950073). Crew Systems Consultants.
- Norman, J., & Ehrlich, S. (1986). Visual accommodation and virtual image displays: Target detection and recognition. *Human Factors*, 28(2), 135–151.
- O'Brien, R. G. (1981). A simple test for variance effects in experimental designs. *Psychological Bulletin*, 89(3), 570.
- Pearson, E. S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika*, 114–133.
- Prinzel III, L. J., & Risser, M. (2004). *Head-up displays and attention capture* (NASA/TM-2004-213000). NASA Langley Research Center.
- Roscoe, S. N. (1985). Bigness is in the eye of the beholder. *Human Factors*, 27(6), 615–636.
- Shelden, S. G., Foyle, D. C., & McCann, R. S. (1997). Effects of scene-linked symbology on flight performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 41(1), 294–298.
- Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition*, 4(3), 215–230.
- Temple, S. D. (2018). *The Tweedie Index Parameter and Its Estimator An Introduction with Applications to Actuarial Ratemaking* (Issue June). University of Oregon.
- Ververs, P. M., & Wickens, C. D. (1998). Head-up displays: effect of clutter, display intensity, and display location on pilot performance. *The International Journal of Aviation Psychology*, 8(4), 377–403.

Weintraub, D. J., & Ensing, M. (1992). *Human Factors Issues in Head-Up Display Design: The Book of HUD (CSERIAC state of art report)*. Crew System Ergonomics Information Analysis Center.

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

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

Wickens, C. D., & Long, J. (1994). Conformal symbology, attention shifts, and the head-up display. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 38(1), 6–10.