Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P5800-cd

Impact of communication delays on pilot workload and performance in RPAS landings: A study of HUD interfaces and workload metrics

Ivan de Souza Rehder

Aeronautics Institute of Technology, São José dos Campos, Brazil. E-mail: ivan.rehder@ga.ita.br

Marina Ronconi de Oliveira

Aeronautics Institute of Technology, São José dos Campos, Brazil. E-mail: marina.oliveira.101443@ga.ita.br

Andrew Gomes Perera Sarmento

Aeronautics Institute of Technology, São José dos Campos, Brazil. E-mail: andrew.g.p.sarmento@gmail.com

Moacyr Machado Cardoso Junior

Aeronautics Institute of Technology, São José dos Campos, Brazil. E-mail: moacyr.cardoso@gp.ita.br

Emilia Villani

Aeronautics Institute of Technology, São José dos Campos, Brazil. E-mail: evillani@ita.br

Studies on managing critical failures during the flight of Remotely Piloted Aircraft Systems (RPAS) must be conducted, especially as long-distance operations between different cities become increasingly feasible. This study analyzes a simulated scenario in which, during a journey between cities, a critical failure occurs that does not affect flight dynamics but requires landing at an unmapped airport, without the ability to remotely adjust the landing settings. In this context, a pilot takes control of the aircraft and performs the landing at an airport along the route. Given that communication between the ground control station and the RPAS is conducted via satellite, a delay of approximately two seconds was observed between the pilot's command and the aircraft's execution. The experiment consisted of three flights, each utilizing a different Head-Up Display (HUD). During the flights, the Instantaneous Self Assessment (ISA) evaluation was applied, while the NASA-TLX was used after each flight to measure workload. After the three flights, the pilots completed the SWORD questionnaire to assess cumulative workload. The objective of this paper is to compare the different subjective evaluations (ISA, NASA-TLX, and SWORD) and investigate their correlation with the pilot's actual performance, aiming to understand the validity of these metrics in the context of RPAS operations under delay conditions.

Keywords: Remotely Piloted Aircraft Systems, Pilot workload, Human performance, Instantaneous Self-Assessment, NASA-TLX, SWORD questionnaire, Human-machine interaction, Performance evaluation metrics.

1. Introduction

One of the most significant technological innovations in modern aviation is the introduction of Unmanned Aerial Vehicles (UAVs).These devices play an increasingly important role in various sectors of developed economies while simultaneously creating unprecedented opportunities in emerging markets (HAULA and AGBOZO (2020)). UAVs and Unmanned Aircraft Systems (UAS), also known as Remotely Piloted Aircraft Systems (RPAS), represent technologies that, although autonomous in many aspects, still rely on pilots, sensor operators, and maintainers for effective operation.

The operation of UAVs presents unique challenges, particularly due to the physical separation between the pilot and the aircraft, remote control via radio frequency, and reliance on digital interfaces. Unlike conventional aviation (LANDRY (2018)), where pilots use all their senses to monitor the aircraft, UAV operators primarily rely on vision, which can increase cognitive load (PESTANA (2011)). Transmission delays, especially in satellite links, exacerbate this load and complicate decision-making in critical situations. Human errors, often associated with these delays, have been responsible for accidents involving UAVs. Technologies such as advanced automation and intuitive interfaces like HUDs (Head-Up Displays) and augmented reality have the potential to mitigate these challenges and improve the remote pilot's situational awareness (FRICKE and HOLZAPFEL (2016)).

This study analyzed a simulated scenario in which, during a journey between cities, a critical failure occurs that requires landing at an unmapped airport. The pilot cannot remotely adjust the landing settings and must take control to perform the landing. Satellite communication introduced a two-second delay between the pilot's commands and the aircraft's execution. The experiment included three flights, each with a different HUD, to assess the impact of the interfaces on operations under delay. Workload was measured in real time using the Instantaneous Self Assessment (ISA), followed by a general assessment using the NASA-TLX tool, and, at the end of the flights, the SWORD questionnaire was applied to measure the accumulated workload.

2. Bibliography Review

The study of human factors in aviation is essential to ensure the safety and efficiency of flight operations. Since the advent of powered flight, issues related to human performance have been a constant concern, representing operational risks that can compromise the effectiveness and safety of missions. These issues are defined as any physiological or psychological condition that negatively impacts the operator's performance, directly affecting the quality of operations (ICAO (2012)).

In the field of human factors, subjective workload assessments are widely used, as they provide a measure of the operator's ability to perform tasks, in addition to evaluating the system's suitability in accommodating the operator efficiently with minimal frustration. Among the most commonly used tools is NASA TLX, which has proven to be an effective approach for comparing different aspects of operator interaction with Unmanned Aerial Systems (UAS) and for making comparisons across different operational systems (SWAN et al. (2010)).

As done by O'CONNELL et al. (2007), some comparative analyses between subjective workload assessment methods can be found in the literature. However, these studies do not include performance as a factor in their analysis. Therefore, this study provides a more holistic view by comparing these techniques, investigating their correlations with pilots' objective performance, and evaluating the validity of these metrics in the context of operations with Remotely Piloted Aircraft Systems (RPAS), especially under communication delay conditions.

3. Method

3.1. *The experiment*

This experiment is part of a larger project called the Air Domain Study (ADS), which followed a design-test-analysis protocol. The tests were conducted in a virtual environment known as the ADS Simulator (SARMENTO et al. (2022)), a computational station designed to simulate aircraft behavior in a controlled flight environment. The interfaces developed aim to assist the pilot in operating the UAV.

The experiment involved the participation of 18 pilots from the Brazilian Air Force (FAB), who performed a simulation in which they were required to land a UAV at an airport outside the usual operating zone of the aircraft. The task required the pilot to perform a lateral correction maneuver, both with and without communication delay. The trajectory was marked by green rings, and the pilot had to maneuver the UAV to pass through the center of these rings.

It is noteworthy that the experiment was approved by the ethics committee of the mentioned institution (Ethics Submission Registration Number: CAAE 77429824.7.0000.5503). All pilots participated in the experiment after signing an Informed Consent Term, ensuring that their participation was fully anonymous and the data collected would be used exclusively for academic purposes.

The experiment began with a briefing, during which the pilot was explained the objective of the simulation, the equipment to be used (simulator, sensors, and analysis methods), and the general organization of the experience. Following this, a training session was conducted to familiarize the pilot with the experimental setup, ensuring that they were comfortable maneuvering the UAV. This training included conditions both without delay and with a 1-second communication delay.

The experimental phase itself consisted of three flight simulations, each using a different HUD with a 2-second delay activated. During the flights, the pilots were required to respond to the ISA scale, which ranges from 1 to 5, indicating the perceived momentary workload. At the end of each flight, the pilot completed the NASA-TLX questionnaire, and upon finishing the experiment, they answered the SWORD.

3.1.1. *HUDs*

Standard HUD

The first flight interface is designed to provide crucial information to the pilot, focusing on the representation of the aircraft's trajectory and behavior. In the center of the screen, two concentric circles symbolize the direction of the aircraft's nose, while three lines—arranged to the left, right, and upward—indicate the aircraft's current trajectory.

The lateral lines also serve as an artificial horizon, allowing the pilot to have a clear reference for the aircraft's pitch. The numbers positioned above and below these lines reflect the pitch angle, offering an accurate view of the vertical flight behavior. Additionally, the numbers on the left of the interface display the speed in knots, while the numbers on the right show the altitude in feet, creating a panel that combines performance and control information.

Alternative HUD 1

The second interface introduces new elements focused on the dynamic aspects of flight. On the left, an indicator displays the roll rate, while on the right, the pitch angle is shown, each accompanied by a number corresponding to the respective angle. This configuration provides an instant view of the aircraft's orientation. At the center of the screen, square reticles represent the predicted trajectory of the aircraft, allowing the pilot to visualize not only the current position but also the path the aircraft is expected to follow. This feature facilitates quick and precise decision-making during the flight.

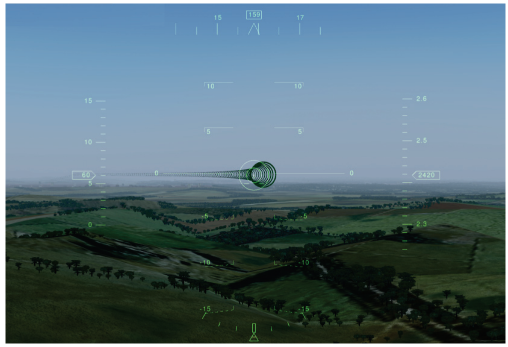

Fig. 1. Standard HUD (Author, 2025).

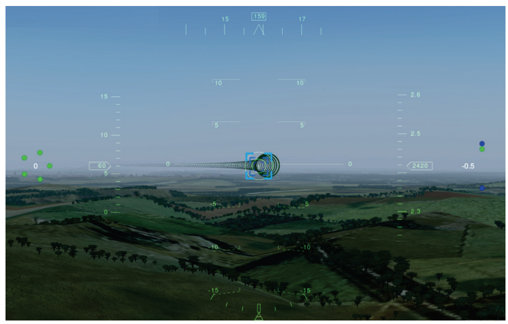

Fig. 2. Alternative HUD 1 (Author, 2025).

Alternative HUD 2

The third interface adopts a more interactive visual approach, featuring indicators on the left and right that display roll and pitch angles, respectively. Each indicator uses a color-coded system to communicate the status of these angles to the pilot: green indicates an ideal condition, yellow signals the need for attention, and red warns of dangerous situations requiring immediate action.

At the center, a fixed reticle and a movable reticle create a dynamic visual cue where the movable reticle, always pink, predicts the aircraft's trajectory, while the fixed reticle changes color to indicate whether the predicted maneuver is safe or not.

In addition, chevrons resembling triangles appear on the right of the reticles. If positioned above the artificial horizon, they indicate that the throttle should be increased; if below, they suggest the throttle should be decreased. This representation also employs color variations to indicate whether the speed is near the ideal range (green), requires attention (yellow), or is considered dangerous (red). Together, these interface elements create an intuitive control environment, aiding the pilot in managing the aircraft during complex flight situations.

Fig. 3. Alternative HUD 2 (Author, 2025).

3.2. *Workload Assessment Methods*

Subjective workload measures are widely used methods in human factors research to understand how operators perceive their workload while performing tasks. These measures focus exclusively on the operator's perception of the level of effort and demand they experience, without directly assessing the task itself or the operator's performance.

According to Casner and Gore (2010), there are two main types of subjective workload measurement techniques:

Subjective Numerical Measurement Techniques

These techniques ask the operator to assign a numerical or ordinal value to the workload they are experiencing in a specific situation. Examples include scales like the NASA-TLX (Task Load Index), where participants evaluate specific dimensions of workload (mental, physical, temporal demand, frustration, effort, and performance) on a numerical scale.

Subjective Comparative Measurement Techniques

In these techniques, the operator compares different task situations and indicates which one involves higher or lower workload. An example is the paired comparison method, where participants analyze two conditions and choose the more demanding one.

3.2.1. *Instantaneous Self-Assessment*

The ISA is a simple technique where operators rate their workload on a scale from 0 to 100 at periodic intervals. This approach helps capture variations in workload over time, providing a dynamic view of the operator's experience. The technique is also easy for the experimenter to apply, as they simply observe the operator and ask the necessary questions, without the need for complex real-time analysis.

3.2.2. *NASA Task Load Index (NASA-TLX)*

The NASA-TLX measurement technique was developed to help mitigate a number of problems arising from differences in how people perceive workload. The NASA-TLX technique is similar to the instantaneous self-assessment technique, as the experimenter must periodically ask the human operator for subjective estimates of their workload. The key difference with the NASA-TLX technique is that, instead of asking participants to subjectively rate their workload on a single scale, they must assess their workload across six different subscales. Each of the six subscales was designed to characterize workload in a distinct way. The six workload subscales are as follows: mental demand (MD), physical demand (PD), temporal demand (TD), performance (PE), frustration (FR) and effort (EF).

3.2.3. *Subjective Workload Dominance (SWORD)*

According to Casner and Gore (2010), the SWORD technique assumes that a more accurate assessment of the workload experienced by human operators in different task contexts can be achieved by comparing task situations with one another rather than evaluating them in isolation.

4. Results

4.1. *Description of the Dataset*

The dataset has the performance and the workload measurements from 18 pilots across three distinct flight conditions: ”Standard HUD (STD HUD)”, ”Alternative HUD 1 (ALT HUD 1)” and ”Alternative HUD 2 (ALT HUD 2)”. The performance variable, Mean Distance, represents the pilot’s average distance from the center of the rings in the flight path. In addition, three different workload assessments were colleted: the Instantaneous Self Assessment (ISA) for real-time workload evaluation, the NASA-TLX for task-level workload ratings, and the SWORD questionnaire for cumulative workload measurement after all flights.

The dataset was analyzed to uncover relationships between performance (Mean Distance) and workload metrics under different flight conditions. Table 1 presents a sample of the dataset with a focus on ISA scores, while Table 2 provides an overview of NASA-TLX and SWORD scores.

Pilot	HUD	Mean ISA Score	Mean Distance	Distance Std
P01	ALT HUD 1	3.65	4.84	2.86
P01	ALT HUD 2	2.57	3.61	1.63
P01	STD HUD	4.14	4.66	3.74
P02	ALT HUD 1	3.07	4.15	2.33
⋮	⋮	⋮	⋮	⋮
P18	STD HUD	1.31	2.88	2.09

4.2. *Data Aggregation, Processing, and Variable Correlation*

The data underwent processing steps and visualization techniques were then employed to deepen understanding of the relationships between workload metrics and performance. Key steps included:

- **Categorization:** The performance metric, Mean Distance, was categorized into three levels—”High,” ”Medium,” and ”Low”—based on distance percentiles. This categorization ensured balanced groups for performance comparisons.
- **Correlation Analysis:** A Pearson correlation analysis was performed to assess the linear relationships between variables. The results were summarized in a heatmap (Figure 4), highlighting associations between performance (Distance) and workload metrics such as ISA, NASA-TLX, and SWORD.
- **Boxplots:** Boxplots were generated to visualize variations in performance across the workload metrics.

These processing and visualization steps revealed meaningful patterns in the data. This comprehensive approach enabled a deeper understanding of the relationships between key variables and performance.

4.3. *Relationship Between Performance and ISA*

The correlation heatmap (Figure 4) indicates a weak positive correlation (0.42) between Mean ISA Scores and Mean Distance.

Figure 5 illustrates the relationship between ISA scores and performance categories. Pilots in high-performance categories (lower Mean Distance) generally reported lower ISA scores. However, a broader range of ISA scores is observed in lower performance categories.

4.4. *Relationship Between Performance and NASA-TLX*

The correlation heatmap (Figure 4) reveals no significant correlation between Mean Distance and Score PE (Perceived Performance, 0.10) or between Mean Distance and the overall NASA-TLX Score (0.26).

In Figure 6, Score PE tends to decrease as performance improves. Similarly, Figure 7 shows that overall NASA-TLX scores decline slightly with better performance categories. Reminder: In

Pilot	**HUD**	**Nota DM**	**Nota DF**	**Nota DT**	**Nota DE**	**Nota ES**	**Nota FR**	**SWORD**
P01	ALT HUD 1	54.00	2.00	12.00	12.00	14.00	10.00	0.25
P01	ALT HUD 2	30.00	4.00	35.00	16.00	7.00	4.00	0.06
P01	STD HUD	60.00	2.00	9.00	12.00	16.00	25.00	0.69
P02	ALT HUD 1	24.00	42.00	6.00	9.00	7.00	10.00	0.72
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
P18	STD HUD	30.00	6.00	2.00	36.00	4.00	5.00	0.71

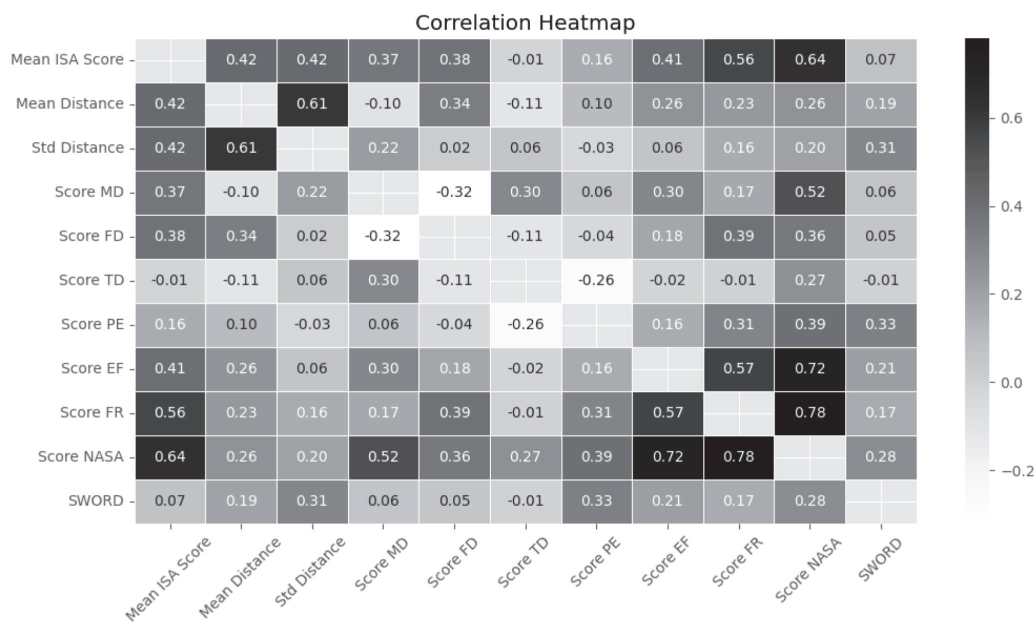

Fig. 4. Correlation heatmap illustrating relationships between variables.

NASA-TLX, the higher the Perfomance Score, the lower is the perceived performance.

4.5. *Relationship Between Performance and SWORD*

The correlation coefficient between SWORD scores and Mean Distance (0.19, Figure 4) indicates a weak positive relationship.

Figure 8 provides additional detail, showing the distribution of SWORD scores across performance categories. Medium performance categories exhibit tighter distributions, while Low and High categories show greater variability.

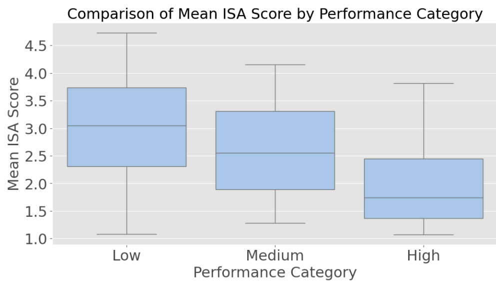

Fig. 5. Comparison of Mean ISA Scores across Performance Categories.

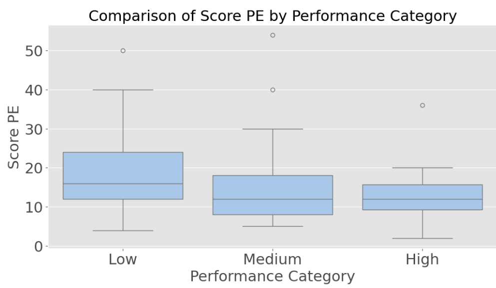

Fig. 6. Comparison of Performance Score (Score PE) across Performance Categories.

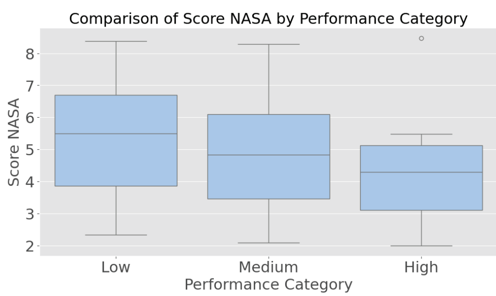

Fig. 7. Comparison of Overall NASA-TLX Scores across Performance Categories.

5. Discussions

The results of this study provide insights into the relationship between subjective workload metrics (ISA, NASA-TLX, and SWORD) and pilot performance during RPAS operations under delay conditions. The simulated scenario with delay demonstrated the cognitive and operational challenges pilots faced, and the results shed light on the effectiveness of these subjective evaluations in this specific context.

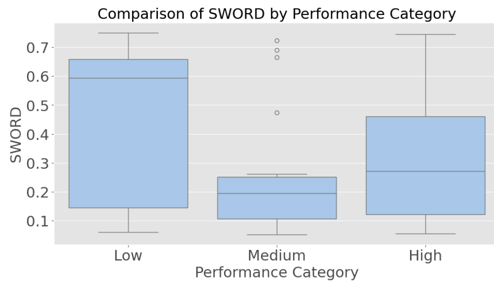

Fig. 8. Distribution of SWORD Scores across Performance Categories.

5.1. *ISA Score*

The weak positive correlation between Mean ISA Scores and Mean Distance (0.42) suggests that higher ISA Scores are loosely associated with reduced performance accuracy. Pilots who performed worse (flew further from the center) reported higher perceived workload during the task. The boxplot (Figure 5) supports this, showing that better performance categories (lower Mean Distance) generally had lower ISA Scores. However, the wide range of ISA Scores within each performance category suggests individual differences in workload perception, possibly influenced by factors such as pilot experience or adaptability to delayed control feedback.

5.2. *NASA Overall and performance*

The weak correlation between NASA-TLX scores and Mean Distance indicates that subjective evaluations of workload and perceived performance (e.g., Score PE) are not strongly tied to objective performance. This suggests that other factors, such as cognitive effort or external pressures, may shape pilots' perceptions more significantly than task accuracy. The boxplots (Figures 6 and 7) reveal subtle trends. For instance, Figure 7 shows that overall NASA-TLX scores slightly decline as performance improves, suggesting that pilots perceived the task as less demanding when they performed well. However, Score PE (Figure 6) lacks a consistent pattern, reflecting variability in how pilots evaluated their own performance.

5.3. *SWORD*

The weak positive correlation between SWORD scores and Mean Distance (0.19) implies that cumulative workload is only marginally influenced by performance accuracy. The boxplot (Figure 8) shows no pattern in SWORD scores across performance categories. Although Medium performance categories exhibit tighter distributions, greater variability in SWORD scores for Low and High performance categories may reflect individual differences in how pilots accumulated and interpreted workload over the three flights, potentially influenced by fatigue or adaptability to the HUD designs.

6. Conclusions

The results demonstrate that subjective workload metrics provide valuable insights into pilots' experiences during RPAS operations but have limited correlation with objective performance measures.

ISA scores, as real time indicators of workload, offer useful insights into pilots' immediate cognitive demands but weakly reflect performance. In contrast, NASA-TLX and SWORD capture retrospective and cumulative workload assessments, but relate very weakly to the performance metrics.

These results suggest that while subjective metrics are important for understanding pilots' workload, they should be complemented with objective performance data to provide a more comprehensive evaluation.

For future work, physiological indicators such as electrocardiogram (ECG), electrodermal activity (EDA), and environmental temperature sensors could be used for a more objective assessment, combined with the subjective measures discussed in the article, to provide a fuller understanding of pilot workload under communication delay conditions.

Additionally, it could be valuable to combine these assessments with factors such as pilot experience, age, and gender to further enrich the research and offer a more nuanced perspective on how these variables may influence workload and performance.

Acknowledgement

Thanks to the Aeronautics Institute of Technology (ITA), in particular, the Competence Center in Manufacturing (CCM) for granting the necessary facilities and structure for the development of the work, to the Funding Authority for Studies and Projects (FINEP) and to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financial support.

References

FRICKE, T. and F. HOLZAPFEL (2016). An approach to flight control with large time delays derived from a pulsive human control strategy. *AIAA*.

HAULA, K. and E. AGBOZO (2020). *Technology in Society 63*, 101357.

ICAO (2012). *Manual of Civil Aviation Medicine*. Montreal, Canada: International Civil Aviation Organization.

LANDRY, S. J. (Ed.) (2018). *Handbook of human factors in air transportation systems*. Human factors and ergonomics. CRC Press, Taylor Francis Group.

O'CONNELL, J. G., S. M. DOHERTY, and I. A. WILSON (2007). Comparison of three subjective workload metrics for a free flight environment. In *International Symposium on Aviation Psychology*, Dayton, Ohio. Wright State University.

PESTANA, M. E. (2011). Flying unmanned aircraft: A pilot's perspective. Technical report, NASA Dryden Flight Research Center, Edwards AFB, California.

SARMENTO, A. G. P., T. R. d. PAULA, A. S. d. OLIVEIRA, E. T. d. SILVA, J. POSSAMAI, H. C. MARQUES, M. M. CARDOSO JUNIOR, and E. VILLANI (2022). A human-machine interface analysis for teleoperation of uav overtime delay. In *ICAS 2022*.

SWAN, B. J., M. ZIARNICK, J. E. MCDONALD, and S. B. HOTTMAN (2010). Fatigue risk management systems (frms): Enhancing aviation safety. *TIB - Leibniz 238*. Print version.