*The Aeronautical Journal* (2025), **129**, pp. 529–558
doi:10.1017/aer.2024.122

ROYAL AERONAUTICAL SOCIETY

RESEARCH ARTICLE

# Eye-tracking analysis to assess the mental load of unmanned aerial system operators: systematic review and future directions

A.C. Russo[1], M.M. Cardoso Junior[2] and E. Villani[2]

[1]Departamento de Engenharia de Minas e de Petróleo da Escola Politécnica da Universidade de Sao Paulo, São Paulo, Brazil
[2]Divisão De Engenharia Mecânica-Aeronáutica, Instituto Tecnológico de Aeronáutica, São José dos Campos, Brazil
**Corresponding author:** A.C. Russo; Email: anacarolinarusso@usp.br

**Received:** 13 March 2024; **Revised:** 3 October 2024; **Accepted:** 9 October 2024

**Keywords:** eye-tracking; mental workload; unmanned aircraft system (UAS); UAV operator

## Abstract
This article presents a systematic review on the use of eye-tracking technology to assess the mental workload of unmanned aircraft system (UAS) operators. With the increasing use of unmanned aircraft in military and civilian operations, understanding the mental workload of these operators has become essential for ensuring mission effectiveness and safety. The review covered 26 studies that explored the application of eye-tracking to capture nuances of visual attention and assess cognitive load in real-time. Traditional methods such as self-assessment questionnaires, although useful, showed limitations in terms of accuracy and objectivity, highlighting the need for advanced approaches like eye-tracking. By analysing gaze patterns in simulated environments that reproduce real challenges, it was possible to identify moments of higher mental workload, areas of concentration and sources of distraction. The review also discussed strategies for managing mental workload, including adaptive design of human-machine interfaces. The analysis of the studies revealed a growing relevance and acceptance of eye-tracking as a diagnostic and analytical tool, offering guidelines for the development of interfaces and training that dynamically respond to the cognitive needs of operators. It was concluded that eye-tracking technology can significantly contribute to the optimisation of UAS operations, enhancing both the safety and efficiency of military and civilian missions.

## Nomenclature

| | |
|---|---|
| UAS | Unmanned Aircraft System |
| UAV | Unmanned Aerial Vehicle |
| NASA-TLX | NASA Task Load Index |
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-analyses |
| ECG | Electrocardiogram |
| EEG | Electroencephalogram |
| fNIRS | Functional Near-Infrared Spectroscopy |
| PERCLOS | Percentage of Eyelid Closure |
| ApEn | Approximate Entropy |
| HRV | Heart Rate Variability |
| SCOUT | Supervisory Control Operations User Testbed |
| CHMI | Cognitive Human-Machine Interfaces |
| CHMI2 | Cognitive Human-Machine Interfaces and Interactions |
| SA&CA | Separation and Collision Avoidance |
| MUM-T | Manned/Unmanned Teaming |
| AAA | Attention Allocation Aid |
| DelCon | Delegation of Control |
| StArt | State of the Art through Systematic Review |

| | |
|---|---|
| PICo | Population, Intervention, Comparison, Outcome |
| AMC | Air Mission Commander |
| GCS | Ground Control Stations |
| SCORCH | Supervisory Control of Remote Crewed and Uncrewed Assets |

## 1.0 Introduction

As in many areas, aviation is progressively adopting unmanned aircraft systems (UAS), as the increasing incorporation of unmanned aircraft into military operations proves to be a game-changer in defense tactics and strategies. These systems can perform long-duration missions in remote and hostile areas, eliminating the need for expensive and bulky life support systems, which results in a higher payload capacity per flight Fricke and Holzapfel [1]. These autonomous systems offer numerous advantages, from advanced tactical reconnaissance to surgical action in high-risk environments. However, the effectiveness of these operations intrinsically depends on the mental load of the operators involved in the cognitive human-machine interfaces.

In military contexts, UAS operators face highly complex and dynamic situations. Target identification, real-time data analysis, crucial decision-making and strategic coordination require a level of attention and mindload management that directly influences mission success. In addition, such operations often occur in harsh environments, where the ability to maintain constant vigilance is of utmost importance for safety and effectiveness.

In this scenario, the assessment of the mental load of UAS operators in cognitive human-machine interfaces emerges as a critical consideration. Mental load, representing the cognitive effort required to perform tasks, plays a vital role in the execution of operations. Maintaining a balanced mental load is essential to allow operators to focus on crucial tasks, ensuring continuous vigilance and accurate decision-making in the face of ever-evolving situations.

Accurate assessment of mental load, however, is a complex challenge, particularly in military settings. Traditional approaches, such as self-assessment questionnaires such as NASA-TLX [2–7], may be limited in terms of accuracy and objectivity. Therefore, it is imperative to employ more advanced methods that enable real-time and continuous understanding of the mental load.

In this article, we will turn to eye tracking analytics, a powerful tool for capturing the nuances of operators' visual attention during UAS operations.

By thoroughly analysing the gaze patterns of operators in simulated environments, which reproduce the real challenges of military operations, it is hoped to identify the moments of greatest mental load, areas of concentration and possible sources of distraction [8]. This in-depth analysis will allow us to understand how the mental load varies throughout operations and how this variation influences critical decision-making.

In addition, it is intended to identify effective strategies to manage the mental load efficiently. This will include the adaptive design of the cognitive human-machine interface, where the distribution of information and alerts can be dynamically adjusted, considering the perceived load of the operators [9]. By optimising data presentation and effective information management, it is hoped to keep operators in a state of mental load suitable for performing tasks, reducing cognitive fatigue and improving performance [10].

Given the complexity and importance of assessing the mental load in UAS operators in cognitive human-machine interfaces, this article employs a systematic review focused on the potential of eye-tracking as a diagnostic and analytical tool. By compiling and analysing relevant studies, we seek not only to understand the effectiveness of this technology in capturing mental load indicators, but also to identify guidelines for the development and improvement of interfaces and training that dynamically respond to the cognitive needs of operators. Such an approach aims to contribute significantly to the optimisation of UAS operations, elevating both the safety and efficiency of military and civilian missions.

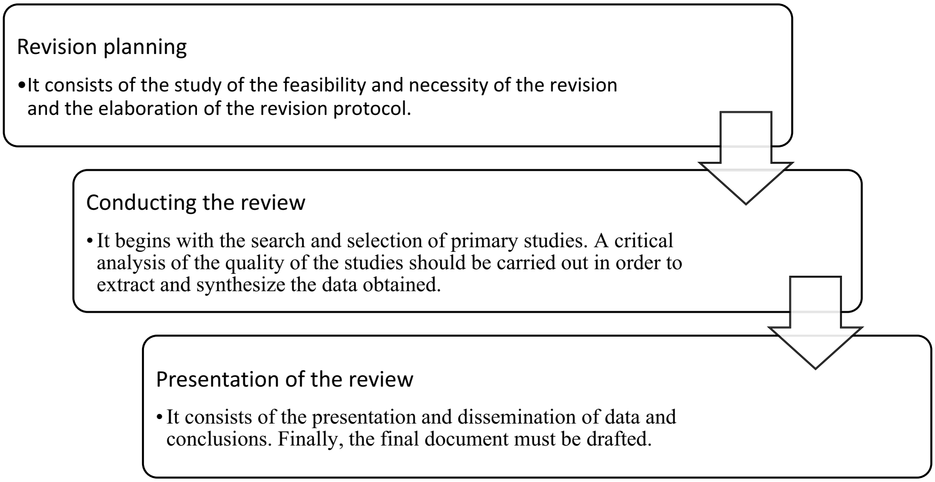


***Figure 1.*** *Stages of the systematic review of the literature. Source: Kitchenham and Charters [12] – modified)*

## 2.0 Materials and methods

This research is a theoretical study through the application of the technical procedure of systematic review of the literature (SRL). This technique was used to identify, evaluate and interpret relevant research on the subject, using a defined methodological sequence that allows the aggregation of knowledge and the construction of knowledge [11, 12].

The design of this SRL was prepared in accordance with the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement [13] and since this was a literature review, it was not necessary to submit it to the Ethics Committee.

The SRL comprises a sequence of three stages: planning, conducting and presenting the review, each with its own actions (Fig. 1).

### *2.1 Search strategy*

The PICo (Population/Problem, Interest and Context) strategy for non-clinical research was used to construct the research question (Table 1), These are: How the use of eye tracking contributes to the assessment of the mental load of operators of UASes and unmanned aerial vehicles (UAVs)?

The choice of a population composed exclusively of UASes and UAVs operators for this systematic review is strategic and justified. UAS and UAV operators play a crucial role, acting as remote pilots who control and monitor aircraft from distant locations, without the physical presence in the cockpit. This role involves managing navigation, making critical decisions in real-time, and analysing complex sensory data, requiring a high cognitive load to maintain safety and operational efficiency.

The complexity and unique cognitive demands faced by these professionals provide a deep spectrum of insights into mental load in cognitively demanding operations, directly relevant to UAS and UAV operation. In addition, the specific literature focused on UAS and UAV operators, especially about mental load assessment via eye tracking, is notoriously sparse. The specific inclusion of these operators allows for a direct understanding of the cognitive demands they face, contributing to filling the gaps identified in existing research and expanding the body of knowledge applicable to the human-machine cognitive interface in UAS contexts.

The searches were carried out in the Web of Science and Scopus databases, chosen for their interdisciplinary nature and for being considered two of the largest reference databases in the world. For this purpose, word combinations were used (Fig. 2).

***Table 1.*** *PICo strategy for the elaboration of the research question*

| Criterion | Definition |
|---|---|
| Population | The target population exclusively includes operators of Unmanned Aircraft Systems (UAS) and Unmanned Aerial Vehicles (UAVs). These professionals are responsible for the remote control and supervision of unmanned aircraft in a variety of contexts, including military operations, surveillance missions, scientific research and commercial applications such as mapping and cargo delivery. The emphasis is on assessing the mental load of these operators, given the complexity and unique cognitive requirements of monitoring and operating UAS or UAV without direct physical interaction with the aircraft. |
| Intervention | The intervention is the use of the eye tracker as a tool to measure and assess mental load. This includes monitoring eye movements, blinking patterns and visual fixation, among other eye parameters such as indicators of mental load. |
| Control | Comparison can be made with other mental load assessment methods used in the aeronautical sector, such as self-reported questionnaires, physiological measures (e.g., heart rate, skin conductance), and performance on specific tasks. |

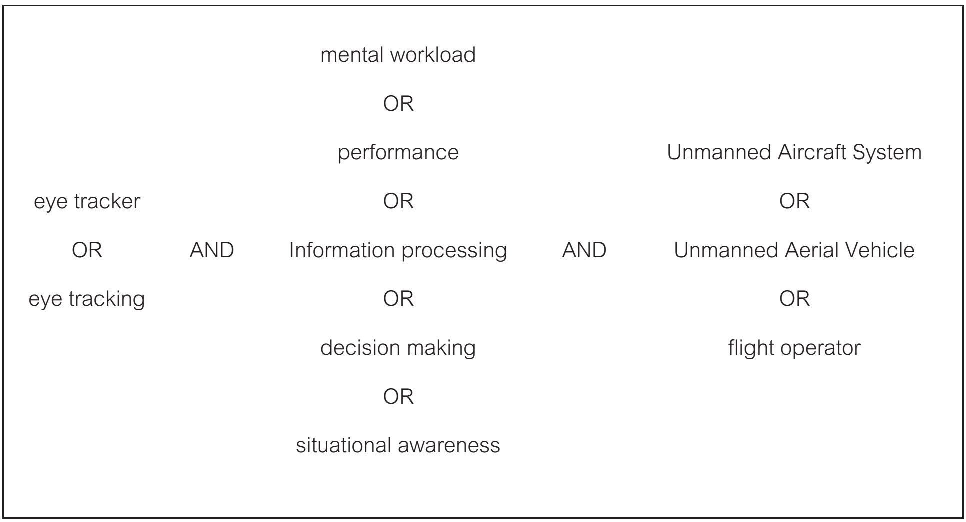


***Figure 2.*** *Combination of keywords used in the search. Source: Authors.*

From the analysis based on the keyword combinations in Fig. 2, it was possible to perform a temporal analysis of the volume of publications and citations related to the use of eye tracking in the assessment of mental load in UAS and UAV operators (Fig. 3).

The Fig. 3 shows a significant increase in the number of published documents over the years, with a particularly sharp rise starting in 2019, reaching a peak in 2024. This trend indicates a growing relevance of the topic, demonstrating increased interest and engagement from the scientific community in recent years.

Figure 4 depicts a pie chart that demonstrates the distribution of scientific publications by areas of knowledge, related to the use of eye tracking to assess mental load. The areas with the highest number of

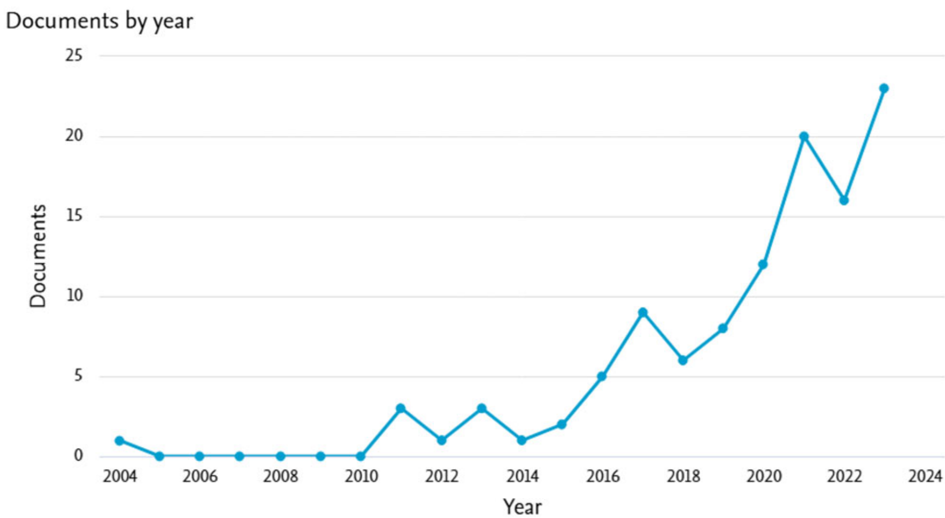


***Figure 3.*** *Years of publication (Scopus). Source: Authors.*

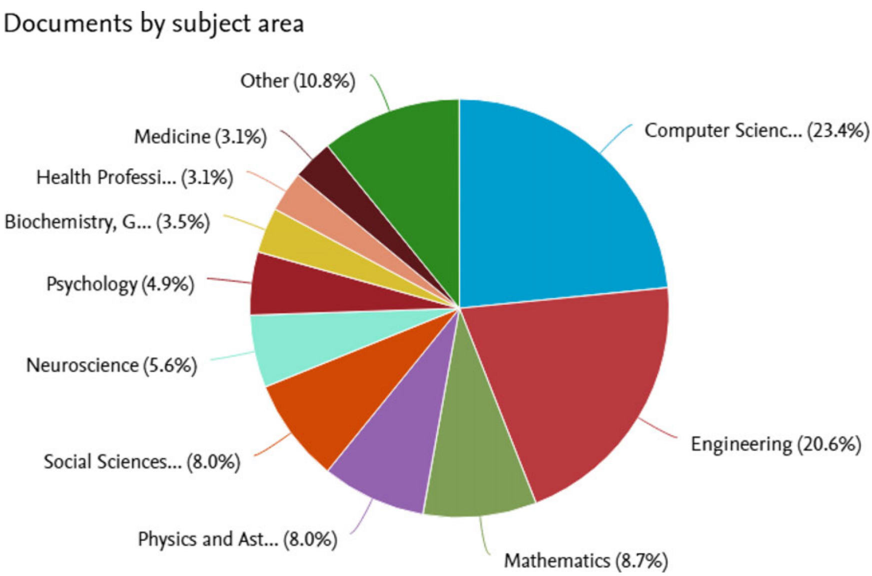


***Figure 4.*** *Main areas of knowledge (Scopus). Source: Authors.*

publications include Computer Science and Engineering (23.4% and 20.6%, respectively). This image was generated by the Scopus platform.

The graph (Fig. 4) demonstrates the interdisciplinarity of the study of mental load and the application of eye tracking in a variety of fields, highlighting the cross-cutting relevance of this technology in understanding complex cognitive phenomena.

### *2.2 Eligibility criteria*

Potentially relevant studies were selected by two independent reviewers according to the following inclusion criteria: only articles in Portuguese and English; full article; review studies, studies with objectives other than the present review were excluded; studies with different audiences; abstracts, technical reports, oral communications, letter to the editor.

The initial selection of the articles occurred independently, through the reading of their titles and abstracts. Subsequently, both reviewers read the full texts of the articles that met the inclusion criteria. Any disagreement about the eligibility of the articles was resolved through consultation with a third researcher. The report of the number of studies included and excluded in the different phases of the systematic review is presented later using the PRISMA flowchart (Chart 3).

### *2.3 Extraction of data from articles*

Data extraction included the following variables: authors, year, objective and main results of the study. The StArt (State of the Art through Systematic Review) software, developed by researchers from the Federal University of São Carlos, was used to manage the selection of articles [14].

### *2.4 Quality assessment*

In the evaluation methodology adopted for the systematic review, two different scales were considered to assess the quality of the selected articles: the scoring scale for the population studied and the scoring scale for the study design.

In the first scale, the articles were evaluated based on the relevance and representativeness of the studied population in relation to the aeronautical sector. The score ranged from 0 for unspecified or irrelevant populations to 5 for those that were exceptionally specified and representative.

In the second scale, the focus was on the methodological rigor of the studies, with the score ranging from 1 to 5, assigned to different study designs, from narrative reviews and expert opinions to experimental studies, valuing studies that allowed a strong causal inference and strict control of variables.

In addition to the two parameters previously mentioned for the evaluation of the selected articles, the software used for data extraction assigned an additional score based on the presence of the keywords defined for this study. Each article could receive up to 5 points if the keywords were present in the title, 3 points if they were in the abstract, and 2 points if they were listed among the keywords. This complementary score served as an indication of the relevance of the article in relation to the focus of the study, allowing for a more refined weighting and a selection of articles highly pertinent to the topic of interest.

This detailed and insightful methodological approach ensured a comprehensive and fair evaluation of the articles, allowing for a reliable qualitative synthesis of existing data on the mental load of unmanned aircraft operators. The scarcity of articles specifically focused on this audience justified the inclusion of professionals from different areas of the aeronautical sector, ensuring a comprehensive view of the application of eye tracking in various cognitive contexts.

## 3.0 Results and discussions

The search strategy identified 137 articles. A total of 16 duplicate articles were eliminated and 64 articles were selected for title and abstract screening, of which 57 were excluded because they did not meet the inclusion criteria.

Of the remaining 64 articles evaluated in full, 38 were excluded because they also did not meet the inclusion criteria. Therefore, 26 articles were included in the present systematic review (Fig. 5).

### *3.1 Analysis of the studies found*

The discussion of the main results found in the studies analysed in this article are highlighted in Chart 1.

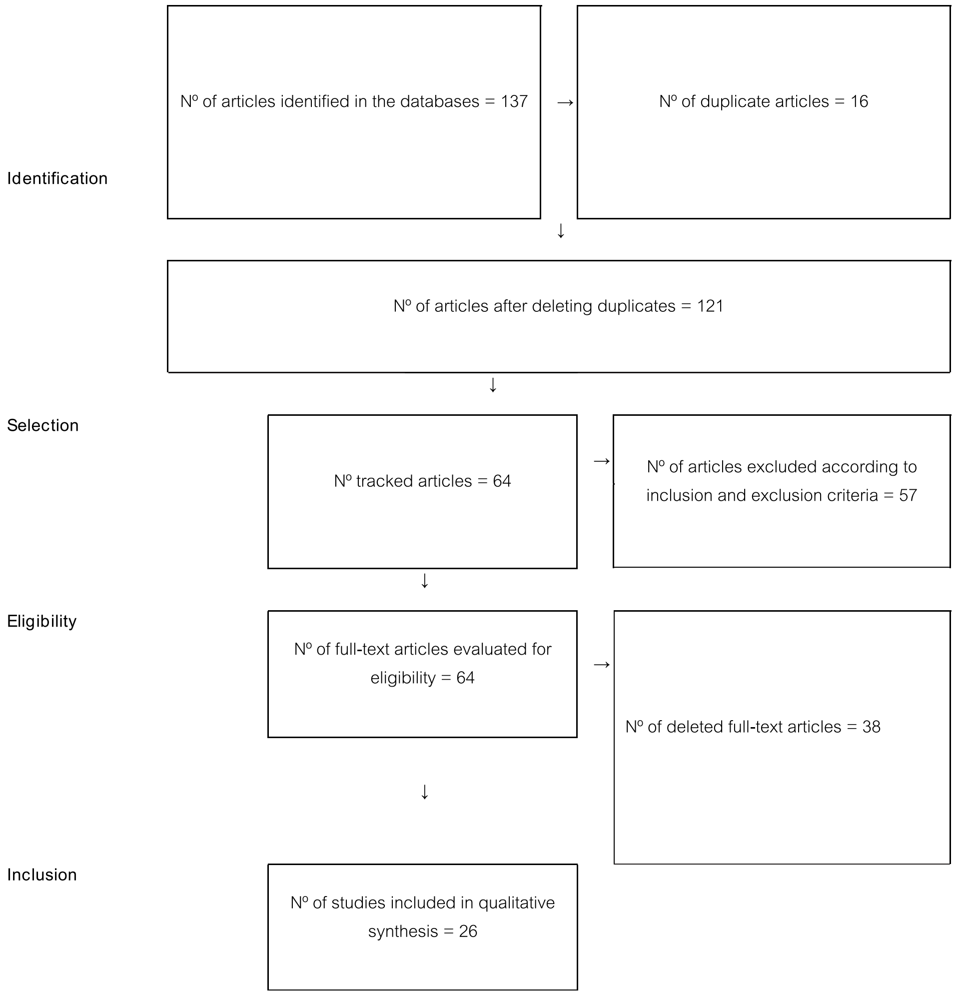


***Figure 5.*** *Flow of information with the different phases of the systematic review. Source: Authors.*

The analysis of the findings of the 26 reviewed studies on the use of eye-tracking in the assessment of the mental load of UAS operators reveals both important convergences and divergences among the authors. This discussion seeks to deepen these points by exploring the contributions of each study and how they interrelate.

*3.1.1 Convergences between the studies*

Eye tracking is widely recognised as a crucial tool for assessing mental load. McKinley et al. [15] demonstrated that the approximate entropy (ApEn) of pupil position is a more sensitive and consistent indicator of fatigue than PERCLOS (Percentage of Eyelid Closure), which measures the percentage of time an individual's eyelids are 80% or more closed. Their findings suggest that fatigue reduces the complexity of eye movements, likely due to longer fixations and slower saccades.

Similarly, Monfort et al. [16] identified pupil dilation, visual dispersion, and reaction time as key metrics for real-time workload prediction. These methods have proven effective, especially in complex and realistic simulation environments.

Roy et al. [17] broaden this perspective by investigating markers of engagement from oculomotor, cardiac, and brain data, finding that blink rate and decrease in the number of fixations are indicative of lower

***Chart 1.*** *Remaining studies are fully evaluated*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| McKinley et al. [15] | To evaluate the effectiveness of ocular metrics, specifically total duration of eye closure (PERCLOS) and approximate entropy (ApEn), as fatigue detectors in Air Force-relevant tasks after sleep deprivation. The study investigated the relationship between these ocular metrics and fatigue-induced performance declines, with the aim of developing reliable. monitoring. devices that can assess when an operator is excessively fatigued and thus mitigate the detrimental effects of fatigue in the execution of operational missions. | 10 volunteer participants (9 male, 1 female) were admitted to the study. They ranged in age from 18–42 years, with a mean age of 29 years. | Eye tracking (total duration of eye closure (PERCLOS) and the approximate entropy (ApEn) of eye movements). | The approximate entropy (ApEn) of pupil position was a more sensitive and consistent indicator of fatigue compared to the PERCLOS metric. As sleep deprivation increased, participants' performance on psychomotor surveillance and target identification tasks decreased significantly, and these reductions in performance were more closely correlated with changes in ApEn than with PERCLOS. This suggested that the complexity of eye movements decreased with fatigue, possibly due to longer eye fixations and slower saccades. | The approximate entropy metric (ApEn) of pupil position is a robust and sensitive indicator of fatigue, superior to the PERCLOS metric, especially in dynamic and noisy environments such as military aviation. The study suggested that ApEn can be used to develop real-time monitoring devices that help identify and mitigate the effects of fatigue on operators, thereby improving safety and performance in operational missions. |
| Monfort et al. [16] | Identify a set of metrics that can be used to assess the workload non-intrusively and in real-time to inform a dynamic task algorithm. This study sought to improve the configuration of UAV operations by allowing various responsibilities to be distributed among fewer pilots, continuously monitoring the operator's workload and adjusting tasks as needed. | 20 individuals (7 women, 13 men) ranging from 18–48 years of age ($M = 30$, $SD = 9.6$) participated in this experiment. | Eye tracking (pupil dilation, visual dispersion) | Pupil dilation (of the right and left eyes), visual dispersion, approximate entropy and reaction time had significant and independent effects on the operator's workload. These metrics were used to predict the 'live' workload at one-minute intervals, resulting in an overall accuracy of 78%. This level of accuracy suggests that physiological workload measurement methods developed in laboratory environments can be applied in more complex and realistic simulation settings. Dynamic workload detection has been identified as a potentially useful way to keep human operators in an optimal workload state, reducing the likelihood of operator errors and improving the efficiency of UAV operations. | Machine learning algorithms were able to accurately classify operators' workload states in a high-fidelity UAV simulation environment. The combination of pupillary dilation, visual dispersion, approximate entropy, and reaction time metrics made it possible to predict the workload of operators at one-minute intervals with high accuracy. These results supported the application of physiological workload measurement methods developed in the laboratory in complex and realistic simulation settings, suggesting that such approaches could be used to dynamically adjust operator tasks and improve operational performance. |

**Chart 1.** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Roy et al. [17] | Investigate which physiological markers could be used to estimate the mental state of operator engagement during prolonged operation of unmanned aerial vehicles (UAVs), the study sought to identify markers of engagement from oculomotor (eye tracking), cardiac (ECG) and brain (EEG) data to improve online estimation of operators' mental state. This aimed to optimise performance and increase the safety of UAV operations through the development of neuro-adaptive systems that adjust to the operator's mental state. | 5 volunteers underwent the experiment (2 males; 24.6 years old – sd 2.6) | Eye tracking (number and duration of fixations exceeding 80 ms and blink rate), EEG (electroencephalogram) and heart rate (HRV). | Operator engagement decreased significantly over time, as evidenced by an increase in blink rate and a decrease in the number of fixations. The power of the EEG alpha band increased after the first 40 minutes of operation, suggesting a decrease in vigilance and engagement. In addition, there was a significant correlation between response time, the cardiac LF/HF ratio, and the number of fixations, with longer response times associated with a lower LF/HF ratio and fewer eye fixations. These findings suggested that eye-tracking and cardiac metrics were effective indicators of mental engagement status during prolonged UAV operations. | The study highlighted that physiological markers, including oculomotor, cardiac and brain metrics, were effective indicators of operators' state of engagement during prolonged UAV operations. Increased blink rate, decrease in number of fixations and changes in EEG alpha-band activity were all associated with a decrease in engagement over time. In addition, the correlation between response time and cardiac LF/HF ratio, as well as the number of fixations, suggests that these markers can be used to monitor and potentially improve operator performance in real time. |
| Sibley et al. [18] | The study aimed to investigate the decision-making challenges involved in managing multiple unmanned aerial vehicles (UAVs) in a dynamic mission environment, utilising the Supervisory Control Operations User Testbed (SCOUT) to explore human performance in the supervised control of UAVs and to develop decision support algorithms. | 20 participants | Eye tracking, ECG (Electrocardiogram) and EEG (Electroencephalogram) | Investigate the challenges of decision-making in the management of multiple unmanned vehicles (UAVs) in a dynamic mission environment, using the Supervisory Control Operations User Testbed (SCOUT) to explore human performance in supervised control of UAVs and develop decision support algorithms. | The study found that physiological metrics, such as eye tracking and heart rate variability, could effectively predict operator performance in managing multiple unmanned aerial vehicles (UAVs). The integration of these metrics with performance data and mission context enabled real-time assessment of whether an operator would successfully complete a mission. Additionally, the study highlighted the ongoing development of route optimisation and decision bias algorithms within the SCOUT testbed, which could provide dynamic decision support by continuously evaluating operator decisions and suggesting adjustments when new information warranted a change in the mission plan. |

***Chart 1.*** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Sibley et al. [20] | Present and validate SCOUT, a flexible testbed developed by the Naval Research Laboratory (NRL) to investigate human performance and automation challenges in the future vision of supervisory control of multiple unmanned systems. SCOUT was designed with input from UAV operators to increase ecological validity and represent the complexity and uncertainty associated with controlling unmanned systems. | | Eye tracking, cardiac monitoring (ECG) and respiratory. | SCOUT is effective for investigating varied psychological phenomena such as decision-making, attention allocation and mission monitoring. The system enables the collection and synchronisation of performance data, eye tracking and real-time physiological monitoring, providing a comprehensive view of the operator's state. The use of SCOUT has shown that it is possible to detect when an operator last checked feeds from specific sensors and provide information about cognitive workload based on pupil size and other physiological parameters. This is crucial for understanding and mitigating complacency and cognitive overload in operations with multiple UAVs. | The conclusion of the study highlights that SCOUT is a powerful tool for conducting ecologically valid supervisory control experiments. It makes it easy to investigate how different independent variables affect system performance and operator state. SCOUT is available free of charge to other researchers, with the aim of improving the safety and management of supervisory control operations in the future. The integration of eye-tracking and physiological monitoring technologies enables continuous evaluation of operator performance, which is essential for the development of adaptive and effective automation systems. |
| Coyne et al. [19] | The goal for this study is to demonstrate that low-cost eye tracking can be used to measure changes in workload via increased pupil diameter as well as a more random gaze pattern as accessed by the Nearest Neighbor Index (NNI) | 19 (18 men and 1 woman) Navy and Marine Corps student pilots and flight officers aged 22–29 were recruited from Naval Air Station Pensacola, Florida. Participants were run in two groups. Eye tracking data for one of the participants was not recorded. | Eye tracking (pupil diameter, duration of fixations, and Nearest Neighbor Index (NNI)). | Participants' pupil diameter increased significantly during periods of high workload, indicating a higher mental demand. However, the duration of fixations did not show significant differences between the periods of low and high workload, suggesting that the allocation of visual attention remained relatively constant. In addition, the Nearest Neighbor Index (NNI) did not show significant changes between the different phases of the task, which may indicate that the dispersion of participants' gazes was not affected by the variation in workload. | Pupil diameter is an effective indicator of mental workload in supervisory control tasks, with significant increases during periods of high load. However, the duration of fixations and the Nearest Neighbor Index (NNI) did not show significant variations with the change in workload, suggesting that these metrics may be less sensitive to mental demand in this context. These results highlight the usefulness of low-cost eye tracking, particularly pupil diameter measurement, as a viable tool for assessing mental workload in automated supervision settings, although more research is needed to fully understand the relationship between gaze distribution and workload. |

***Chart 1.*** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Devlin and Riggs [22] | Develop a Markovian framework to analyse eye movement on different panels during the execution of a simulated Unmanned Aerial Vehicle (UAV) control task. The study aims to understand how participants transition between tasks and whether there are differences in eye scan patterns between individuals. | 10 undergraduate students participated in this study and had self-reported normal or corrected-to-normal vision (six males, four females; mean age = 21.4 years, standard deviation = 1.1 years). Three participants were dismissed due to eye tracker calibration issues. | Eye tracking (fixations and saccades). | The increased workload adversely affected the performance of the participants, but did not alter the individual eye scan patterns, which were analysed using a Markovian framework. Among the participants, five distinct eye scan patterns were identified, each with different levels of success in terms of response time and accuracy. The top four participants adopted different scanning patterns. These results suggest that eye tracking can provide unique insights into the differences in performance between individuals and can be used to develop algorithms that optimise performance by taking into account these individual differences. | Analysis of eye-tracking data using a Markovian framework can provide meaningful insights into the cognitive state of operators that are not captured by response time and accuracy measures alone. The identification of five different visual transition patterns among participants, with the top four performers adopting different patterns, suggests that there is no single optimal method for all individuals in data-rich environments. This emphasises the need to develop adaptive display systems that take into account individual differences, potentially improving performance in various complex areas such as the military, aviation, healthcare and the automotive industry. |
| Lim et al. [23] | Compare physiological metrics and task performance between different workload conditions when operating multiple UAVs (Unmanned Aerial Vehicles). With the advancement of UAV automation technology, one operator can control multiple UAVs simultaneously. However, the increase in the number of controlled UAVs can intensify the operator's workload and potentially reduce the performance in controlling UAVs. | 30 volunteers participated in the experiment. All experiment participants were males, and their average age (Korean age) was 24.6 (standard deviation: 2.22). | Eye tracking and heart rate data | Preliminary results indicate statistically significant differences between high and low workload conditions, observed in both eye-tracking data and heart rate data. This suggests that these metrics could be potential indicators of the operator's workload level. | The conclusion of the study points to the feasibility of constructing online workload metrics based on the comparative results obtained. Future research may develop these metrics to improve the supervision and control of multiple UAVs to optimise operator performance and operation efficiency. |

***Chart 1.*** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Jian et al. [24] | Propose a collaborative human-machine support scheduling system for the information intelligence of multiple unmanned aerial vehicles (UAVs) based on eye tracking. This system aims to improve operators' decision-making by taking into account the operator's attention to obtain the optimal solution in the sequence of images collected by UAVs. | – | Eye tracking | The proposed system, which incorporates the operator's attention through ocular attachment points, allowed for more efficient scheduling of images collected by UAVs. The heuristic algorithms developed proved to be effective, with an absolute error of approximately 1% in relation to the optimal solution. The experimental tests performed under six different distributions indicated that the proposed algorithm is not sensitive to the different distributions of processing time and has a negligible computational time. | The conclusion of the study is that the incorporation of eye attachment points into the human-machine collaborative scheduling system significantly improves the processing efficiency of the images collected by UAVs. The proposed heuristic algorithms are close to the optimal solution and show consistent performance under different processing time distributions. In addition, the study suggests the need to include the operator's workload in the system to further optimise the performance of the human-machine interaction system. |
| Lim et al. [8] | The objective of this work is to present the simulation environment of the Human Factors Engineering Laboratory (HFE-Lab) of RMIT University, detailing its network architecture and demonstrating its usefulness as a research tool in cognitive ergonomics and human factors engineering. The study aims to describe the different components of the HFE-Lab, including the air traffic management and pilot/remote simulation environment, as well as the psychophysiological data collection modules and scenario management tools. The research highlights an experimental case study that evaluates the workload of a single pilot during a circuit flight, validating HFE-Lab's data management functionality. | 1 licensed pilot | Eye tracking (gaze dispersion, fixation time, and gaze speed) and cardiorespiratory measurements (respiratory rate, interval between beats (RR), and heart rate). | The results of the study indicated that during the approach/landing and take-off phases, there was a high visual workload, characterised by low gaze dispersion and high fixation times. The analysis of cardiorespiratory data showed that heart rate variability (HRV) did not show a clear correlation with the different phases of flight, but rather with the circuit numbers. Eye-tracking metrics showed some ability to discriminate between phases of flight, while cardiorespiratory measurements did not show a variation consistent with phases of flight. | Preliminary results suggest that eye-tracking metrics can discriminate between different phases of flight, while cardiorespiratory measurements showed correlation with circuit numbers but not with phases of flight. The study validates HFE-Lab's ability to manage and analyse complex data from multiple sensors in real time. These findings indicate that the use of psychophysiological metrics, such as eye tracking, can provide meaningful insights into operator workload and performance, but also underscore the need for more complex experimental scenarios to effectively assess human-machine interactions in dynamic and demanding environments. |

***Chart 1.*** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Turpin et al. [21] | Develop and evaluate a system that allows a single crewmember to effectively manage multiple unmanned aerial systems (UAS) from a manned helicopter cockpit. The research aims to implement autonomous behaviors and human-centered design principles to alleviate the cognitive processing bottleneck experienced by operators when controlling multiple UAS simultaneously. The study evaluates these concepts through performance assessments in a multi-UAS manned/unmanned teaming (MUM-T) environment. | 16 U.S. Army aviators with prior MUM-T experience. | Eye tracking. | The research demonstrated that a single crewmember could manage at least three UAS assets while executing complex multi-UAS MUM-T tactical missions. The DelCon capability allowed participants to more efficiently perform a subset of mission tasks. Subjective ratings from participants indicated a willingness to accept the AAA and DelCon systems. Although few hypotheses were supported with statistical significance, the data suggested that the SCORCH system could improve performance by reducing mission duration times and maintaining or improving situational awareness and workload levels compared to conditions without DelCon and AAA. | The SCORCH system, comprising the DelCon behaviours and the Attention Allocation Aid (AAA), showed promise in increasing the efficiency, effectiveness and safety of an Air Mission Commander (AMC) performing advanced teaming or MUM-T missions. The automation can be refined further to enhance the AMC's battlefield capabilities while reducing exposure to harm. Participants quickly adapted to the transition from direct control to supervision of automation without an increase in workload, demonstrating the potential of using autonomous systems to manage cognitive bottlenecks and achieve greater system efficiency. |
| Lim et al. [10] | To present the concept of human-machine cognitive interfaces and interactions (CHMI2) for Unmanned Aircraft Systems (UAS) ground control stations. This CHMI2 system aims to improve the performance of operators by measuring cognitive states based on psychophysiological observables and adapting command, control and display functionalities. The research seeks to predict and improve operator performance in aviation tasks, increasing the efficiency and effectiveness of human-machine staff. | 3 RMIT aircraft/UAS pilots | Brain activity (EEG and fNIRS), heart rate (ECG), heart rate variability, breathing rate and volume, and eye tracking (EOG, blink rate, pupil dilation). | Specific eye-tracking variables, such as visual entropy, can discriminate between different control modes and task difficulty levels. A case study focused on the sensing and extraction layers showed that eye-tracking data is effective in discerning these differences. Another case study evaluated the adaptation of the GCS interface in Collision Separation and Avoidance (SA&CA) scenarios, showing that dynamic transitions in the interface can improve situational awareness and support automation in challenging scenarios with high time pressure or low-visibility obstacles. Preliminary results highlight the advantages of CHMI2 in improving the safety and efficiency of UAS operations. | The introduction of CHMI2 functionalities in future UAS can significantly reduce reaction time and improve the operational effectiveness of unmanned aircraft responses to collision and separation loss events. This, in turn, improves the safety and overall efficiency of operations. The cognitive adaptations of the interface and the complementary human-machine interactions provided by CHMI2 have demonstrated potential to enhance cooperation between SA&CA functions and human pilots, promoting better human-automation collaboration for safer and more efficient operations in non-segregated civil airspace. |

**Chart 1.** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Bektaş et al. [25] | Investigate how the use of Gaze Contingent Displays (GCDs) can improve visual search performance on large screens. The study seeks to understand the effectiveness of different visual perception models (contrast, color and depth perception) combined in a GCD, especially in geographic image interpretation tasks. | 39 participants (11 female, 16 with corrected vision, aged 23–45). | Eye tracking | Adjusted COMBO (GCD) models significantly improved participants' performance on the counting task, reducing completion time compared to uniform resolution displays and the COMBO model Specifically, participants completed counting tasks faster with the adjusted COMBO display, without compromising accuracy. In addition, participants' subjective evaluations in terms of image quality, visual comfort, confidence and perception of visual artifacts showed no significant differences between the different types of displays, suggesting that adjusted GCDs are as effective as uniform resolution displays | GCD models, especially adjusted COMBO, can improve performance in geographic imagery tasks without sacrificing perceived quality or visual comfort. The successful implementation of GCDs shows that it is possible to reduce spatial and chromatic detail at the periphery of vision without negatively impacting users' performance. These findings suggest that fine-tuned GCDs can be effective on large screens and in tasks that require intensive visual interpretation, such as counting targets in aerial imagery, providing improvements in task execution time without compromising accuracy |
| Foroughi et al. [26] | Determine how individuals perform their duties and allocate their visual attention when monitoring multiple automated screens with varying levels of reliability. The study investigated how variation in automation reliability affects participants' ability to detect automation failures, such as false alarms and detection failures, under both high and low workload conditions. The research seeks to better understand human-automation interaction and provide insights to improve the design of automated systems in supervisory control environments. | 96 students (M age ¼ 20.4 years, SD age ¼ 4.9 years, 70 females) from George Mason University participated in this research for course credit. | Eye tracking (pupil size and the total viewing time of each display). | Participants failed to detect automation failures (misses) approximately 2.5 times more often than false alarms, in both workload conditions. Participants' performance was worse in the high-workload condition compared to the low-workload condition. Automation fault detection remained relatively stable regardless of the reliability of the automation, except for a reduction in the false alarm detection rate in the high workload condition. The eye-tracking data revealed that participants distributed their attention relatively equally among the three automated sensor feeds throughout the experiment, suggesting a broad reliance approach to the system. | Operators' ability to detect automation failures is heavily influenced by the workload, resulting in significantly worse performance in high-load conditions. Automation reliability had a smaller impact on fault detection, with participants demonstrating similar performance at different levels of reliability. In addition, eye-tracking data revealed that participants adopted a systemic trust strategy, distributing their attention equally among the automated displays, regardless of their varying reliability. |

***Chart 1.*** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Devlin et al. [27] | Explore how patterns of peer-to-peer shared visual attention evolve over time in response to changes in workload. Specifically, the study uses a simulation environment to train multi-UAV (Unmanned Aerial Vehicles) missions and observe the dynamics of visual attention in situations of high cognitive demand. | 10 pairs of undergraduate students (20 students total) from Clemson University were recruited for the study (M = 21.3 years, SE = 0.24 years). | Eye tracking | The study analysed visual attention data in pairs of participants and followed the guidelines of ISO/TS 15007-2:2014, accepting a data loss of up to 15%. The results showed that the better-performing pairs had a greater overlap of the gaze, especially when the workload increased, indicating more effective visual coordination. In addition, the maximum percentage recurrence, which reflects the coordination of visual attention sequences, was also consistently higher for the best-performing pairs. This suggests that shared and well-coordinated visual attention patterns are crucial for performance on tasks that involve changes in workload. | This work suggests better performance during workload increases may be attributed to shared visual attention that is large in magnitude and consistent over time. The findings support the potential of technology to rely on these metrics, to inform and improve collaboration. |
| Moacdieh et al. [28] | Analyse how performance and attention allocation, evidenced by eye movements, are affected by gradual and sudden workload transitions compared to constant low or high workload in a multitasking environment. | 21 students participated in this study (13 men and 8 women; mean age = 20.9, SD = 1.5). | Eye tracking (pupil size, fixations and gaze dispersion). | There was no significant difference between sudden and gradual workload transitions in terms of performance or attention allocation. However, both load transitions altered the participants' strategy in executing the primary and secondary tasks, compared to constant low and high workloads. Under high workload, the performance of the primary task was worse, while under transitional conditions (sudden or gradual), the performance of the secondary task was better. Eye-tracking metrics revealed that under low workload, participants more widely distributed their visual attention, while under high workload and transition conditions, attention allocation was more focused and efficient. | Contrary to expectations, sudden workload transitions are no more detrimental to performance than gradual transitions. Both sudden and gradual transitions affected participants' strategy, leading to better performance on secondary tasks during transitions compared to constant workloads. In addition, eye-tracking metrics, especially those related to gaze dispersion and efficiency, provided valuable insights into how workload transitions influence attention allocation. These findings suggest the importance of considering varied eye-tracking metrics to better understand the effects of workload transitions on attention allocation in multitasking environments. |

**Chart 1.** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Niu et al. [29] | Propose a new method to monitor the performance of operators of multiple UAVs by assessing their workload levels and fatigue through eye movement patterns. The research aims to detect potential operator errors that may be caused by abnormal conditions, fatigue or overload, using machine learning techniques to classify eye movement patterns and analyse characteristics such as the proportion of serving time and pupil closing time. | | Eye tracking (saccade rate, eye movement patterns, pupil closure time ratio) | The main results of the study showed that eye movement patterns can be used effectively to monitor the performance and workload of operators of multiple UAVs. Analysis of eye-tracking data revealed that the ratio of withdrawal time and pupil closure time are effective indicators of fatigue and cognitive overload. In addition, the detection of abnormal states, such as the tunnel effect, was possible through the analysis of ocular fixation patterns. The proposed method significantly reduced operational errors, demonstrating the effectiveness of early detection of fatigue and abnormal conditions at different levels of task complexity. | The conclusion of the study indicates that the proposed method for monitoring the performance of operators of multiple UAVs, using eye movement patterns and machine learning techniques, is effective in detecting fatigue and cognitive overload. The analysis of eye movements, such as saccade rate and pupil closure time, provided an accurate assessment of the cognitive status of the operators. The results showed that the implementation of this method can reduce operational errors and improve the safety and efficiency of UAV operations by allowing dynamic adjustments to the operator's workload. The study suggests that future research should focus on applying this method in real-world operational scenarios to further validate its effectiveness. |
| Planke et al. [30] | Develop and validate a cyber-physical-human (CPH) system that uses accessible physiological sensors and artificial intelligence techniques to measure the mental workload (MWL) of operators in real-time during multiple UAS operations. | 5 participants that took part in the experiment comprising of 4 males and 1 female. The participants were aerospace students at Royal Melbourne Institute of Technology (RMIT) University and were selected based on their prior experience in aviation and aerospace engineering. | EEG (electroencephalogram) and eye tracking. | The mental workload (MWL) of operators increases with the complexity and number of UAVs managed. The EEG data showed significant fluctuations in the frontal and parietal regions of the brain during more complex missions. Eye-tracking analysis revealed that stationary gaze entropy and fixation duration are effective predictors of operator performance, with greater gaze dispersion associated with poorer performance. The study demonstrated that combining physiological measurements with artificial intelligence techniques can significantly improve workload understanding and management in complex multi-UAV operations. | The developed cyber-physical-human (CPH) system has been shown to be effective in measuring mental workload (MWL) in real time during operations with multiple UAVs. The integration of physiological sensors, such as EEG and eye tracking, combined with artificial intelligence techniques, has made it possible to monitor and adapt the human-machine interface to optimise operators' performance. The results showed moderate to high correlations between EEG measures, visual entropy, and task performance, validating the use of these metrics to infer the cognitive status of the operators. The fusion of physiological data has improved the accuracy of workload estimates, indicating that this approach may be beneficial for future applications in complex operating environments. |

**Chart 1.** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Devlin et al. [31] | Determining whether eye-tracking analysis, specifically the type of visual attention (focal or ambient), can be a non-invasive and quantitative measure to better understand how operators are impacted by workload transitions over time, known as the workload history effect. The study investigated how participants' visual attention patterns affected their performance in an unmanned aerial vehicle (UAV) command and control scenario, where the workload transitions gradually and suddenly. The eye-tracking metric, called the K-coefficient, is used to distinguish between focal and ambient visual attention, providing insights into how operators manage the demands of variable tasks in complex environments. | 21 undergraduate students participated in this study (13 male; mean age = 20.9 years, SD = 1.5). | Eye tracking (fixations and saccades). | Results of the study indicate that there is an effect of workload history on the performance of participants, but it depends on the level of workload, the load transition rate and the elapsed time. During periods of low workload, response times were faster at first and increased over time, especially in gradual transitions. Accuracy was higher at the beginning and decreased over time in gradual transitions, while remaining stable in sudden transitions. The analysis of the K coefficient showed that visual attention was more focused during periods of high workload and more dispersed during periods of low workload. | Eye-tracking technology, specifically through the K-coefficient metric, can be a powerful tool for understanding how workload transitions affect operators' visual attention and performance over time. The results suggest that inadequate visual strategies adopted during periods of low workload may have immediate and long-lasting adverse effects on performance. The K-coefficient metric has been shown to be sensitive to changes in workload, effectively distinguishing between focal and ambient attention. This indicates that eye tracking can be used in real-time to detect and potentially mitigate the negative effects of workload history, informing the design of adaptive technologies for complex and dynamic work environments. |
| Foroughi et al. [32] | Evaluate performance, confidence and allocation of visual attention while monitoring a near-perfect automated system. This study seeks to understand how humans interact with highly reliable systems, which will be essential in high-risk environments, by evaluating three information flows essential for the success of monitoring and detecting rare failures in automation: performance, subjective confidence and allocation of visual attention. | 73 students with normal or corrected to- normal vision (M age = 20.5 years, SD age = 4.2 years, 51 females) from George Mason University participated in this research for course credit. | Eye tracking (specifically metrics such as total fixation time, number of visual transitions, and entropy of gaze transition) | 34% of participants correctly identified the 'miss' automation failure and 67% correctly identified the 'false alarm'. Subjective confidence increased when participants did not detect the flaws and decreased when they did. Those that detected the 'false alarm' showed a more complex visual scan pattern in the 2 minutes around the automation failure, as well as longer fixation times and more frequent transitions to the central sensor feed. | This work highlights the human limitations in monitoring near-perfect automated systems, quantifying subjective experience and human attentional cost. The results emphasise the need to reassess the role of the operator in future high-risk environments and to understand the human being at an individual level in order to design systems that best meet the needs of each operator. It is recommended to collect multiple measurements at the operator level in real-time to monitor their status and provide individualised assistance. |

***Chart 1.*** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Sibley et al. [33] | Demonstrate how pupillometry and gaze transition analysis can augment the evaluation of user interface visualisations. | 119 U.S. Naval and Marine Corps officers (Mage = 23.82, SDage = 2.12, 26 females) enrolled in aviation school in Pensacola, FL, voluntarily participated in this experiment. | Eye tracking (pupil diameter, fixations, saccades, and gaze transition). | Scanpaths between team members were more similar during high-workload tasks than low-workload tasks, indicating stronger visual coordination under pressure. The analysis revealed that the similarity of gaze trajectories positively correlated with task performance, suggesting that teams with more aligned visual scanning patterns tended to be more effective. In addition, it was observed that pupillary dilation increased with workload, confirming the use of pupillometry as an indicator of mental load. These findings highlight the importance of eye-tracking metrics for understanding performance dynamics in collaborative tasks, offering valuable insights for interface design and team training. | Comparing scanpaths provides valuable insights into the performance of team tasks, especially under different workload levels. Teams exhibited greater similarity in gaze trajectories during high-workload tasks, which was associated with better performance. In addition, pupillary dilation increased with workload, validating pupillometry as an effective indicator of mental load. These results highlight the usefulness of eye-tracking metrics to improve team coordination and efficiency in collaborative environments, suggesting that scan path analysis can be a powerful tool for interface design and training strategies in contexts of high cognitive demand. |
| Singh et al. [34] | Characterise and estimate mental workload (MW) in a Manned-Unmanned Teaming (MUM-T) scenario using physiological signals and evaluate the impact of the validation procedure on classification accuracy. This study was designed to provide an estimate of the mental workload through the classification applied to these physiological measures and to consider the impact of non-stationarity of physiological signals on the accuracy of the classification. | 14 healthy volunteers (6 females, 8 males; 24.4 y.o. ± 1.95) | Electroencephalogram (EEG), electrocardiogram (ECG) and eye tracking. | The main results of the study indicate that mental workload (MW) had a significant impact on all measures analysed. The analysis showed a significant increase in heart rate (HR) and a decrease in heart rate variability (HRV) with increasing workload EEG measurements revealed an increase in the potency of the beta and gamma bands, as well as an increase in the Engagement Index (EI) with increasing workload. In addition, pupil dilation increased, and the average duration of fixations decreased with increasing workload. Intra-subject classification achieved an accuracy of 75% using ECG features alone or in combination with EEG and ET, using the AdaBoost, Linear Discriminant Analysis (LDA) or Support Vector Machine (SVM) classifiers. However, accuracy dropped significantly with eco-validation, indicating the need for additional developments to monitor mental workload in operational scenarios. | The conclusion of the study highlights those physiological characteristics, such as EEG, ECG and eye-tracking measurements, are effective in estimating mental workload in MUM-T scenarios. The accuracy of the classification was significantly affected by the non-stationarity of the physiological signals, with a notable drop in accuracy when using ecological validation compared to traditional validation. These results underline the importance of developing more robust methods for monitoring mental workload in real-world operational contexts, potentially exploring techniques such as learning transfer, automatic trait selection, deep learning and co-learning to improve the accuracy of mental workload estimation. |

***Chart 1.*** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Schwerd and Schulte [18] | To develop a framework for estimating the cognitive state of a pilot in real time, in order to assess attention allocation and situational awareness (SA) in a Manned-Unmanned-Teaming (MUM-T) application. This approach aims to identify potential flaws in situational awareness that could cause performance decrements and errors. The work also presents the design of a MUM-T cockpit simulator to describe how this cognitive state estimation framework is integrated into a human-autonomy team environment, aiming to improve interaction and cooperation between pilots and automated systems. | 15 participants (11 males, 4 females) in the age between 22 and 33. The participants did not have any flying experience to ensure equal difficulty in the flight task. | Eye tracking (gaze direction and fixations, to infer visual attention allocation and situational awareness). | The main results of the study indicate that the proposed cognitive state estimation model has a moderate predictive capacity, with an overall accuracy of 83.9% in predicting participants' correct answers in a situational awareness assessment questionnaire (SAGAT). However, the model also generated a high proportion of false positives, suggesting the need for improvements in the perception model and the robustness of eye tracking. The results highlighted that, despite moderate predictive accuracy, the model was able to identify situations where participants failed to recognise important information, suggesting that notifications based on this estimate could improve operators' performance. | The study suggests that estimating the operator's cognitive state, based on eye-tracking measures and situational awareness models, could be an effective tool to improve human-autonomous interaction in complex MUM-T environments. The approach showed moderate predictive performance, with the ability to identify situations in which participants failed to recognise important information, which could have been used for adaptive interventions. |
| Devlin et al. [35] | Determine whether scanning pattern-based eye tracking metrics can predict performance trends during workload transitions. The study investigates whether operators' visual attention allocation patterns can inform both theory and design guidelines for workload transitions by examining various eye-tracking metrics as predictors in performance growth curve models in an unmanned aerial vehicle (UAV) test environment. | 60 student Naval Aviators participated in this study (age: $M = 24.5$ years, $SD = 2.3$ years, 51 males). | Eye tracking (stationary gaze entropy, average fixation duration, gaze transition rate and spatial gaze density). | The results of the study indicate that eye tracking metrics based on scan patterns are significant predictors of performance trends during workload transitions. Stationary gaze entropy was a consistent predictor of mean performance and trends over time for all transition rates, indicating that greater gaze dispersion was associated with poorer performance. During slow transitions, the mean duration of fixation and the entropy of the stationary gaze were predictors of response time and accuracy. In medium transitions, the entropy of the stationary gaze was a predictor of the average performance. For fast transitions, stationary gaze entropy and average fixation duration predicted average accuracy, while gaze transition rate predicted accuracy trends over time. | Eye tracking metrics based on scan patterns are significant predictors of performance trends during workload transitions. Stationary gaze entropy was a consistent predictor of mean performance and trends over time for all transition rates, indicating that greater gaze dispersion was associated with poorer performance. During slow transitions, the mean duration of fixation and the entropy of the stationary gaze were predictors of response time and accuracy. In medium transitions, the entropy of the stationary gaze was a predictor of the average performance. For fast transitions, stationary gaze entropy and average fixation duration predicted average accuracy, while gaze transition rate predicted accuracy trends over time. |

***Chart 1.*** *(Continued)*

| Authors (year) | Objective | Participants | Physiological parameters analysed | Main results | Conclusion |
|---|---|---|---|---|---|
| Dalilian and Nembhard [36] | To investigate the underlying affective behaviours related to the acquisition of drone piloting skills, with the goal of developing methods to improve human performance. Electroencephalography (EEG) and eye-tracking instruments are used to measure human affect in a series of simulated drone piloting experiments, examining performance through behavioural variables, controller input variables, as well as measures of individual cognitive ability. | The average age of the participants was 21.3 years; 10 were males and 5 were females | Electrical activity of the brain as measured by EEG and eye tracking. | Biometric and behavioural measures significantly affected task completion times. Approximately half of the vertical control's performance profile was determined by individual differences. Cognitive load and focused attention had a significant effect on completion times of vertical control and transfer and landing tasks. Introducing additional task weight increased difficulty and negatively impacted early performance, but continued practice improved completion times for both tasks, both at standard and increased difficulty levels. | The study concluded that drone assistance technologies and the increasing use of drones for civilian purposes raise interesting questions about pilots' interactions with drones and their training. Modeling pilots' control actions and mental states is essential for the successful adaptation of drones in Industry 4.0. The study identified biometric and behavioral indicators for drone pilots' skill learning at different tasks and difficulty levels. A potential next step could be the development of AI-based training paradigms that incorporate these factors as inputs to create engaging and personalised learning experiences. |
| El Iskandarani et al. [37] | To explore whether and to what extent the similarity of the eye-scan of two people working together on a task is linked to their performance in a complex environment and multitasking in different workloads. | 10 pairs of undergraduate students at the University of Virginia (20 students total) were recruited for the study ($M = 21.3$ years, $SE = 0.24$ years). | Eye tracking (pupil diameter, frequency and duration of blinks, average duration of fixations, amplitude of saccadic movements and the Nearest Neighbor Index (NNI)) | Participants exhibited greater pupillary dilation when facing mathematical problems under high workload, indicating a higher mental load. Analysis of the eye-tracking data revealed that during high levels of workload, participants experienced fewer fixations on the relevant region of the screen and shorter duration fixations compared to lower workloads. In addition, the dispersion of fixations was lower under high load, suggesting a more focused allocation of attention. These results indicate that eye-tracking metrics, such as pupil size, duration of fixations, and gaze dispersion, are predictive of situational awareness and can be used to dynamically measure operators' mental workload and attention allocation on complex tasks. | Eye-tracking metrics, such as pupil size, duration of fixations, and gaze dispersion, are effective tools for assessing mental workload and attention allocation in complex task environments. Workload transitions, both sudden and gradual, impact operators' attention strategies, but there is no significant difference in overall performance between these types of transitions. The use of these metrics can provide valuable insights for interface design and operator training, helping to improve situational awareness and operational efficiency in high cognitive demand scenarios. |

mental engagement during prolonged UAV operations. Sibley et al. [18] and Coyne et al. [19] explore heart rate variability (HRV) and other eye metrics, such as pupil size, to monitor mental load, concluding that these measurements can be used to predict whether an operator will be able to successfully complete the mission.

The application of eye tracking in dynamic environments is highlighted by several studies. Sibley et al. [20] present SCOUT, a testbed designed to investigate human performance and automation challenges, demonstrating its effectiveness in detecting when an operator has scanned specific sensor feeds, and providing insight into cognitive workload based on pupil size. Turpin et al. [21] demonstrate that a single crew member can manage multiple UASes in complex tactical missions, with the help of automated systems that improve operational efficiency and safety.

There was also a consensus on the correlation between eye tracking and other physiological measures. Lim et al. [10] combine eye-tracking with EEG and ECG to assess cognitive states and adapt command-and-control functionalities, showing that specific eye-tracking variables, such as visual entropy, can discriminate between different control modes and task difficulty levels. Singh et al. [34] highlight that pupil dilation and the average duration of fixations decrease with increasing workload, suggesting that these metrics are effective in estimating mental load.

### *3.1.2 Divergences between the studies*

The variations in measurement methods employed across the reviewed studies reflect different approaches to assessing mental workload in UASoperators. McKinley et al. [15] used approximate entropy as a metric to detect signs of fatigue, focusing on the complexity of eye movements as a response to cognitive demand. In contrast, Coyne et al. [19] focused on pupil diameter and the Nearest Neighbor Index (NNI) to measure cognitive effort and gaze dispersion, respectively. While approximate entropy can capture subtle variations in the regularity of eye movements, pupil diameter is associated with changes in mental effort, and NNI provides information on the spatial distribution of eye fixations. These differing methodological choices suggest that there is no consensus on the most appropriate metrics for assessing mental workload, reflecting the diversity of approaches available in the literature.

In addition to measurement methods, the contexts in which the studies are conducted also vary widely. Studies such as those by Devlin et al. [31] Sibley et al. [20] focused on military scenarios where operations are characterised by high complexity, requiring maximum attention and cognitive performance from operators. In these studies, mental workload is often associated with situations that demand rapid decision-making and the simultaneous execution of multiple tasks, which can affect both the workload measurements, and the results obtained.

In contrast, studies exploring applications in simulation and training environments, such as those by Devlin and Riggs [22] and Niu et al. [29], operate under controlled conditions that allow for the manipulation and control of specific variables. These simulation environments are designed to replicate critical aspects of real-world operations but differ in terms of stressors present in real operational scenarios, such as time pressure and the unpredictability of situations. The difference between these contexts can impact the validity of the results obtained in simulation studies when compared to real-world operational scenarios.

The diversity in methodological choices and application contexts reflects the inherent complexity of research on mental workload in UAS operators. Each methodological approach and operational context brings with it a specific set of advantages and limitations that influence both data collection and the interpretation of results. For example, while pupil diameter measurement may be sensitive to rapid changes in cognitive load, approximate entropy might capture the evolution of fatigue over time. Similarly, application in military versus simulation scenarios can lead to results that vary not only in precision but also in practical relevance.

These methodological and contextual variations raise questions about the comparability of studies. The absence of standardisation in mental workload measurement metrics and the experimental conditions used may hinder the construction of a cohesive knowledge base and the extrapolation of results to different operational scenarios. Standardising metrics and harmonising application contexts could facilitate comparison between studies and synthesis of results, contributing to a more integrated understanding of mental workload in UAS operators.

These divergences highlight the importance of considering both the nature of the measurement methods and the operational context when interpreting study results. The choice of metrics and the environment in which the study is conducted can have significant implications for the findings and the conclusions that can be drawn about operators' mental workload. Therefore, when evaluating the literature on mental workload in UAS, it is essential to account for this diversity to better understand how different approaches may complement or contrast with one another.

#### *3.1.3 Results on operational efficiency*

Some studies, such as the one by Monfort et al. [16], show high accuracy in predicting the "live" workload with the use of eye tracking, while others, such as that of Devlin et al. [27], highlight the complexity of predicting performance trends during workload transitions. This suggests that the effectiveness of eye tracking may vary depending on the experimental conditions and study design.

Wanyan et al. [38] introduced a multidimensional perspective in the assessment of mental workload by combining eye tracking with behavioural and physiological measures for a more complete understanding of the mental state of pilots. The authors broadened the scope of application of eye tracking by focusing on mental workload prediction, which highlights the importance of adaptive flight interfaces and procedures to avoid cognitive overload. Gomolka et al. [39], Rudi et al. [40] and Schriver et al. [41] explored the applicability of eye tracking to better understand pilots' attention, pointing to improvements in training and interface design.

The discussion deepened with Pongsakornsathien et al.(42) and Singh et al.(34) who investigated the use of eye tracking in operations with UAVs and human-machine systems, respectively, suggesting new possibilities for optimising cooperation and operational efficiency. Lefrançois et al. [43], Li et al. [44, 45], Scannella et al. [46] and Yu et al. [47] underscored the value of eye tracking in pilot training and incident investigation while Diaz-Piedra et al. [48] and Lounis et al. [49] discussed fatigue detection and the impact of experience on operational efficiency.

Lim et al. [10] conclude this comprehensive review by introducing cognitive human-machine interfaces for UAS, marking a significant advance in adapting air operations to the cognitive needs of pilots.

The reviewed studies offer a comprehensive overview of the potential and limitations of eye tracking in assessing the mental load of UAS operators. The convergence in findings highlights the usefulness of this technology as a valuable tool for improving safety and operational efficiency. However, methodological and contextual divergences underline the need for standardisation and a more integrated approach that considers multiple sources of physiological data for a more accurate and holistic assessment. Future research should focus on harmonising the metrics and exploring the applicability of eye tracking in diverse operational contexts to maximise its effectiveness. These investigations promise to enhance the expertise and effectiveness of UAS operators and pave the way for future innovations in aviation.

### ***3.2 Evaluation of the quality of the studies found***

Chart 2 presents the classification of the articles analysed in the systematic review, evaluated based on three main criteria: score (S), study population (P) and study design (D).

- Score (S) represents the congruence of the publications with the research terms. Articles were rated based on how well their titles, abstracts and keywords aligned with the predefined research terms.
- Study Population (P) assesses the adequacy of the research groups concerning the scope of the study. The scoring for this criterion is divided as follows:

***Chart 2.*** *Assessment of the quality of the studies analysed in full*

| Authors (year) | Score (S) | Population studied (P) | Study design (D) | Total (S + P + D) |
|---|---|---|---|---|
| McKinley et al. [15] | 22 | 1 | 5 | 28 |
| Monfort et al. [16] | 6 | 2 | 5 | 13 |
| Roy et al. [17] | 0 | 1 | 5 | 6 |
| Sibley et al. [18] | 33 | 0 | 5 | 38 |
| Sibley et al. [20] | 16 | 0 | 5 | 21 |
| Coyne et al. [19] | 28 | 2 | 5 | 35 |
| Devlin and Riggs [22] | 27 | 2 | 5 | 34 |
| Lim et al. [23] | 8 | 2 | 5 | 15 |
| Jian et al. [24] | 20 | 0 | 5 | 25 |
| Lim et al. [8] | 3 | 0 | 5 | 8 |
| Turpin et al. [21] | 11 | 3 | 5 | 19 |
| Lim et al. [10] | 6 | 1 | 5 | 12 |
| Bektaş et al. [25] | 11 | 3 | 5 | 19 |
| Foroughi et al. [26] | 6 | 3 | 5 | 14 |
| Devlin et al. [27] | 8 | 2 | 5 | 15 |
| Moacdieh et al. [28] | 32 | 3 | 5 | 40 |
| Niu et al. [29] | 17 | 0 | 5 | 22 |
| Planke et al. [30] | 14 | 1 | 5 | 20 |
| Devlin et al. [31] | 25 | 3 | 5 | 33 |
| Foroughi et al. [32] | 21 | 3 | 5 | 29 |
| Sibley et al. [33] | 28 | 3 | 5 | 36 |
| Singh et al. [34] | 12 | 2 | 5 | 19 |
| Schwerd and Schulte [50] | 3 | 3 | 5 | 11 |
| Devlin et al. [35] | 15 | 3 | 5 | 23 |
| Dalilian and Nembhard [36] | 21 | 2 | 5 | 28 |
| El Iskandarani et al. [37] | 27 | 2 | 5 | 34 |

  - Unspecified or irrelevant population: 0 points
  - Minimum specification: 1 point
  - Specified but not very representative: 2 points
  - Adequately specified: 3 points
  - Very representative: 4 points
  - Exceptionally specified and representative: 5 points
- Study Design (D) refers to the methodology used in the articles, with the following scoring system:
  - Experimental studies: 5 points
  - Quasi-experimental studies: 3.5 points
  - Cross-sectional studies: 3 points
  - Control case studies: 2.5 points
  - Case studies: 2 points
  - Narrative reviews and expert opinions: 1 point

These criteria were applied to ensure a comprehensive and objective analysis of the studies, which allowed for a consistent evaluation of their quality based on methodological rigor, relevance, and the alignment of their research focus with the terms used in this systematic review.

***Chart 3.*** *PRISMA checklist*

| Section and topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| TITLE | | | |
| Title | 1 | Identify the report as a systematic review. | Page 1 |
| ABSTRACT | | | |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | Page 1 |
| INTRODUCTION | | | |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | Page 2 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Page 2 |
| METHODS | | | |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Page 3 |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | Page 3 |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | Page 3 |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | Page 4 |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | Page 4 |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | Page 4 |
| | 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | Page 4 |

***Chart 3.*** *(Continued)*

| Section and topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | Page 4 |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | Page 4 |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | Page 4 |
| | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | Page 4 |
| | 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | Page 4 |
| | 13d | Describe any methods used to synthesise results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | Page 4 |
| | 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | Page 4 |
| | 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesised results. | Page 4 |
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | Page 4 |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | Page 4 |
| RESULTS | | | |
| Study selection | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. | Page 11 |
| | 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. | Page 11 |
| Study characteristics | 17 | Cite each included study and present its characteristics. | Page 11 |

**Chart 3.** *(Continued)*

| Section and topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| Risk of bias in studies | 18 | Present assessments of risk of bias for each included study. | Page 11 |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. | Page 11 |
| Results of syntheses | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | Page 12 |
| | 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | Page 12 |
| | 20c | Present results of all investigations of possible causes of heterogeneity among study results. | Page 12 |
| | 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesised results. | Page 12 |
| Reporting biases | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | Page 28 |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | Page 28 |
| DISCUSSION | | | |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence. | Page 32 |
| | 23b | Discuss any limitations of the evidence included in the review. | Page 32 |
| | 23c | Discuss any limitations of the review processes used. | Page 33 |
| | 23d | Discuss implications of the results for practice, policy, and future research. | Page 33 |
| OTHER INFORMATION | | | |
| Registration and protocol | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered. | – |
| | 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | Page 33 |
| | 24c | Describe and explain any amendments to information provided at registration or in the protocol. | – |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | – |

***Chart 3.*** *(Continued)*

| Section and topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| Competing interests | 26 | Declare any competing interests of review authors. | – |
| Availability of data, code and other materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | Page 33 |

The StArt software (State of the Art through Systematic Review), used to manage the selection of articles, automatically generates the Score (S), which represents the congruence of the articles with the predefined research terms. This score is based on the match between the titles, abstracts and keywords of the articles with the terms of the research in question.

However, good articles may receive a score of zero. This can happen when a relevant article for the field of study does not present titles, abstracts or keywords that directly match the predefined research terms. This situation reflects the limitations of a purely textual search, as high-quality articles may be excluded due to the lack of strict alignment with the terms used in the search process. Therefore, it is important to recognise that, although the score provided by the software is useful for initial filtering, it should not be the sole criterion for exclusion or inclusion, as it may fail to capture the more complex nuances of certain studies' relevance to the research.

The studies of McKinley et al. [15] and Coyne et al. [19] stood out for their high scores, reflecting a strong congruence with research terms, well-defined populations and robust methodologies. McKinley et al. [15] presented an in-depth analysis of approximate entropy (ApEn) as an indicator of fatigue, while Coyne et al. [19] focused on pupil diameter and the NNI to measure mental load.

The studies of Sibley et al. [20] and Devlin et al. [31] also received high scores, highlighting the effectiveness of SCOUT, a testbed designed to investigate human performance and automation challenges in UAS operations. These studies have demonstrated the ability of eye tracking to provide valuable data on operators' cognitive load and attention allocation.

On the other hand, some studies, such as those by Jian et al. [24] and Lim et al. [23], received lower scores in terms of the population studied, indicating a need for greater specification and representativeness of the research groups. However, these studies still contributed significantly to the understanding of mental load in UAS operations.

The variability in scores reflected methodological and contextual differences between studies. Studies such as those by Monfort et al. [16] and Roy et al. [17] have used specific ocular metrics, such as pupil dilation and blink rate, to predict the real-time workload and mental engagement of operators, highlighting the usefulness of these measurements in complex simulation settings.

Devlin and Riggs [22] used a Markovian framework to analyse eye-scan patterns, providing insights into individual differences in operator performance. Studies such as those by Niu et al. [29] have proposed the use of machine learning techniques to classify eye movement patterns and detect states of fatigue and cognitive overload, showing the applicability of eye tracking in various operational contexts.

The studies also varied in terms of application contexts. Sibley et al. [20] and Devlin et al. [31] focused on military scenarios and highly complex operations, while others, such as Devlin et al. [27] and Foroughi et al. [32], explored human-automation interaction in supervisory control environments. This diversity of contexts reinforces the versatility of eye tracking, although it highlights the need for standardisation in the metrics used to assess mental load.

The analysis of the quality of the reviewed studies evidenced the methodological robustness and relevance of the findings for the assessment of the mental load of UAS operators. The highest quality studies provided detailed insights into mental load indicators and highlighted the importance of rigorous methodologies and well-defined populations. However, the variability in scores and application contexts indicated the need for standardisation of metrics and a more integrated approach that considers multiple sources of physiological data for a more accurate and holistic assessment. Future research should focus on harmonising the metrics and exploring the applicability of eye tracking in diverse operational contexts to maximise its effectiveness.

## 4.0 Final thoughts

The findings of this systematic review highlighted the relevance and methodological robustness of studies investigating the use of eye tracking as a tool to assess the mental workload of UAS operators. The diversity and complexity of the studied contexts demonstrated the versatility of eye tracking in capturing critical nuances of cognitive load, especially in demanding environments such as military operations and air traffic control.

High-quality studies, such as those by Lefrançois et al. [43] provided detailed insights into mental workload indicators and emphasised the importance of rigorous methodologies and well-defined populations. These works showed a strong correlation between specific ocular metrics and cognitive load, validating the use of eye tracking as a reliable indicator. Additionally, research by Devlin et al. [31] and Sibley et al. [20] highlighted the effectiveness of systems like SCOUT and CHMI2, which combine physiological sensors with artificial intelligence techniques to improve workload management in complex operations.

In contrast, studies such as those by Behrend and Dehais (2020) and Scannella et al. [46], which received lower scores, indicated the need for greater specificity and representativeness in the studied populations. However, even these studies contributed significantly to understanding mental workload, suggesting methodological improvements and standardisation of the metrics used.

The variability in study scores reflected the methodological and contextual differences. Studies like those by Monfort et al. [16] and Roy et al. [17] used ocular metrics such as pupil dilation and blink rate to predict real-time workload, highlighting the utility of these measures in complex simulation environments. Devlin and Riggs [22] applied a Markovian framework to analyse eye scan patterns, providing valuable insights into individual differences in operator performance.

One potential limitation of this review process was the reliance on studies available in specific databases and the exclusion of non-English publications, which might have resulted in a selection bias. Additionally, variations in the methodologies and metrics used across different studies could have influenced the comparability and generalisability of the findings. The review protocol used in this study is available upon request from the authors.

Future research should focus on harmonising metrics and exploring the applicability of eye tracking in various operational contexts. Integrating multiple sources of physiological data would provide a more precise and holistic assessment of mental workload, contributing to the development of more intuitive interfaces and training programmes that mitigate cognitive overload, thus enhancing the safety and effectiveness of operations.

In conclusion, eye tracking is a valuable and promising tool for assessing the mental workload of UAS operators. The research underscored the importance of rigorous methodologies and well-defined populations in understanding the nuances of mental workload in this specific context. Furthermore, the consistency of the results supported the use of eye tracking as a reliable indicator of mental workload, allowing for the improvement of cognitive human-machine interfaces and suggesting a fertile field for future investigations.

## References

[1] Fricke, T. and Holzapfel, F. An approach to flight control with large time delays derived from a pulsive human control strategy, AIAA Atmospheric Flight Mechanics Conference, 2016. https://doi.org/10.2514/6.2016-1033

[2] Alaimo, A., Esposito, A., Orlando, C. and Simoncini, A. Aircraft pilots workload analysis: Heart rate variability objective measures and NASA-task load index subjective evaluation, *Aerospace*, 2020, **7,** (9). https://doi.org/10.3390/aerospace7090137

[3] Hart, S.G. and Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, *Human Mental Workload*, 1988, pp 139–179. https://doi.org/10.1007/s10749-010-0111-6

[4] NASA, FAA Improve Air Traffic Safety, Efficiency. *Aviation Space Environ. Med.*, 2000, **71**, (4), p 463. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0034024167&partnerID=40&md5=9c6afb4e4543e62212561f9bab26af06

[5] National Aeronautics and Space Administration. *NASA TLX: Task Load Index*, 2020.

[6] Yiyuan, Z., Tangwen, Y., Dayong, D. and Shan, F. Using NASA-TLX to evaluate the flight deck design in design phase of aircraft, *Procedia Eng.*, 2011, **17**, pp 77–83. https://doi.org/10.1016/j.proeng.2011.10.010

[7] Zheng, Y.Y., Yin, T.W., Dong, D.Y. and Fu, S. Using NASA-TLX to evaluate the flight deck design in design phase of aircraft, in Shuguang, Z., Shuiting, D., Yanlai, Z. and Junmin, D. (Eds), 2nd International Symposium on Aircraft Airworthiness (ISAA), vol. 17, 2011. https://doi.org/10.1016/j.proeng.2011.10.010. WE - Conference Proceedings Citation Index - Science (CPCI-S).

[8] Lim, Y., Gardi, A., Ramasamy, S., Vince, J., Pongracic, H., Kistan, T. and Sabatini, R. A novel simulation environment for cognitive human factors engineering research, AIAA/IEEE Digital Avionics Systems Conference - Proceedings, 2017-September, 2017a. https://doi.org/10.1109/DASC.2017.8102126

[9] Sibley, C., Foroughi, C., Brown, N., Drollinger, S., Phillips, H. and Coyne, J. Augmenting traditional performance analyses with eye tracking metrics, in A. H. and A. U. (Eds), *Advances in Intelligent Systems and Computing: Vol. 1201 AISC*, Springer, 2021a, pp 118–125. https://doi.org/10.1007/978-3-030-51041-1_17

[10] Lim, Y., Ramasamy, S., Gardi, A., Kistan, T. and Sabatini, R. Cognitive human-machine interfaces and interactions for unmanned aircraft, *J. Intell. Robot. Syst. Theory Appl.*, 2018a, **91,** (3–4), pp 755–774. https://doi.org/10.1007/s10846-017-0648-9

[11] Greenhalgh, T. How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses), *BMJ*, 1997, **315,** (7109), pp 672–675. https://doi.org/10.1136/bmj.315.7109.672

[12] Kitchenham, B. and Charters, S. Guidelines for performing Systematic Literature Reviews in Software Engineering, 2007. https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf %3E

[13] Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P. and Stewart, L.A.S. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement, *Jpn Pharmacol. Ther.*, 2019, **47,** (8), pp 1177–1185.

[14] Fabbri, S., Silva, C., Hernandes, E., Octaviano, F., Di Thommazo, A. and Belgamo, A. Improvements in the StArt tool to better support the systematic review process, ACM International Conference Proceeding Series, 01-03-June-2016, 2016. https://doi.org/10.1145/2915970.2916013

[15] McKinley, R.A., McIntire, L.K., Schmidt, R., Repperger, D.W. and Caldwell, J.A. Evaluation of eye metrics as a detector of fatigue, *Hum. Factors*, 2011, **53,** (4), pp 403–414. https://doi.org/10.1177/0018720811411297

[16] Monfort, S.S., Sibley, C.M. and Coyne, J.T. Using machine learning and real-time workload assessment in a high-fidelity UAV simulation environment, *Next-Gener. Anal. IV*, 2016, **9851**, p 98510B. https://doi.org/10.1117/12.2219703

[17] Roy, R.N., Bovo, A., Gateau, T., Dehais, F. and Carvalho Chanel, C.P. Operator engagement during prolonged simulated UAV operation, *IFAC-PapersOnLine*, 2016, **49,** (32), pp. 171–176. https://doi.org/10.1016/j.ifacol.2016.12.209

[18] Sibley, C., Coyne, J., Avvari, G.V., Mishra, M. and Pattipati, K.R. Supporting multi-objective decision making within a supervisory control environment, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9744, 2016, pp 210–221. https://doi.org/10.1007/978-3-319-39952-2_21

[19] Coyne, J.T., Sibley, C., Sherwood, S., Foroughi, C.K., Olson, T. and Vorm, E. Assessing workload with low cost eye tracking during a supervisory control task, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10284 11th, 2017, pp 139–147. https://doi.org/10.1007/978-3-319-58628-1_12

[20] Sibley, C., Coyne, J. and Thomas, J. Demonstrating the supervisory control operations user testbed (SCOUT), *Proc. Hum. Factors Ergon. Soc.*, 2016, pp 1323–1327. https://doi.org/10.1177/1541931213601306

[21] Turpin, T., Surana, A., Alicia, T. and Taylor, G.S. *Removing the bottleneck: Utilizing autonomy to manage multiple UAS sensors from inside a cockpit*, 2018, 22. https://doi.org/10.1117/12.2303915

[22] Devlin, S.P. and Riggs, S.L. Analyzing eye tracking data using a Markovian framework to assess differences in scan patterns, Proceedings of the Human Factors and Ergonomics Society, 2017-October, 2017, pp 1814–1818. https://doi.org/10.1177/1541931213601935

[23] Lim, H.-J., Choi, S.-H., Oh, J., Kim, S., Lee, J.-W., Kim, B.S., Kim, S. and Yang, J.H. Comparison study of potential workload index in a simulated multiple-UAV operation environment, 17th International Conference on Control, Automation and Systems (ICCAS), 2017.

[24] Jian, L., Yin, D., Shen, L. and Niu, Y. Human machine collaborative support scheduling system of intelligence information from multiple unmanned aerial vehicles based on eye tracker, *J. Shanghai Jiaotong Univ. (Sci.)*, 2017, **22,** (3), pp 322–328. https://doi.org/10.1007/s12204-017-1838-0

[25] Bektaş, K., Çöltekin, A., Krüger, J., Duchowski, A.T. and Fabrikant, S.I. GeoGCD: Improved visual search via gaze-contingent display, Eye Tracking Research and Applications Symposium (ETRA), 2019. https://doi.org/10.1145/3317959.3321488

[26] Foroughi, C.K., Sibley, C., Brown, N.L., Rovira, E., Pak, R. and Coyne, J.T. Detecting automation failures in a simulated supervisory control environment, *Ergonomics*, 2019, **62,** (9), pp 1150–1161. https://doi.org/10.1080/00140139.2019.1629639

[27] Devlin, S.P., Flynn, J.R. and Riggs, S.L. How shared visual attention patterns of pairs unfold over time when workload changes, Eye Tracking Research and Applications Symposium (ETRA), 2020. https://doi.org/10.1145/3379156.3391339
[28] Moacdieh, N.M., Devlin, S.P., Jundi, H. and Riggs, S.L. Effects of workload and workload transitions on attention allocation in a dual-task environment: Evidence from eye tracking metrics. *J. Cogn. Eng. Decision Making*, 2020, **2020**, (2).
[29] Niu, J., Wang, C., Niu, Y. and Wang, Z. Monitoring the performance of a multi-UAV operator through eye tracking, Proceedings - 2020 Chinese Automation Congress, CAC 2020, 2020, pp 6560–6565. https://doi.org/10.1109/CAC51589.2020.9326955
[30] Planke, L.J., Lim, Y., Gardi, A., Sabatini, R., Kistan, T. and Ezer, N. A cyber-physical-human system for one-to-many UAS operations: Cognitive load analysis, *Sensors (Switzerland)*, 2020, **20,** (19), pp 1–21. https://doi.org/10.3390/s20195467
[31] Devlin, S.P., Byham, J.K. and Riggs, S.L. Does what we see shape history? Examining workload history as a function of performance and ambient/focal visual attention, *ACM Trans. Appl. Percept.*, 2021, **18,** (2). https://doi.org/10.1145/3449066
[32] Foroughi, C.K., Brown, N.L., Sibley, C. and Coyne, J.T. Near-perfect automation: Investigating performance, trust, and visual attention allocation, *Human Factors*, 2021, **65,** (4), pp 546–561. https://doi.org/10.1177/00187208211032889
[33] Sibley, C., Foroughi, C., Brown, N., Drollinger, S., Phillips, H. and Coyne, J. Augmenting traditional performance analyses with eye tracking metrics, Advances in Intelligent Systems and Computing, 1201 AISC, 2021b, pp 118–125. https://doi.org/10.1007/978-3-030-51041-1_17
[34] Singh, G., Chanel, C.P.C. and Roy, R.N. Mental workload estimation based on physiological features for pilot-UAV teaming applications, *Front. Hum. Neurosci.*, 2021, **15**. https://doi.org/10.3389/fnhum.2021.692878
[35] Devlin, S.P., Brown, N.L., Drollinger, S., Sibley, C., Alami, J. and Riggs, S.L. Scan-based eye tracking measures are predictive of workload transition performance, *Appl. Ergon.*, 2022, **105**. https://doi.org/10.1016/j.apergo.2022.103829
[36] Dalilian, F. and Nembhard, D. Biometrically measured affect for screen-based drone pilot skill acquisition, *Int. J. Hum.-Comput. Interact.*, 2023. https://doi.org/10.1080/10447318.2023.2208991
[37] El Iskandarani, M., Atweh, J.A., McGarry, S.P.D., Riggs, S.L. and Moacdieh, N.M. Does it multiMatch? What scanpath comparison tells us about task performance in teams, *J. Cogn. Eng. Decision Making*, 2023, **17,** (3), pp 294–309. https://doi.org/10.1177/15553434231171484
[38] Wanyan, X., Zhuang, D. and Zhang, H. Improving pilot mental workload evaluation with combined measures, *Bio-Med. Mater. Eng.*, 2014, **24,** (6), pp 2283–2290. https://doi.org/10.3233/BME-141041
[39] Gomolka, Z., Kordos, D. and Zeslawska, E. The application of flexible areas of interest to pilot mobile eye tracking, *Sensors (Switzerland)*, 2020, **20,** (4). https://doi.org/10.3390/s20040986
[40] Rudi, D., Kiefer, P., Giannopoulos, I. and Raubal, M. Gaze-based interactions in the cockpit of the future: A survey, *J. Multimodal User Interfaces*, 2020, **14,** (1), pp 25–48. https://doi.org/10.1007/s12193-019-00309-8
[41] Schriver, A.T., Morrow, D.G., Wickens, C.D. and Talleur, D.A. Expertise differences in attentional strategies related to pilot decision making, *Human Factors*, 2008, **50,** (6), pp 864–878. https://doi.org/10.1518/001872008X374974
[42] Pongsakornsathien, N., Lim, Y., Gardi, A., Hilton, S., Planke, L., Sabatini, R., Kistan, T. and Ezer, N. Sensor networks for aerospace human-machine systems, *Sensors (Switzerland)*, 2019, **19,** (16). https://doi.org/10.3390/s19163465
[43] Lefrançois, O., Matton, N. and Causse, M. Improving airline pilots' visual scanning and manual flight performance through training on skilled eye gaze strategies. *Safety*, 2021, **7,** (4). https://doi.org/10.3390/safety7040070
[44] Li, J., Oksama, L. and Hyönä, J. Close coupling between eye movements and serial attentional refreshing during multiple-identity tracking, *J. Cogn. Psychol.*, 2018, **30,** (5–6), pp 609–626. https://doi.org/10.1080/20445911.2018.1476517
[45] Li, W.C., Zhang, J., Le Minh, T., Cao, J. and Wang, L. Visual scan patterns reflect to human-computer interactions on processing different types of messages in the flight deck, *Int. J. Indus. Ergon.*, 2019, **72**, pp 54–60. https://doi.org/10.1016/j.ergon.2019.04.003
[46] Scannella, S., Peysakhovich, V., Ehrig, F., Lepron, E. and Dehais, F. Assessment of ocular and physiological metrics to discriminate flight phases in real light aircraft, *Hum. Factors*, 2018, **60,** (7), pp 922–935. https://doi.org/10.1177/0018720818787135
[47] Yu, C.S., Wang, E.M.Y., Li, W.C., Braithwaite, G. and Greaves, M. Pilots' visual scan patterns and attention distribution during the pursuit of a dynamic target, *Aerospace Med. Hum. Perform.*, 2016, **87,** (1), pp 40–47. https://doi.org/10.3357/AMHP.4209.2016
[48] Diaz-Piedra, C., Rieiro, H., Suárez, J., Rios-Tejada, F., Catena, A. and Di Stasi, L.L. Fatigue in the military: Towards a fatigue detection test based on the saccadic velocity, *Physiol. Meas.*, 2016, **37,** (9), pp N62–N75. https://doi.org/10.1088/0967-3334/37/9/N62
[49] Lounis, C., Peysakhovich, V. and Causse, M. Visual scanning strategies in the cockpit are modulated by pilots' expertise: A flight simulator study, *PLoS ONE*, 2021, **16,** (2). https://doi.org/10.1371/journal.pone.0247061
[50] Schwerd, S. and Schulte, A. Operator state estimation to enable adaptive assistance in manned-unmanned-teaming, *Cogn. Syst. Res.*, 2021, **67**, pp 73–83. https://doi.org/10.1016/j.cogsys.2021.01.002

**Cite this article:** Russo A.C., Cardoso Junior M.M. and Villani E. (2025). Eye-tracking analysis to assess the mental load of unmanned aerial system operators: systematic review and future directions. *The Aeronautical Journal*, **129**, 529–558. https://doi.org/10.1017/aer.2024.122