Contemporary Ergonomics and Human Factors 2026. Eds. D Golightly, R Charles & F Dixon, CIEHF.

Beyond the Lab: Ten-Day Eye Tracking and Cardiovascular Monitoring in Live Air-Traffic Control

Jediah Clark, Wing-Yee Cheung, Lisa Aldridge & Adam Smith

NATS, UK

SUMMARY

As the largest investigation of its kind in UK Air Traffic Control (ATC), 25 participants took part in a ten-day continuous pilot trial using desktop eye tracking, heart rate monitors and sleep diaries across multiple shifts in NATS Swanwick Terminal Control. The study's primary objective was to assess the technical viability for measuring real-time biometrics in a live ATC operational environment, with a secondary objective of validating these measures against other environmental and behavioural factors. Although challenges were present, a wealth of data was collected from this study, showing that there is potential in real-time, non-invasive, data collection and assimilation in live-ATC. This contribution highlights recent developments in biometric devices and the key challenges associated with real-time biometric measurement. This paper presents initial findings on how these insights may support more advanced approaches to monitoring workload, fatigue, and future application opportunities.

KEYWORDS

Air Traffic Control, Eye Tracking, Workload, Fatigue

Introduction

As the largest investigation of its kind in UK Air Traffic Control, this study provides a unique contribution by directly addressing technical challenges and practical applications for biometrics in a live ATC operation, across multiple shifts and times of day, for one of the world's busiest portions of airspaces.

In operational environments, subjective measures for workload and fatigue are often relied upon due to their low cost and relative ease of application (Bartulović et al., 2023). Such measures are often inconsistent, may cost staff time during their breaks, and do not provide the data precision required to progress human factors towards a data-informed age of predicting high workload and fatigue. As such, alternative measures such as eye tracking and cardiovascular measures should be considered for use in real-time ATC operations.

Historically, due to the types of devices being available on the market, biometric recording devices have been deemed as being 'invasive'. However, recent technological developments have meant that eye trackers can be integrated into workstations with minimal interference, and sports-watches can be worn to track a wide range of biometrics with minimal interference to controller workflow. If successfully integrated, the introduction of 'objective' human performance data would better enable ATC operators to react tactically to evolving safety-critical situations or use this information strategically to inform future improvement projects.

This study outlines the technical boundaries facing integration into operations and assesses the validity of biometrics for gaining insights into human performance, specifically workload and fatigue.

Method

Twenty-five participants took part in a ten-day continuous pilot trial of using heart-rate monitors (n = 21), sleep diaries (n = 22), and desktop eye tracking (n = 25) across multiple shifts in NATS Swanwick Terminal Control (6f, 19m). Controllers were an average of 44.4 years of age (min = 31, max = 56., std = 7.32). The study featured 24-hour data collection from volunteers covering the entire rostering cycle of morning, afternoon and night shifts.

Apparatus

Two workstations (Swanwick Terminal Control South-West) were each equipped with two eye trackers (Tobii Pro Fusions; Tobii AB, 2026) for ten-days, one on top of the radar display, and one on top of the flight-strip panel (EXCDS). Eye trackers were connected to a desktop PC, installed with the Tobii Python SDK, behind the workstations where 60Hz resolution data was streamed and stored in an SQLite database.

Controllers wore Polar Ignite 3 Sport watches (Polar Electro, 2026) to track heart rate and were asked to 'track activity' before each controlling session to enable datapoints to be provided every second.

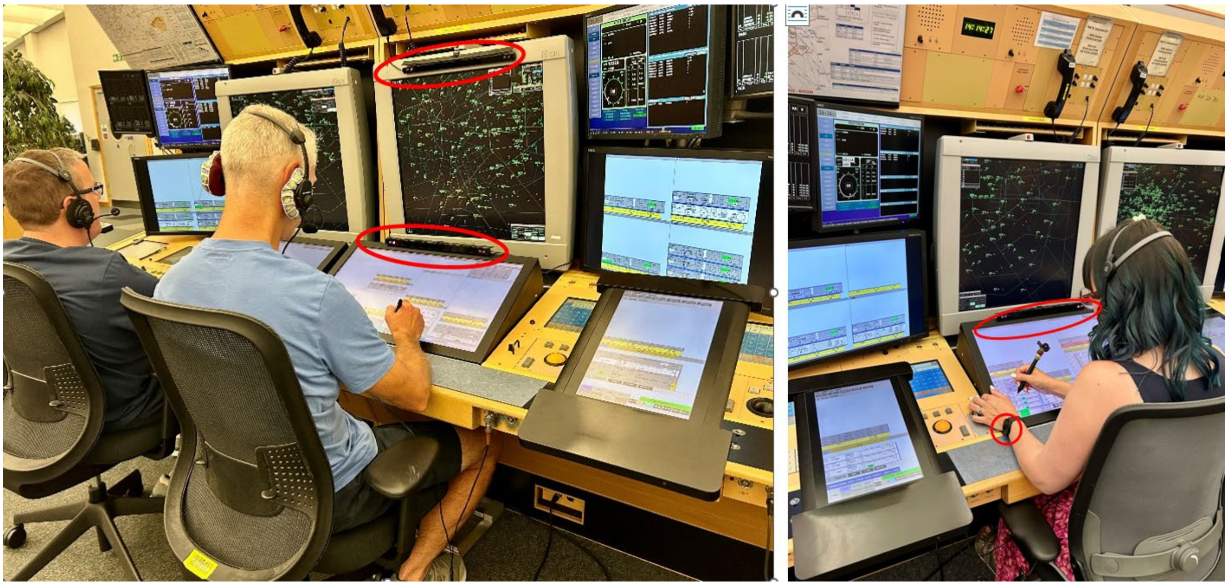

Figure 1. Controller TC workstations with eye trackers and sports watches highlighted in red

Method of Analysis

Workload was assessed through the measurement of pupil diameter, a well-documented workload correlate (van der Wel & van Steenbergen, 2018) that when baselined can indicate percentage increased and decreases of workload from that baseline. Pupil diameter baselines were calculated as the 95th percentile for that session (indicating the lowest workload reading for future comparisons). Pupil diameter used interpolation to remove blinks and consisted of a centred 1-minute rolling mean to filter out noise. Additionally, heartrate was tracked at a second-by-second interval when tracking was enabled by participants.

Using methods employed in automobiles (Ji et al., 2004), eye openness provided the potential for tiredness detection by using a one-minute rolling mean eye-openness (with a baseline of 5th percentile being calculated as 'fully open'). Percentage Closure (PERCLOS) indicated the % at which an eye was open over time. A threshold was set at 20% for PERCLOS before being flagged as a 'tiredness indicator'. Additionally, long blink events were classified as anything above 500ms, providing a second indicator of tiredness.

Subjective workload, fatigue and situation awareness were collected for every session across the ten-days to support findings and provide additional context and measure validation. The questionnaires used were adapted from the Bedford Workload Scale (Roscoe, 1984), the Karolinska Sleepiness Scale (Åkerstedt & Gillberg, 1990) and the Situation Awareness Rating Technique (SART; Taylor, 1990) and are widely used at NATS when time availability of operational controllers is limited. Additionally, the workload measure asks controllers to provide 'average' and 'peak' workload scores on a 1-10 scale, and the difference of these scores can reflect the 'peakiness' of the controlling session where peaks deviate more greatly from the average.

Additionally, controllers completed sleep diaries over the 10-day period, and risk factors associated with shift patterns were calculated by FAID (Johnston, 2025), a model that predicts controller fatigue risk based on controller rest opportunities and roster.

Given that this was a real-time operational trial, where realities of controlling meant that controllers were not always looking at radar, for a session to be considered for workload and fatigue analysis, a minimum of 60% of datapoints must have been populated when participants were thought to be looking at radar (cross-checked with researcher observations).

Results

Eye Tracker Integration

It was hypothesised that due to the reality of an ATC operational environment, the trial would suffer data-loss due to operational demands and controller posture. A logistic regression found that the angle in which controller's set their display had a significant effect on eye tracking data quality (β = -17%, p = .007). For every 10 degrees increase in EXCDS height, data quality experiences an additional 17% data loss.

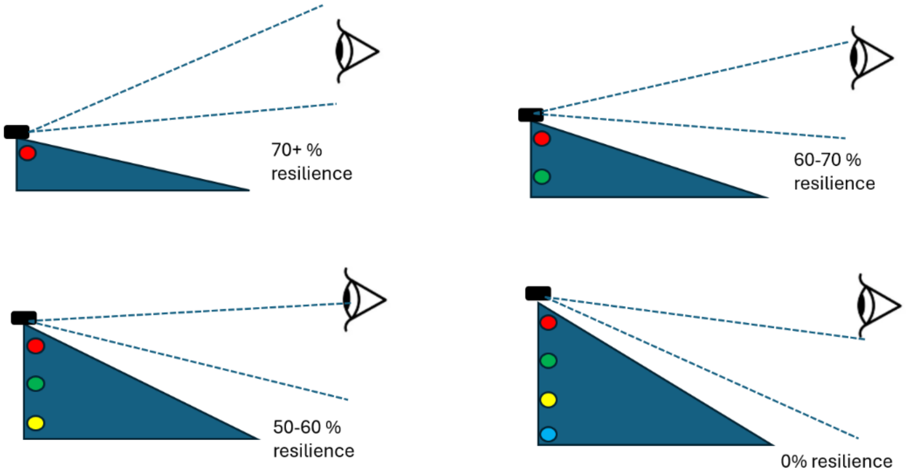

Figure 2. Change of data quality based on angle of EXCDS panel

Workload

A multiple regression predicting maximum pupil diameter (% deviation from baseline) was significant, $F(7, 24) = 4.17$, $p = .004$, explaining 54.9% of the variance ($R^2 = .55$, adj. $R^2 = .42$). Subjective average-peak difference ($p < .05$), frequency of aircraft switching ($p < .05$), and number of controller actions ($p < .05$) were significant positive predictors.

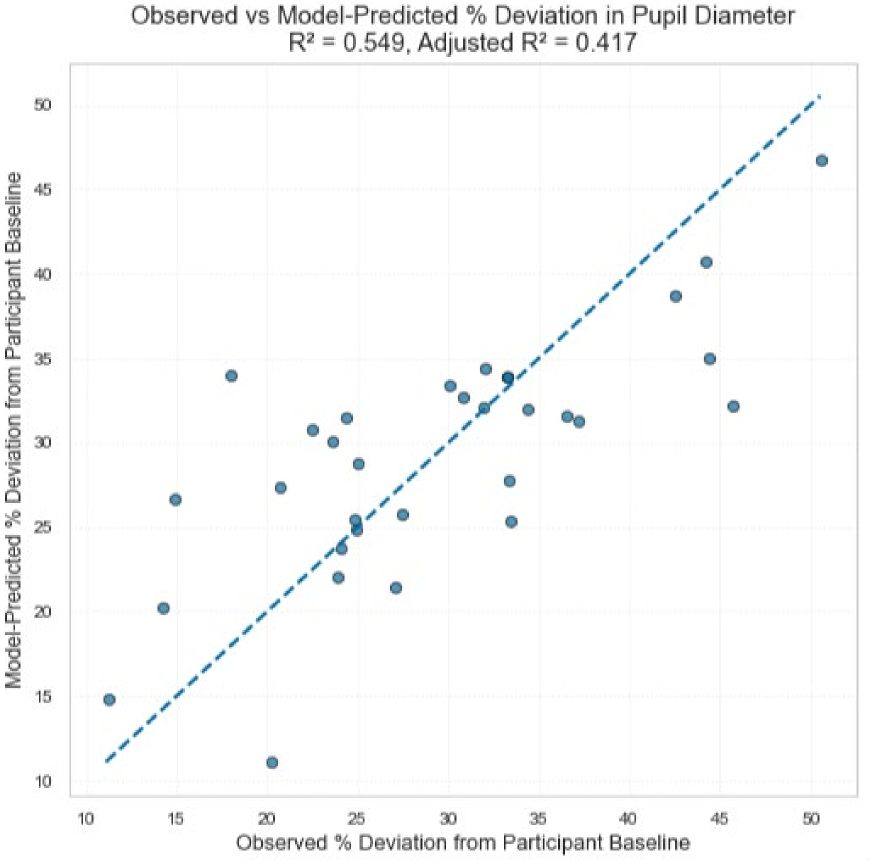

Figure 3. Multiple Linear Regression using Subjective Workload responses and number of tasks to predict pupil-diameter peak.

Eye Openness

Due to the rarity of tiredness indicators in the dataset, it was not deemed appropriate to run statistical modelling for eye-openness data. However, data quality for eye-openness was high, as this metric operates in parallel with pupil detection and therefore does not incur the same drawbacks. When tiredness indicators did occur during the trial, they represented a clear deviation from normal.

One night session showed pronounced signs of tiredness and difficulty maintaining alertness. This session lasted 2 hours and 40 minutes. In the final 25 minutes, partial eye closure exceeded the 20% threshold, averaging 40% (double the set threshold), with a peak of 60%. In total, 32 long-blink events were recorded in this session. The controller reported workload as “very low,” indicating that they were unlikely to have been required to perform any tasks during this period, which likely contributed to the observed tiredness indicators.

By contrast, all other sessions over the 10-day period showed fewer than two long blink events, and eye closure consistently remained below 20%. Figure 4 shows this session (bottom of image) alongside a “normal” session (top of image) for comparison.

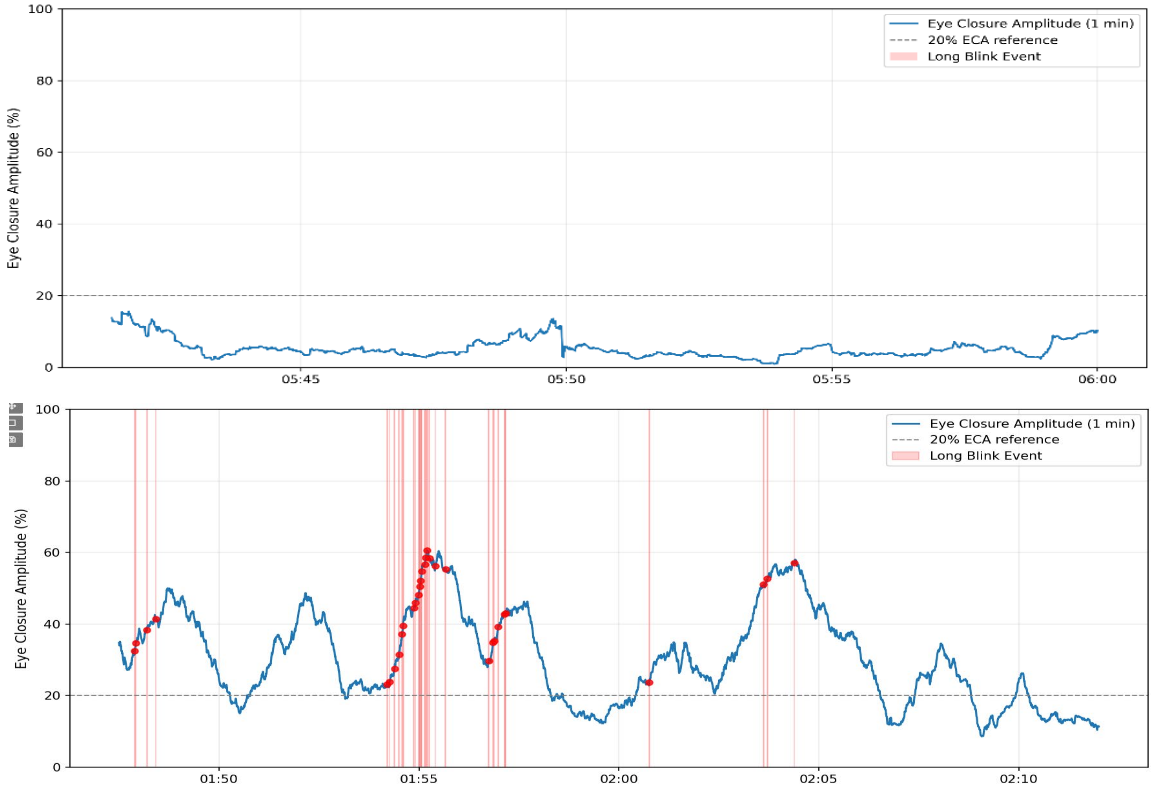

Figure 4. Timeline for sessions showing no signs of tiredness (top) and signs of tiredness during a night-shift (bottom). Dotted line indicates PERCLOS threshold. Red lines indicate long blink events.

Heart Rate

Heart rate was included as part of a larger mixed-effects model that included the amount of sleep recorded in participant sleep diaries, and calculated fatigue risk from FAID rostering analysis. The mixed-effects model predicting subjective fatigue (KSS) was significant, $\chi^2(6) = 69.35$, $p < 0.001$, explaining 49% of the variance ($R^2 = .49$, *adj.* $R^2 = .47$). Poorer sleep quality ($p < .001$), higher rostering risk factor ($p < .001$), heart rate outside of moderate zone ($p < .01$), and lower subjective average workload ($p < .01$) were significant predictors.

Discussion

For those working in high-intensity operational environments, the challenges of collecting real-time operational biometric data are profound. It is hoped that this contribution inspires and informs the human-factors community who face similar challenges in their sector, where their operational environment may not be representative of a lab-controlled environment.

Technological Implementation

Although data quality is significantly influenced by workstation configuration and how controllers interact with their environment, we propose several technological requirements to improve reliability and feasibility for real-time operational deployment:

1. Bespoke hardware mounts capable of maintaining optimal sensor angles, even when screens or equipment are adjusted.
2. Automated user identification to dynamically calculate an individual's 95th percentile baseline pupil diameter.

3. Dynamic gaze-to-traffic alignment, enabling an ‘autocalibration’ feature to improve gaze-coordinate accuracy without the need for lab style manual calibration.
4. Sports watches that do not require start-stop functionality to enable high resolution data collection.
5. Sports watches that are operationally friendly and feature high battery-life to enable better data coverage.

It is important to note that, depending on device type, certain eye tracking and cardiovascular measures are better suited to specific use cases than others. Some data streams impose stringent tracking requirements and therefore lose data richness if minimum thresholds are not met (e.g., headbox constraints, eyelid occlusion, or heartrate sensors requiring manual start stop input).

Eye-openness data demonstrated the greatest reliability, making it a dependable *reactive* metric. In contrast, the noisier nature of pupil-diameter and heart-rate data suggests they may be better suited for longer-term monitoring of controller workload and fatigue. Heart-rate data quality improved when controllers remembered to manually start tracking; consequently, devices that automate this process may be more appropriate in operational settings where time and cognitive resources are limited.

Theoretical Discussion

Although biometric measures were modelled against task measures and subjective responses, biometric data may be more closely aligned with controller performance than other human performance indicators. Models that perform poorly may reflect the fundamental distinction between *taskload* and *workload* (SKYbrary, 2026), and the extent to which system level factors—such as traffic complexity or high stakes events—drive physiological responses independently of taskload. These influences may remain undetected in recorded datasets.

It is therefore essential that contextual information is consistently incorporated into discussions of biometric outputs, and that additional system-level metrics (e.g., incidents, startles, operational pressures) are included in ongoing investigations. In complex systems such as air traffic control, neat correlations are unlikely, and oversimplified interpretations can be misleading.

Enabling future eye tracking

Future work will expand data collection to other ATC operational environments (e.g., airports, Area Control). An especially promising direction involves the mathematical classification of gaze behaviours (McClung & Zhang, 2016), allowing trainees, On the Job Training Instructors (OJTIs), and safety specialists to better understand visual strategies, workload indicators, and peripheral awareness.

For example, Figure 5, adapted from McClung and Kang (2016), illustrates how scan patterns from the trial can be translated into insights regarding controller visual behaviour. Systems capable of performing this analysis dynamically and automatically could enable real-time insights into controller performance and identify areas for skill development.

The long-term objective of this work is to equip a wide range of ATC disciplines with high-fidelity human performance data to support more informed tactical and strategic decision-making, and to monitor long-term trends associated with new technology or airspace changes. Achieving this requires resolving foundational questions of technical feasibility. Through collaboration between industry, academia, and suppliers, NATS has demonstrated the significant potential of biometrics in real-time operations, where safety, efficiency, and staff wellbeing remain central.

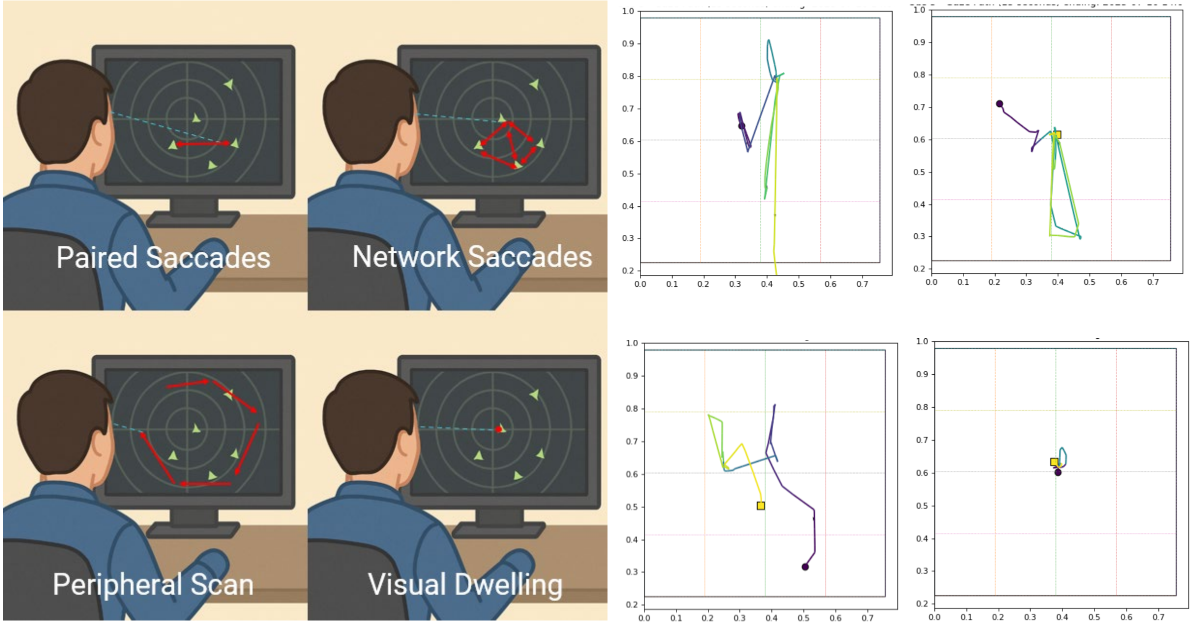

Figure 5. Example visual behaviours (left) and associated visual scan patterns captured during the operational trial (right).

Conclusion

This ten-day live-operations study demonstrates that continuous, non-invasive biometric monitoring—combining desktop eye-tracking, heart-rate sensing, and sleep diaries—can be integrated into air-traffic control without disrupting workflow, while revealing both the promise and practical limits of real-time human performance assessment. Pupil-diameter modelling, supported by operational and subjective measures, accounted for meaningful variance in workload; eye-openness proved a reliable reactive indicator for rare alertness lapses; and mixed-effects analyses showed that heart-rate features, when considered alongside sleep quality and roster-derived fatigue risk, predict subjective sleepiness.

At the same time, screen-angle–dependent data loss and manual device workflows highlight design targets for operational maturity, including stable mounting, automated user identification/session start, and dynamic gaze-to-traffic auto-calibration. More broadly, the findings emphasise the need to interpret biometrics through the lens of system context and to distinguish taskload from workload when modelling complex sociotechnical environments. Future work will extend to other ATC settings and advance automated gaze-pattern classification and device automation to deliver dependable human-performance telemetry that supports safer operations and more informed decision-making.

References

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience, 52*(1–2), 29–37.

Bartulović, D., Steiner, S., Fakleš, D., & Mavrin Jeličić, M. (2023). *Correlations among fatigue indicators, subjective perception of fatigue, and workload settings in flight operations. Aerospace, 10*(10), 856. https://doi.org/10.3390/aerospace10100856.

Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE transactions on vehicular technology, 53(*4), 1052-1068.

Johnston, F. (2025). *FAID: A biomathematical approach to fatigue risk.* Shiftwork Services. https://shiftwork.co.nz/faid-a-bio-mathematical-approach-to-fatigue-risk/

McClung, S. N., & Kang, Z. (2016). Characterization of visual scanning patterns in air traffic control. *Computational Intelligence and Neuroscience*, 2016, Article 8343842. https://doi.org/10.1155/2016/8343842.

Polar Electro. (2026). *Polar Ignite 3: Fitness & wellness watch* [Product documentation]. https://www.polar.com/us-en/ignite3

Roscoe, A. H. (1984). *The Bedford workload scale: A method for measuring the workload of aircraft pilots and aircrew.* Royal Aircraft Establishment, Bedford.

SKYbrary**.** (2026). *Workload (OGHFA BN).* Flight Safety Foundation / EUROCONTROL. https://skybrary.aero/articles/workload-oghfa-bn

Taylor, R. M. (1990). *Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design.* AGARD-CP-478.

Tobii AB. (2026). *Tobii Pro Fusion: Screen-based eye tracker* [Product documentation]. Retrieved from https://www.tobii.com/products/eye-trackers/screen-based/tobii-pro-fusion [tobii.com]

van der Wel, P., & van Steenbergen, H. (2018). *Pupil dilation as an index of effort in cognitive control tasks: A review. Psychonomic Bulletin & Review, 25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y