

EXPERIMENTAL ASSESSMENT OF COGNITIVE LOAD THROUGH EYE TRACKING AND SUBJECTIVE MEASURES: A PRELIMINARY PILOT STUDY

Chiara Forgione^{1*}, Antonio Maria Coruzzolo², Elia Balugani², Francesco Lolli^{2,3}, Rita Gamberini^{2,3}, Riccardo Magistroni⁴

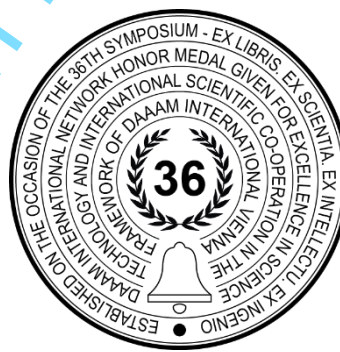
¹ Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Via Giuseppe Campi 287, 41125 – Modena, Italy.

² Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, Via Amendola 2, Padiglione Morselli, 42122 Reggio Emilia, Italy

³ Artificial Intelligence Research and Innovation Center (AIRI), University of Modena and Reggio Emilia, Via Pietro Vivarelli 10, 41125 – Modena, Italy.

⁴ Nephrology Dialysis and Transplant Unit, University Hospital of Modena, Azienda Ospedaliero - Universitaria di Modena, Via del Pozzo 71, 41124 – Modena, Italy.

* Corresponding author



This Publication has to be referred as: Forgione, C[hiara]; Coruzzolo, A[ntonio] M[aria]; Balugani, E[lia]; Lolli, F[rancesco]; Gamberini, R[ita] & Magistroni, R[iccardo] (2026). Experimental Assessment of Cognitive Load Through Eye Tracking and Subjective Measures: a Preliminary Pilot Study, Proceedings of the 37th DAAAM International Symposium, pp.xxxx-xxxx, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-xx-x, ISSN 1726-9679, Vienna, Austria
DOI: 10.2507/37th.daaam.proceedings.xxx

Abstract

Cognitive load is a key factor in human-machine interaction, as it affects an individual's ability to acquire, process, and integrate information. With increasing digitalization, task execution has evolved, requiring users to learn methodologies and perform complex activities, increasing cognitive demands. Thus, monitoring cognitive load and designing human-centred interfaces and tools become essential. Objective measures derived from eye tracking were combined with the NASA Task Load Index to assess perceived workload. Cognitive load was evaluated during tasks with varying levels of complexity, including a Psychomotor Vigilance Task, an N-back task assessing working memory performance, and an assembly-disassembly task performed using bench tools. Task performance was assessed to examine the relationship between cognitive load and behavioural outcomes. Results showed the assembly-disassembly task generated the highest

physiological activation, while participants perceived the N-back tasks as more demanding. Reaction times in the vigilance task were fast and accuracy was high, confirming its lower perceived difficulty. In the N-back, performance decreased as task complexity increased, with more frustration reported. Moreover, execution times in the assembly task improved in the second attempt, suggesting a learning effect. Combining physiological, subjective, and performance data provided a comprehensive understanding of cognitive load, supporting the development of human-centred design solutions.

Keywords: Cognitive Load; Eye Tracking; NASA-TLX; Human-Centred Design; Assembly Task.

1. Introduction

The advent of Industry 4.0 has transformed human-machine interaction through advanced information systems capable of processing large volumes of data [1]. Despite improvements in technological performance, less attention has been devoted to human reliability and cognitive resource management [2]. As task complexity and information demand increase, operators may experience elevated cognitive load, making its assessment essential to support human-centred work environment.

Among emerging approaches, eye-tracking (ET) has emerged as one of the most promising methods for assessing cognitive load, capable of capturing real-time indicators of attention and workload during task execution [3]. Combined with subjective measures such as NASA Task Load Index (NASA-TLX) [4], it enables a more comprehensive view of cognitive demand by integrating perceived and behavioural responses. However, existing studies are often limited to specific task categories or isolated experimental paradigms, with limited investigation of cognitive workload across heterogeneous tasks within a unified framework. Moreover, the combined use of ET and subjective workload measures remains relatively underexplored in comparison to more complex physiological monitoring setups.

In this context, this preliminary pilot study investigates cognitive workload across tasks with different levels of complexity by integrating ET metrics, NASA-TLX, and performance indicators, focusing on psychomotor, working memory, and manual assembly activities. The study addresses the following research questions:

- RQ1: How does cognitive workload vary across psychomotor, working memory, and manual assembly activities using ET, subjective, and performance measures?
- RQ2: How consistent are objective and subjective workload assessments across different task types?

The remainder of this paper is structured as follows. Section 2 presents the related work on cognitive workload assessment methods and industrial applications. Section 3 describes the experimental settings, the sensor used, the procedures, and the methods for data processing. Section 4 presents the main findings from the sensor data, task performance measures, and subjective evaluations. Section 5 discusses the results and provides their interpretation. Finally, Section 6 concludes with implications, limitations, and directions for future research.

2. Related Work

This section provides an overview of the main approaches for cognitive workload assessment, including subjective, physiological, and multimodal methods. The focus is placed on objective and ET-based techniques, which are most relevant to the scope of this study.

2.1. Cognitive Load Assessment Methods

According to Cognitive Load Theory, human performance is influenced by workload, which depends on task complexity, individual characteristics, and the work environment conditions [5]. To assess cognitive load, both subjective and objective measurement approaches have been developed. Traditional approaches are mainly subjective, such as the NASA-TLX [4], widely used to evaluate perceived mental demand and task difficulty across different operational contexts. However, traditional methods are limited by their reliance on self-reporting and their inability to provide real-time monitoring [6]. To overcome these limitations, objective methods based on physiological signals have been developed, allowing the assessment of cognitive load through measurable biological responses [3]. These methods rely on indicators such as heart rate (HR), galvanic skin response (GSR), pupillometry, and electroencephalography (EEG) which provides insights into neural activity associated with cognitive processes [7]. Among physiological approaches, EEG is widely adopted due to its sensitivity to cognitive processes, while ET has emerged as a promising non-invasive alternative for assessing cognitive workload in a continuous and unobtrusive manner [8]. ET enables the analysis of both pupillary responses and eye movement behaviour during task execution. Pupil dilation is commonly associated with cognitive load, although it can also be influenced by external and individual factors, making multimodal integration often necessary [9]. In addition, metrics such as fixations and saccades provide complementary information on visual attention and information processing [10], [11]. Recent studies have increasingly adopted multimodal approaches integrating physiological, behavioural, and subjective measures to improve the robustness of cognitive workload estimation [12], [13]. These systems are particularly relevant in industrial and human-machine interaction contexts, although their adoption is still limited by issues related to sensor intrusiveness, synchronization complexity, and computational cost. Comparative studies of wearable and markerless sensing technologies have also shown that device selection involves a

trade-off between measurement accuracy, cost, and ease of implementation in real settings [14], [15]. Overall, the literature highlights a shift toward integrated multimodal frameworks while emphasizing the need for scalable and application-oriented solutions suitable for real industrial environments [16], [17].

2.2. Industrial Application

Cognitive workload assessment methods have been widely applied across different industrial and experimental contexts, reflecting the diversity of tasks and operational demands in modern human-machine systems. In industrial assembly environments, common sub-tasks include pick-and-place operations [18], [19], blasting [20], rotary drilling [21], which typically involve repetitive actions combined with varying levels of precision and attentional demand. Moreover, assistance systems such as augmented reality are increasingly adopted to support manual assembly, although they may impose additional cognitive load on workers if not designed around their needs [22]. In parallel, human-robot interaction studies frequently investigate teleoperation [23] and robotic manipulation [24], [25], where operators are required to manage continuous monitoring, decision-making, and adaptation to autonomous systems. Beyond applied industrial scenarios, controlled experimental studies have also investigated cognitive workload through paradigms such as cognitive capacity tests [26], [27], Stroop-based tasks [28], and mental workload elicitation tasks [29], which are typically designed to isolate working memory, attention, and cognitive control processes. Overall, these task categories are particularly relevant to the present study, as they encompass both psychomotor and cognitively demanding activities, which are directly comparable to the experimental conditions considered here (manual assembly and computer-based tasks). This alignment allows for a more consistent interpretation of cognitive workload variations across different task types within a unified experimental framework.

3. Materials and Methods

The aim of this preliminary pilot study was to analyse changes in pupil diameter and eye movement across tasks with different complexity levels, comparing these measures with task performance and perceived cognitive load assessed through NASA-TLX.

3.1. Experimental Setting

Seven participants (2 females and 5 males), including students and research staff from the University of Modena and Reggio Emilia, took part in the study. Participants were aged between 22 and 35 years.

ET data were collected using the Tobii Pro Glasses 3, a wearable device capable of recording both temporal and spatial gaze data. To reduce variability in pupillary responses due to environmental factors, all experimental sessions were conducted in the same indoor environment under constant artificial lighting conditions throughout the data collection process. Three of the four tasks were performed on an HP Pavilion Gaming Laptop 15 (15.6-inch display) positioned 50 cm from the participant, using a mouse and keyboard. The fourth task took place at a separate workstation equipped with a hammer, an electric screwdriver, an assembly platform for five components, and 10 wooden dowels. A printed NASA-TLX questionnaire was provided at each workstation for immediate post-task completion.

3.2. Procedure

Before the experiment began, participants were informed about the main objective of the study, the experimental procedures, the task instructions, and how to complete the NASA-TLX questionnaire. Participants then wore the ET, and a brief calibration was performed. Four tasks were administered in the following order:

- Task 1: Assembly and disassembly task. This activity consisted of three sub-tasks: (1) tightening and loosening four screws on the assembly platform using an electric screwdriver; (2) striking a marked point on the platform five times with a hammer; (3) assembling five wooden components using wooden dowels, following symbolic instructions to reproduce a final configuration (Fig. 1). After completion, participants disassembled the structure. The task was performed twice, and execution time was recorded using a stopwatch.

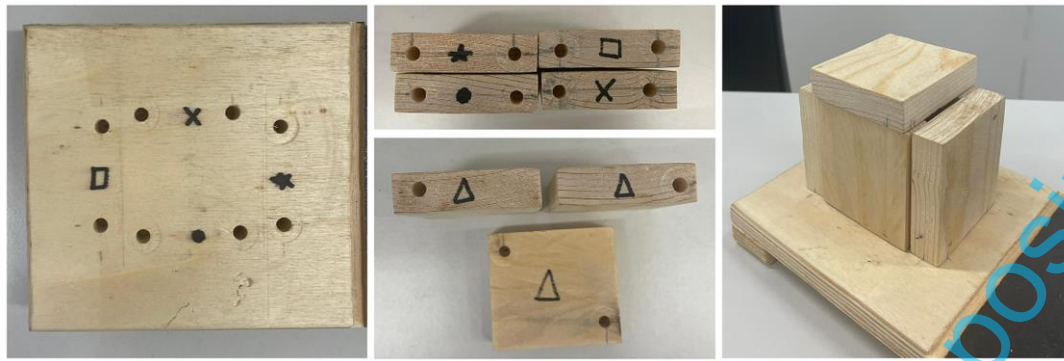


Fig. 1. Assembly/disassembly task workstation: (left) base platform with marked positions; (centre) wooden components with symbolic indicators; (right) final assembled configuration.

- Task 2: Psychomotor Vigilance Task (PVT). An online application [30] was used to measure reaction times through randomly presented visual stimuli (1.5 x 1.5 cm). Participants pressed the spacebar as quickly as possible when a stimulus appeared. The task included 20 trials, with stimulus durations up to 1500 ms and inter-stimulus intervals randomly varying between 1000 and 2000 ms. Reaction times were summarized at the end of the task.
- Task 3: N-back (3x2 grid). A visuospatial N-back task based on the Italian version of the Jaeggi-Buschkuehl dual N-back paradigm [31] was administered via web application. A red square appeared sequentially in a 3x2 grid, and participants responded when its position matched that shown two trials earlier. The task included 20 trials with a 2000 ms stimulus interval, and responses were given via mouse. Accuracy was reported as correct, incorrect, and omitted responses.
- Task 4: N-back (5x5 grid). This task replicated the previous N-back paradigm, with an expanded 5x5 grid and identical procedures and timing parameters.

After completing each task, participants filled out the NASA-TLX questionnaire to assess perceived workload. For subsequent analysis, experimental activities were grouped into two categories according to their operational characteristics. The assembly and disassembly activity (Task 1), which involved manual interaction with physical components and tool use, was classified as a manual assembly task. In contrast, the remaining activities, including the PVT (Task 2) and the visuospatial N-back tasks (Task 3 and Task 4), were performed via computer interface and primarily involved visual search, sustained attention, and working memory processes. These tasks were therefore collectively classified as computer-based tasks.

3.3. Data Processing

Data collected through the Tobii Pro Glasses 3 Controller were first pre-processed to remove invalid samples caused by temporary tracking loss or blinking. Missing but valid segments were interpolated by estimating values based on adjacent samples. Subsequently, a low-pass filter was applied to reduce high-frequency noise and retain only the slower, meaningful fluctuations in pupil size. For each experimental acquisition, the mean pupil diameter, standard deviation, minimum, and maximum values were calculated. Pupil diameter data were analysed to assess autonomic nervous system responses associated with task complexity, as increases in pupil size are commonly linked to greater cognitive effort [32]. Mean pupil dilation values were computed for each task and participant and then aggregated at the group level for statistical analysis. Furthermore, fixations were identified using the Dispersion-Threshold Identification (I-DT) algorithm, which defines fixations as clusters of consecutive points within a specified spatial dispersion. This method provides robust results even in the presence of ET noise [33] and is well-suited for the 50 Hz sampling frequency used in this experiment. For each participant, descriptive and inferential statistics were calculated for all oculomotor parameters and subsequently aggregated to enable comparisons across experimental conditions. Task performance metrics were analysed according to task type: execution time was recorded for manual assembly task, while accuracy, error rates, and success percentages were calculated for computer-based tasks. NASA-TLX responses were processed by calculating the overall workload score (0-100) and analysing individual subscales (mental demand, physical demand, temporal demand, performance, effort, and frustration). Finally, statistical analyses were conducted to compare oculomotor parameters across tasks. This integrated approach provided a comprehensive assessment of the relationship between eye-movement behaviour and perceived cognitive load under different experimental conditions.

4. Results

This section presents the results related to task performance, ET data, and NASA-TLX scores. Performance was evaluated through execution time, accuracy, error rates, and success percentage, while ET analysis focused on pupil

diameter variability and fixations. NASA-TLX scores were analysed to assess perceived workload across the experimental conditions.

4.1. Task Performance

For Task 1, the assembly and disassembly activity, performance was evaluated by measuring execution time (Fig. 2). The mean execution time for the first trial was 3 minutes and 53 seconds, decreasing to 3 minutes and 9 seconds in the second trial. Consistently, the minimum execution time was recorded during the second trial (2 minutes and 52 seconds), whereas the maximum execution time (4 minutes and 29 seconds) was observed in the first trial for the first participant. Overall, the reduction in execution times from the first to the second trial suggests a progressive familiarization with the task procedure across repetitions.

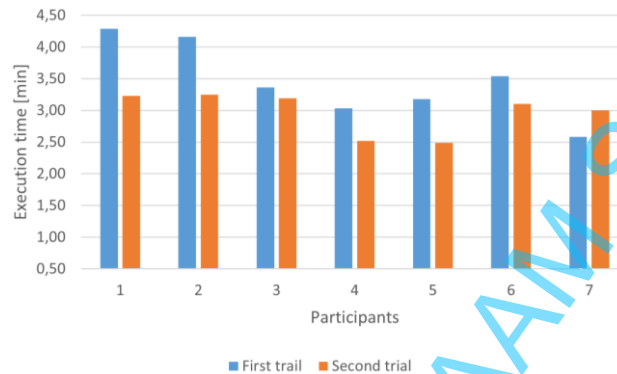


Fig. 2. Execution time (minutes) for each participant during the first and second trial of Task 1.

For computer-based tasks (Task 2, Task 3 and Task 4), performance was evaluated by calculating the number of errors and the corresponding success percentage (Fig. 3).

Task 2 showed the highest mean success rate (98%), consistent with its lower level of complexity. Task 3 yielded a mean success percentage of 81%, whereas Task 4 showed the lowest mean performance (64%), with values ranging from a minimum of 45% to a maximum of 75%. Overall, the progressive decrease in success rates across tasks reflects the increasing cognitive demands, with Task 4 representing the most complex condition among the computer-based tasks.

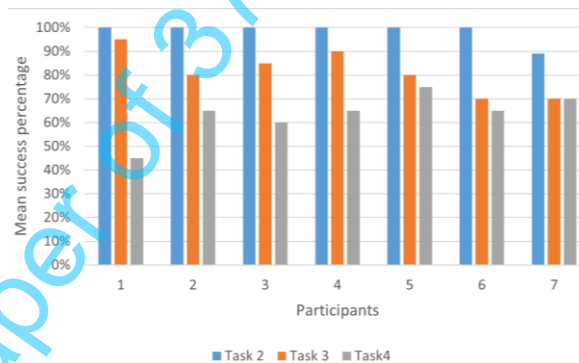


Fig. 3. Mean success percentage for each participant in Task 2 (PVT), Task 3 (N-back 3x2), and Task 4 (N-back 5x5).

4.2. Eye Tracking

Following data preprocessing, pupil dilation values from the left and right eyes of each participant and for each task were recorded and filtered to remove noise and artifacts. The mean pupil diameter between right and left eye and for each acquisition was computed, and the standard deviation of these means was used to quantify intra-subject and intra-task variability in pupil response throughout the experiment. Furthermore, gaze coordinates were used to identify fixations and compute fixation metrics, including total number of fixations, fixation rate, and fixation duration statistics. To support inter-task comparisons, fixation metrics were aggregated at the participant level by grouping tasks into manual assembly activity (Task 1) and computer-based activities (Task 2, Task 3, Task 4). Mean pupil diameter values, reported in Table 2, showed a clear variation across the experimental activities. Larger pupil diameters were consistently observed during the manual assembly task compared to the computer-based cognitive tasks. Mean values during assembly were generally above 5 mm, while computer-based tasks showed lower values (3.1-3.9 mm). This pattern was consistent across

participants, indicating a systematic modulation of pupil response by task characteristics. The higher dilation in manual assembly suggests increased physiological activation linked to tool interaction and visuomotor coordination, whereas computer-based tasks elicited lower pupil diameters, likely reflecting more stable visual engagement with screen-based stimuli. The standard deviation of pupil diameter provided additional insight into intra-subject variability across tasks. Overall, higher variability was observed during the assembly task compared to the computer-based activities. This suggests that pupil responses during manual tasks were more dynamically modulated, likely reflecting alternating phases of visual search, manipulation, and brief attentional disengagement. In contrast, computer-based tasks showed lower pupil variability, suggesting more sustained visual attention and a more stable physiological response during interaction with graphical interfaces.

Fixation analysis revealed systematic differences in visual behaviour between manual assembly and computer-based cognitive tasks, as shown in Table 2. Median fixation duration was generally shorter during the assembly activities and increased during computer-based tasks. Across participants, fixation durations during manual assembly were typically concentrated around 140-180 ms, whereas longer fixations were observed during computer-based tasks, often exceeding 250-400 ms. This pattern suggests that manual assembly requires more dynamic visual exploration and frequent gaze shifts between tools, components, and spatial targets. In contrast, computer-based tasks promoted longer periods of visual stability, likely reflecting sustained attention to discrete screen locations and increased cognitive processing demands.

Differences were also observed in fixation rate across task types. Assembly activities generally showed higher fixation rates, indicating more rapid alternation between fixation and saccadic movements. Conversely, computer-based tasks were associated with lower fixation rates, consistent with the presence of longer fixation events and reduced gaze scanning behaviour. These findings support the interpretation that manual interaction with physical objects induces a more active visuomotor strategy, whereas computer-based tasks favour prolonged attentional engagement with specific visual stimuli.

As shown in Fig. 4, notable inter-individual variability in fixation behaviour emerged across conditions. Although fixation duration generally increased during computer-based tasks, the magnitude of this effect varied between participants. Some showed strong increases, while others displayed more moderate changes, with occasional prolonged fixations during demanding tasks.

Additionally, inferential statistics were applied to assess differences across tasks using a Friedman test for repeated measures. Significant effects were found for pupil dilation, total number of fixations, and mean fixation duration (p -value < 0.001), whereas fixation rate did not show significant differences. Effect size estimation using Kendall's W indicated a strong concordance among participants ($W = 0.796$), suggesting that task-related effects were consistent across individuals.

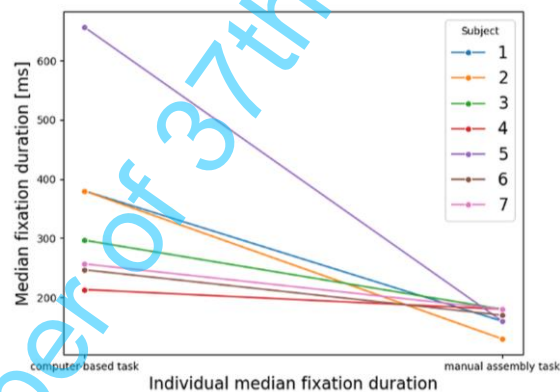


Fig. 4. Individual trends in median fixation duration across manual assembly and computer-based tasks. Lines connect task-type for each participant.

4.3. NASA-TLX

Participants completed the NASA-TLX questionnaire after each task to assess perceived cognitive workload. The overall workload score is derived from six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. For each task, a total NASA-TLX score was calculated, and mean values were compared across participants and tasks.

The results indicate that Task 4 (N-Back 5x5) elicited the highest perceived cognitive workload, with a mean score of 77.3 out of 100. Individual scores ranged from a minimum of 68.3 to a maximum of 94.0. Task 3 (N-Back 3x2) showed slightly lower perceived workload, with a mean score of 61.6 and a wider range of responses (46.7-90.7).

The manual assembly task (Task 1) yielded a moderate workload level, with a mean score of 49.6 varying between 25.7 and 70.0. Finally, Task 2 (PVT) resulted in the lowest perceived workload, with a mean score of 39.8 and scores varying between 20.3 and 82.3.

A detailed overview of the results is provided in Table 1.

ID participant	Task 1	Task 2	Task 3	Task 4
1	25.67	22.67	55.33	71.30
2	50.33	82.30	64.67	70.00
3	70.00	47.67	65.00	76.00
4	28.00	20.30	46.67	89.00
5	51.00	54.00	90.67	94.00
6	62.00	47.67	48.33	68.33
7	60.00	66.00	61.30	72.67
Mean	49.57	39.83	61.60	77.32

Table 1. Total scores of NASA-TLX for each participant.

Inferential analysis confirmed significant differences in perceived workload across tasks (Friedman $X^2 = 11.23$, p -value = 0.0105), suggesting that NASA-TLX scores varied depending on task complexity.

Analysing the specific dimensions of NASA-TLX, interesting insights were observed. In the manual assembly task (Task 1), the highest NASA-TLX scores were observed for performance (61%) and frustration (54.2%), while the remaining dimensions showed intermediate values; physical demand was the lowest (18%). For the PVT (Task 2), performance was again the highest (64.3%), followed by frustration (50%), with the other dimensions contributing less and physical demand remaining minimal. In N-Back 3x2 (Task 3), scores were relatively homogeneous (between 47% and 59%) and mental demand (56%) slightly higher than the other dimensions, while physical demand remained the lowest. Finally, in N-Back 5x5 (Task 4), mental demand was dominant (88%), whereas the other dimensions showed intermediate or low values, with physical demand again minimal (10.7%).

Overall, the results indicate that cognitive related dimensions (mental demand, performance, and effort) were the main drivers of perceived workload across tasks, whereas physical demand consistently played a marginal role.

Task	Mean Pupil Diameter (mm)	Pupil Standard deviation (mm)	Median fixations duration (ms)	Fixation rate (Hz)	Mean NASA-TLX score	Performance
Task 1, first trial	5.23	0.45	162.86	2.08	-	233 s (execution time)
Task 1, second trial	5.18	0.52	168.57	2.05	49.57	189 s (execution time)
Task 2	3.47	0.61	437.14	1.72	39.83	98% (success percentage)
Task 3	3.41	0.48	307.14	1.85	61.60	81% (success percentage)
Task 4	3.37	0.50	297.14	2.03	77.32	64% (success percentage)

Table 2. Multimodal summary of ET, performance, and NASA-TLX measures across experimental tasks.

5. Discussion

The findings are discussed in relation to cognitive load assessment across tasks with different complexity levels, integrating ET metrics and subjective workload measures, with particular attention to fixation behaviour and pupillary responses.

5.1. Manual Assembly Task

The manual assembly task (Task 1) showed the highest mean pupillary diameter, likely due to several interacting factors. First, the task involved multiple visual stimuli and continuous manipulation of physical components, which may have increased both mental workload and cognitive processing demands. Second, the remaining tasks were performed on a laptop in a more uniformly illuminated workstation, potentially contributing to lower pupillary responses. Furthermore, assembling and disassembling objects using hand tools mainly requires visuospatial processing; previous studies suggest that pupil dilation tends to be greater during visuospatial activities compared to verbal or decision-making tasks [34].

At the same time, mean fixation duration was lower during assembly activities, while fixation rates were higher, indicating more frequent alternation between fixations and saccadic movements. This pattern suggests a more dynamic

visuomotor exploration strategy when interacting with physical objects, consistent with the increased pupillary response highlighted by the pupillometry analysis.

Performance data, expressed in terms of task completion time, showed a reduction in the second trial, suggesting the presence of a learning effect, although this improvement was not clearly reflected in the ET indicators. Subjective workload obtained through NASA-TLX scores indicated lower perceived workload only for the PVT task. Overall, these preliminary findings suggest distinct patterns of cognitive load across task types, which are supported by the inferential statistics on the multimodal parameters.

5.2. Computer-based tasks

Computer-based tasks (Task 2, Task 3, and Task 4) were characterized by lower mean pupil diameters compared to the manual assembly task. This may be related to the need for rapid responses to visual stimuli, which likely increased attentional demands while promoting more stable visual engagement. This interpretation is supported by the observation of longer mean fixation durations and lower fixation rates. Reduced fixation rates indicate fewer gaze shifts and decreased visual scanning behaviour, suggesting sustained attentional focus on specific on-screen elements.

Performance outcomes, measured as the percentage of correct responses, were highest for the PVT task (Task 2) and progressively decreased in the N-Back task (Task 3 and Task 4). This trend reflects the increasing cognitive complexity of the tasks and is consistent with subjective workload ratings obtained through the NASA-TLX, which showed increasing perceived workload from Task 2 to Task 4.

These results are supported by inferential statistical analysis. The Friedman test confirmed significant differences across tasks for both physiological and subjective measures, while no significant effect was found for fixation rate, suggesting that visual sampling frequency remained consistent across conditions. It should be emphasised, however, that these inferential results were obtained from a pilot sample of seven participants ($N = 7$). Although Kendall's W indicated a strong concordance among participants ($W = 0.796$), suggesting that task-related effects were consistent at the individual level, this agreement refers only to the present cohort, which was relatively homogeneous in age (22–35 years) and composed of university students and research staff. With such a limited sample, the statistical power of the Friedman test is reduced and effect size estimates may be unstable. The present findings should therefore be regarded as baseline evidence, and expanded cohort testing, including industrial operators with different ages, levels of experience and task expertise, is required to generalise them to broader industrial populations.

The observed discrepancy between physiological and subjective measures may reflect their different sensitivities to cognitive load components in computer-based tasks, with pupil dilation being more closely related to immediate visuomotor and attentional demands, whereas NASA-TLX reflects consciously perceived cognitive effort [35].

6. Conclusions

Cognitive load is a key factor in many applied domains, particularly in contexts where human-machine interaction is central and high precision and operational efficiency are required. Understanding and managing cognitive load is essential for improving operator well-being, optimizing performance, and reducing the risk of errors, especially in environments with high informational demands.

Previous studies have highlighted that relying on a single type of measurement can present limitations, emphasizing the need for a multimodal analysis that combines both objective and subjective measures. In this context, ET emerges as a promising methodology for real-time cognitive load assessment, as it allows monitoring of ocular parameters such as pupil diameter and fixations. The correlation of these parameters with cognitive effort provides reliable and non-intrusive indicators.

In this preliminary pilot study, ET data were used to analyse cognitive load during specific tasks. Results showed that the manual assembly task elicited larger pupil diameters and higher fixation rates, reflecting dynamic visuomotor engagement, whereas computer-based tasks were associated with longer fixation durations and lower fixation rates, suggesting sustained attentional focus on screen-based stimuli. These preliminary findings are supported by inferential statistics, which confirmed significant differences across tasks for pupil dilation and fixation metrics. These results suggest that ocular behaviour may be associated with cognitive effort, highlighting the potential of ET as a complementary assessment tool.

The presented preliminary findings have potential implications for industrial settings, where ET could be used to monitor operator cognitive load in real time, support adaptive human-machine interaction systems, and human-centred task and workstation design.

However, several limitations should be acknowledged. The small sample size ($N = 7$), due to the preliminary pilot nature of the study, limits the generalisability of the results. Although the Friedman test revealed significant task-related effects and Kendall's W indicated strong participant concordance ($W = 0.796$), these findings refer to a relatively homogeneous group of university students and research staff aged 22–35 years and should therefore be considered exploratory baseline evidence. Task modality may also introduce confounding factors related to motor and environmental differences between manual and computer-based conditions. Additionally, other physiological and contextual factors that can influence cognitive effort and ocular parameters were not monitored, and these could have affected the observed patterns.

Future studies should extend the experimental protocol to larger and more heterogeneous cohorts, including industrial operators with different ages, levels of experience and task expertise, in order to validate these baseline findings and generalise them to broader industrial populations. The integration of additional physiological and behavioural measures with ET data could further enhance the robustness of the assessment, while applying ET in real work environments and more complex scenarios could provide further insights into its effectiveness for continuous cognitive load monitoring. Overall, the study suggests that ET represents a promising and complementary approach for assessing cognitive load, supporting the development of more effective and adaptive monitoring systems.

7. Acknowledgments

This work was funded by the European Regional Development Fund (ERDF) through the Regional Operational Programme of Emilia-Romagna (Italy) under Project “LUME PlannER” (CUP: E82I15002660002, POR-FESR 2014-2020).

8. References

- [1] Chen, F.; Ruiz, N.; Choi, E.; Epps, J.; Khawaja, M. A.; Taib, R.; Yin, B. & Wang, Y. (2012). Multimodal behavior and interaction as indicators of cognitive load, *ACM Transactions on Interactive Intelligent Systems*, Vol. 2, No. 4, pp. 1-36, DOI: 10.1145/2395123.2395127
- [2] Markova, P.; Prajova, V.; Homokyova, M. & Horvathova, M. (2019). Human factor in Industry 4.0 in point of view ergonomics in Slovak Republic, *Proceedings of the 30th DAAAM International Symposium*, pp. 284-289, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-22-8, ISSN 1726-9679, Vienna, Austria, DOI: 10.2507/30th.daaam.proceedings.037
- [3] Hessels, R. S.; van Doorn, A. J.; Benjamins, J. S.; Holleman, G. A. & Hooge, I. T. C. (2020). Task-related gaze control in human crowd navigation, *Attention, Perception, & Psychophysics*, Vol. 82, No. 5, pp. 2482-2501, DOI: 10.3758/s13414-019-01952-9
- [4] Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, In: *Human Mental Workload*, *Advances in Psychology*, Vol. 52, Hancock, P. A. & Meshkati, N. (Eds.), pp. 139-183, North-Holland, Amsterdam, DOI: 10.1016/S0166-4115(08)62386-9
- [5] Jex, S. M. (1998). *Stress and job performance: Theory, research, and implications for managerial practice*, Sage Publications, Thousand Oaks
- [6] Gopher, D. & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures?, *Human Factors*, Vol. 26, No. 5, pp. 519-532, DOI: 10.1177/001872088402600504
- [7] Gkintoni, E. & Halkiopoulou, C. (2025). Mapping EEG metrics to human affective and cognitive models: An interdisciplinary scoping review from a cognitive neuroscience perspective, *Biomimetics*, Vol. 10, No. 11, 730, DOI: 10.3390/biomimetics10110730
- [8] Katona, J. (2022). Measuring cognition load using eye-tracking parameters based on algorithm description tools, *Sensors*, Vol. 22, No. 3, 912, DOI: 10.3390/s22030912
- [9] Bhatia, N.; Sen, D. & Pathak, A. V. (2015). Visual behavior analysis of human performance in precision tasks, In: *Engineering Psychology and Cognitive Ergonomics*, *Lecture Notes in Computer Science*, Vol. 9174, Harris, D. (Ed.), pp. 95-106, Springer, ISBN 978-3-319-20372-0, Cham, DOI: 10.1007/978-3-319-20373-7_10
- [10] Hessels, R. S.; Niehorster, D. C.; Nyström, M.; Andersson, R. & Hooge, I. T. C. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers, *Royal Society Open Science*, Vol. 5, No. 8, 180502, DOI: 10.1098/rsos.180502
- [11] Ahmadi, N.; Sasangohar, F.; Yang, J.; Yu, D.; Danesh, V.; Klahn, S. & Masud, F. (2024). Quantifying workload and stress in intensive care unit nurses: Preliminary evaluation using continuous eye-tracking, *Human Factors*, Vol. 66, No. 3, pp. 714-728, DOI: 10.1177/00187208221085335
- [12] Vanneste, P.; Raes, A.; Morton, J.; Bombeke, K.; Van Acker, B. B.; Larmuseau, C.; Depaepe, F. & Van den Noortgate, W. (2021). Towards measuring cognitive load through multimodal physiological data, *Cognition, Technology & Work*, Vol. 23, No. 3, pp. 567-585, DOI: 10.1007/s10111-020-00641-0
- [13] Zhu, L. & Lv, J. (2023). Review of studies on user research based on EEG and eye tracking, *Applied Sciences*, Vol. 13, No. 11, 6502, DOI: 10.3390/app13116502
- [14] Coruzzolo, A. M.; Forgione, C.; Balugani, E.; Gamberini, R. & Roversi, A. (2025). Comparison of different motion capture technologies for gait analysis and suggestions for implementation in tourism applications, *Proceedings of the 36th DAAAM International Symposium*, pp. 122-130, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-47-1, ISSN 1726-9679, Vienna, Austria, DOI: 10.2507/36th.daaam.proceedings.016
- [15] Lolli, F.; Coruzzolo, A. M.; Forgione, C.; Peron, M. & Sgarbossa, F. (2024). Auto-AzKNIOSH: An automatic NIOSH evaluation with Azure Kinect coupled with task recognition, *Ergonomics*, pp. 1-17, DOI: 10.1080/00140139.2024.2433027
- [16] Ioniță, A.-R.; Anghel, D.-C. & Boudouh, T. (2025). Mind, machine, and meaning: Cognitive ergonomics and adaptive interfaces in the age of Industry 5.0, *Applied Sciences*, Vol. 15, No. 14, 7703, DOI: 10.3390/app15147703

- [17] Ricci, A.; Ronca, V.; Capotorto, R.; Giorgi, A.; Vozzi, A.; Germano, D.; Borghini, G.; Di Flumeri, G.; Babiloni, F. & Aricò, P. (2025). Understanding the unexplored: A review on the gap in human factors characterization for Industry 5.0, *Applied Sciences*, Vol. 15, No. 4, 1822, DOI: 10.3390/app15041822
- [18] Arif, A.; Zakeri, Z.; Omurtag, A.; Breedon, P. & Khalid, A. (2024). Neurophysiological approach for psychological safety: Enhancing mental health in human-robot collaboration in smart manufacturing setups using neuroimaging, *Information*, Vol. 15, No. 10, 640, DOI: 10.3390/info15100640
- [19] Zakeri, Z.; Arif, A.; Omurtag, A.; Breedon, P. & Khalid, A. (2023). Multimodal assessment of cognitive workload using neural, subjective and behavioural measures in smart factory settings, *Sensors*, Vol. 23, No. 21, 8926, DOI: 10.3390/s23218926
- [20] John, A. R.; Singh, A. K.; Gramann, K.; Liu, D. & Lin, C.-T. (2024). Prediction of cognitive conflict during unexpected robot behavior under different mental workload conditions in a physical human-robot collaboration, *Journal of Neural Engineering*, Vol. 21, No. 2, 026010, DOI: 10.1088/1741-2552/ad2494
- [21] Shao, J.; Zhan, Y.; Zhu, H.; Zhang, M.; Qin, L.; Tian, S. & Qi, H. (2024). The effects of representation of industrial icons on visual search performance, *Displays*, Vol. 82, 102661, DOI: 10.1016/j.displa.2024.102661
- [22] Konkol, K.; Kießling, N.; Engeländer, N. & Schwarz, A. (2024). Development of a framework for adaptable augmented reality assistance systems for manual assembly tasks, *Proceedings of the 35th DAAAM International Symposium*, pp. 230-237, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-44-0, ISSN 1726-9679, Vienna, Austria, DOI: 10.2507/35th.daaam.proceedings.031
- [23] Lee, H.; Han, J. & Yang, G.-H. (2024). Development of variable scaling teleoperation framework for improving teleoperation performance, *International Journal of Control, Automation and Systems*, Vol. 22, No. 3, pp. 936-945, DOI: 10.1007/s12555-022-1099-z
- [24] Bussolan, A.; Baraldo, S.; Avram, O.; Urcola, P.; Montesano, L.; Gambardella, L. M. & Valente, A. (2025). MultiPhysio-HRC: A multimodal physiological signals dataset for industrial human-robot collaboration, *Robotics*, Vol. 14, No. 12, 184, DOI: 10.3390/robotics14120184
- [25] Teng, J.; Cho, S. & Lee, S. (2025). Human-robot interaction: Integrating BCI-controlled virtual robotic arm with manual control in VR environment, *Computer Animation and Virtual Worlds*, Vol. 36, No. 5, e70072, DOI: 10.1002/cav.70072
- [26] Eesee, A. K. & Ruppert, T. (2026). Evaluating the reliability of physiological and kinematic features for tracking cognitive load changes during repetitive task execution, *Results in Engineering*, Vol. 29, 108657, DOI: 10.1016/j.rineng.2025.108657
- [27] Luzzani, G.; Pogliano, M.; Buraioli, I.; Colavincenzo, M.; Martorana, S.; Guglieri, G. & Demarchi, D. (2025). ECG, respiration, fNIRS, and eye tracking for stress and mental workload monitoring in human-machine interaction, *IEEE Access*, Vol. 13, pp. 122726-122741, DOI: 10.1109/ACCESS.2025.3588384
- [28] Trapero, J. A.; Thinnis, D.; Wagner, E. & Strauss, D. J. (2026). Haptic vest-attention assistance for outside field-of-view guidance and enhanced human-robot interaction, *IEEE Transactions on Industrial Informatics*, Vol. 22, No. 3, pp. 2099-2107, DOI: 10.1109/TII.2025.3635535
- [29] Niu, H.; Zhai, Y.; Huang, Y.; Wang, X. & Wang, X. (2022). Investigating the short-term cognitive abilities under local strong thermal radiation through EEG measurement, *Building and Environment*, Vol. 224, 109567, DOI: 10.1016/j.buildenv.2022.109567
- [30] Milanese, A. & Lavermicocca, V. (n.d.). Training cognitivo. Attenzione e funzioni esecutive, Available from: <https://www.trainingcognitivo.com/gamecenter/gamecenter-funzioni-esecutive/> Accessed: 2025-03-20
- [31] Jaeggi, S. M.; Buschkuhl, M.; Perrig, W. J. & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure, *Memory*, Vol. 18, No. 4, pp. 394-412, DOI: 10.1080/09658211003702171
- [32] Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources, *Psychological Bulletin*, Vol. 91, No. 2, pp. 276-292, DOI: 10.1037/0033-2909.91.2.276
- [33] Salvucci, D. D. & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols, *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*, pp. 71-78, Association for Computing Machinery, ISBN 1-58113-280-8, New York, DOI: 10.1145/355017.355028
- [34] Jyotsna, C.; Amudha, J.; Ram, A. & Nollo, G. (2023). IntelEye: An intelligent tool for the detection of stressful state based on eye gaze data while watching video, *Procedia Computer Science*, Vol. 218, pp. 1270-1279, DOI: 10.1016/j.procs.2023.01.105
- [35] La Bara, L. M. A.; Meloni, L.; Giusino, D. & Pietrantoni, L. (2021). Assessment methods of usability and cognitive workload of rehabilitative exoskeletons: A systematic review, *Applied Sciences*, Vol. 11, No. 15, 7146, DOI: 10.3390/app11157146