Measuring Mental Demand Using Pupil Dilation in an Augmented Reality Environment

Siddarth Mohanty[1], Jung Hyup Kim[1], Sara Mostowfi[1],
Fang Wang[1], Danielle Oprean[1], Kangwon Seo[1], and Yi Wang[1]

Abstract

This study investigates the method for measuring cognitive workload in augmented reality-based biomechanics lectures by analyzing pupil dilation. Using Dikablis Glasses 3 and Microsoft HoloLens, we recorded physiological and subjective data across learning and problem-solving phases. Pupil dilation was normalized and segmented, enabling a comparison of cognitive demands between phases. The results indicated significant correlations between pupil dilation and NASA TLX cognitive demand, particularly in lectures that primarily involved procedural knowledge. These findings suggest that instructional design and content complexity have a significant impact on cognitive load, providing valuable insights for optimizing AR-based learning environments to support cognitive efficiency and student engagement.

Keywords

augmented reality, pupil dilation, mental demand

Introduction

Augmented Reality (AR) is transforming educational experiences by providing immersive and interactive learning environments. Although the potential of AR to significantly boost engagement and improve learning outcomes has been investigated, the cognitive demands posed by this technology are still surrounded by uncertainty. Further research is still needed to understand these cognitive demands and explore effective strategies to alleviate them during AR-based learning experiences. Hence, this current study investigates how to measure cognitive load in AR-based biomechanics lectures, focusing on pupil dilation as an indicator of mental demand.

Prior research (Kim et al., 2024) has predominantly examined learning outcomes and user experience, often overlooking how cognitive effort fluctuates between learning and problem-solving phases. In our AR learning modules (as illustrated in Figure 1), students are immersed in two distinct cognitive states: the expedition of acquiring new knowledge and the dynamic problem-solving process. The learning phase requires intense focus and engagement, as students wrestle with novel concepts. On the other hand, the problem-solving stage needs strong thinking skills to analyze situations and make thoughtful decisions to address challenges. Students frequently encounter various obstacles that compel them to transform their theoretical knowledge into practical solutions. These situations not only require them to think critically and creatively but also to demonstrate a comprehensive understanding of their academic learning. By engaging with real-world applications, they can showcase their diverse intellectual skills, including analytical reasoning, problem-solving, and effective communication, thereby illustrating the breadth of their educational experience. To explore and quantify the differences in mental workload across these distinct phases, we turn to eye-tracking technology alongside the NASA Task Load Index (NASA TLX). We propose a compelling hypothesis: with its intense demands of information processing and the establishment of conceptual understanding, the learning phase will reveal a stronger correlation with mental demand than the problem-solving phase. This research opens doors to invaluable insights for optimizing AR-based instructional design. By carefully balancing cognitive load with effective learning strategies, we

[1]University of Missouri, Columbia, USA

Corresponding Author:
Jung Hyup Kim, Department of Industrial & Systems Engineering,
University of Missouri, E3437F Thomas & Nell Lafferre Hall, Columbia,
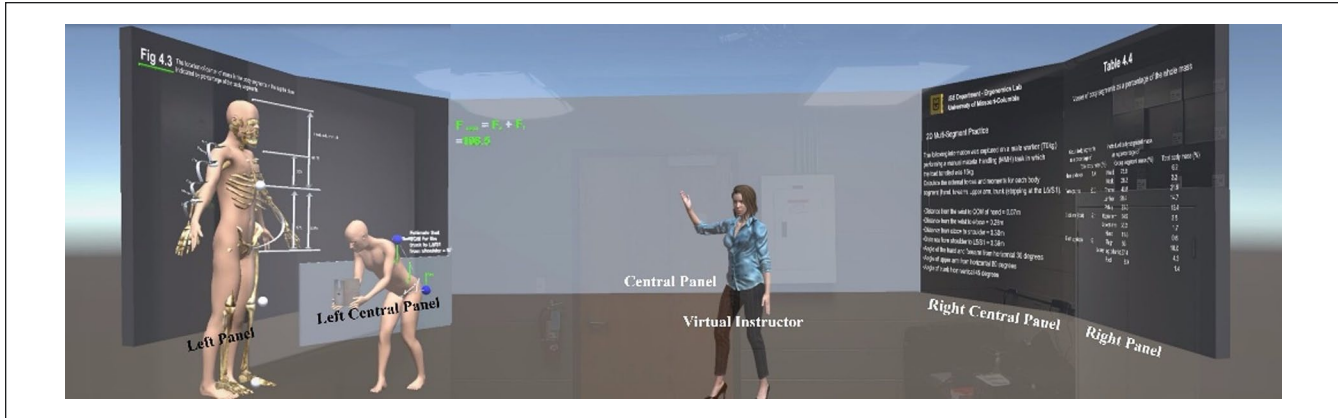Mo 65211, USA.
Email: kijung@missouri.edu

Figure 1. 3D virtual contents used in AR learning Environment (Mohanty et al., 2024).

aim to enhance educational outcomes, ensuring that students are as enriching and transformative as it is engaging.

Background

While AR learning environments have been extensively studied for their impact on knowledge retention and learner motivation, their cognitive demands remain less explored. AR's ability to overlay digital information onto the physical world introduces unique benefits and challenges to cognitive processing. According to Cognitive Load Theory (Sweller, 2011), learning effectiveness is determined by the interplay of three types of loads: intrinsic (task complexity), extraneous (irrelevant or poorly designed information), and germane (mental effort invested in learning). AR has the potential to reduce extraneous load by providing intuitive, multimodal, and context-rich instructional cues, yet it can also increase intrinsic load due to the novelty and interactivity of the medium, as well as the requirement for divided visual and attentional resources (Buchner et al., 2022). This makes an accurate assessment of mental workload in AR critical for optimizing learning environments and reducing the risk of cognitive overload.

Recent advancements in physiological measurement techniques, particularly eye-tracking, have enabled researchers to monitor learners' mental states during interactive tasks in real time. Among these techniques, pupillometry the measurement of pupil size has proven to be a robust and non-invasive indicator of mental effort (Gorin et al., 2024; Othman & Romli, 2016; Yang & Kim, 2019). Pupil dilation has been shown to correlate with the activation of the locus coeruleus–norepinephrine (LC-NE) system, which governs attentional control and cognitive arousal (Kahneman & Beatty, 1966). Pupillary responses have also been linked to task engagement, mental fatigue, and attentional switching, making them useful for assessing dynamic cognitive workload during complex activities (Hopstaken et al., 2015; Kim & Yang, 2020; Nazareth & Kim, 2020). When used alongside subjective assessments like the NASA-TLX, pupillometry can offer a

dual-layered view of workload, capturing both perceived effort and physiological strain (Xie & Salvendy, 2000).

Within AR learning, students typically engage in two distinct cognitive phases: a learning phase, where they absorb and make sense of new information, and a problem-solving phase, where they apply that knowledge in a task-specific context. The learning phase often involves conceptual integration and schema building, which can impose significant intrinsic and germane load. Conversely, the problem-solving phase may involve working memory overload, error correction, and adaptive reasoning, all of which can increase cognitive demands. While some studies have suggested that pupil dilation is more pronounced during problem-solving (Marshall, 2002), the mental workload imposed by conceptual learning, especially in immersive AR environments remains underexamined. A key research gap persists in directly comparing cognitive load across these phases using real-time physiological indicators in combination with validated subjective instruments.

To tackle this gap, we embraced a comprehensive framework for categorizing AR instructional modules into two distinct types: declarative and procedural knowledge. Declarative knowledge encompasses information, descriptive insights, and conceptual understanding, imagine grasping the intricate principles of biomechanics or pinpointing various anatomical structures in vivid detail. In contrast, procedural knowledge is rooted in action; it involves the dynamic sequences of tasks and the cognition necessary to execute them to think of conducting ergonomic assessments. These two forms of knowledge are recognized for engaging different cognitive mechanisms, resulting in varying levels of cognitive load that depend heavily on both the nature of the task at hand and the surrounding context. By aligning eye-tracking data with our classification of these modules, this study paves the way for an intricate examination of how the type of knowledge and the cognitive phase, whether learning or problem-solving interact to shape mental workload within AR environments. The insights gained promise practical applications for refining AR instructional design,

Figure 2. Participant equipped with Dikablis eye tracker and Microsoft HoloLens undergoing lectures.

ensuring that it is finely tuned to meet learners' needs and optimizing the delivery of content to enhance understanding and cognitive efficiency.

Methodology

Experimental Setup

This study was conducted using two structured AR lectures focused on biomechanics and ergonomics, delivered through the Microsoft HoloLens 2 headset (Guo & Kim, 2021; Yu et al., 2023). The first lecture unfolded across seven crafted modules, each designed to immerse students in the essential concepts of biomechanics and the principles of ergonomic practices. Building upon this solid foundation, the second lecture expanded into eight engaging modules that presented dynamic problem-solving tasks, challenging participants to apply the knowledge they had just acquired.

To obtain an objective measurement of cognitive workload, we incorporated the Dikablis Glasses 3 eye tracker into our HoloLens setup, allowing for the capture of real-time pupil dilation data, as illustrated in Figure 2. A group of 27 participants, all students enrolled in an industrial engineering program at the University of Missouri, participated in this study. Each participant embraced both lectures, with a carefully mandated 24-hour rest interval between sessions, ensuring that the effects of cognitive fatigue were minimized and that they could approach each lecture with fresh minds and renewed focus.

In each lecture, participants engaged with immersive AR content directly from the designated locations, engaging in an interactive AR learning experience designed to facilitate conceptual understanding. Following these AR modules, they tackled concept-specific questions on the laptop to assess comprehension and apply learned concepts in a problem-solving context. To create a distinct and seamless transition from the realm of learning to that of application, we designed the sessions to fluidly shift from dynamic AR presentations to thoughtful, related tasks.

Throughout both phases (learning and solving), the eye-tracking system closely monitored pupil dilation, unveiling invaluable insights into their cognitive engagement. At the conclusion of each lecture, participants reflected on their experiences using the NASA-TLX, filing in on the mental demand subscale to create a synergy with our pupil dilation data.

Data Analysis

Pupil area data gathered using the Dikablis Glasses 3 eye tracker was analysed to explore cognitive workload during AR learning sessions. Recognizing the inherent variability in pupil size among individuals, which typically ranges from 800 to 2,500 square millimetres, we employed Equation 1 to normalize all pupil data between 0 and 1. The original pupil size is called P_i (i=1. . .n) and $Pnorm_i$ is the normalized pupil data for P_i.
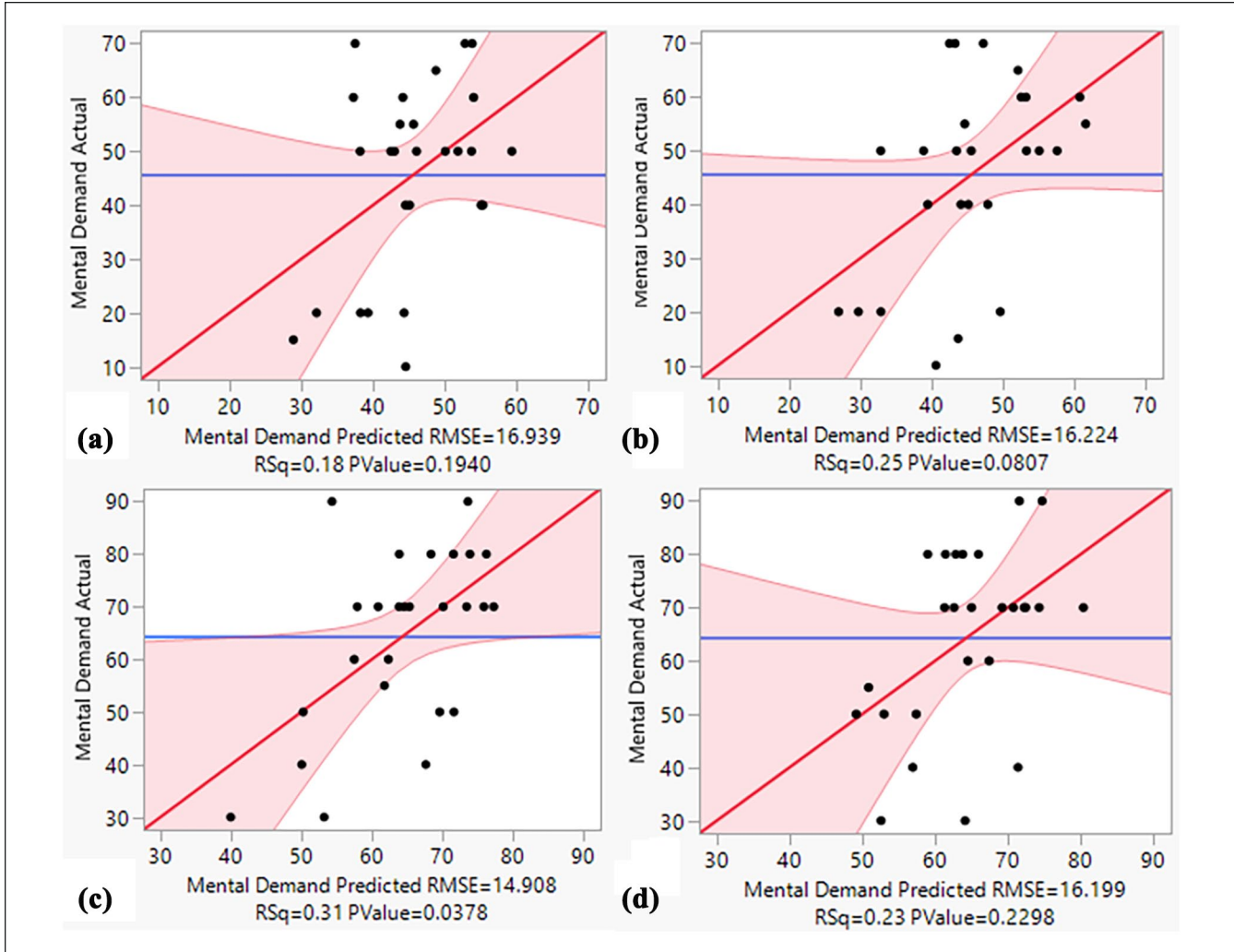
Figure 3. Scatter plots comparing actual NASA TLX Mental Demand values against predicted Mental Demand.

$$\text{Pnorm}_i = \frac{P_i - \min(P)}{\max(P) - \min(P)} \qquad (1)$$

The data from each learning module was thoughtfully segmented into three distinct phases: baseline (B), learning (L), and problem-solving (S). The baseline phase was characterized by a serene three-second idle period preceding the introduction of AR content, serving as a reference point for participants' resting pupil sizes. The learning phase kicked off as participants fully engaged with the AR instructional scene, powered by the Microsoft HoloLens. Subsequently, the problem-solving phase commenced when they shifted gears to tackle a quiz question on the laptop.

Precise phase transitions were determined using time-stamps from Microsoft HoloLens logs, which tracked scene engagement and user navigation across the dynamic AR modules. From the normalized pupil data, we computed the absolute differences between the baseline and learning phases (B–L) and between the baseline and solving phases (B–S). These critical values were extracted for each module

and labelled B-L-1 through B-L-7 for the seven modules in Lecture 1, and B-S-1 through B-S-8 for the eight modules in Lecture 2. To assess the relationship between these physiological indicators and perceived mental effort, we conducted a regression analysis on the B–L and B–S values against the mental demand subscale scores collected from the NASA-TLX, which participants completed after each lecture. This analytical approach provided an avenue for exploring how shifts in pupil area—interpreted as indicators of mental workload—correspond with participants' subjective evaluations of task difficulty. By connecting eye-tracking data to cognitive phases (learning versus solving), this study unveils an understanding of how mental demand fluctuates within AR-driven educational environments.

Results

Figure 3 presents the results of the Fit model analysis, which explores the relationship between the predicted mental demand derived from regression models utilizing normalized

pupil dilation differences (B–L and B–S) and self-reported NASA-TLX Mental Demand scores during both the learning and problem-solving phases of two AR-based lectures. The term "predicted" is employed to denote the modeled regression estimates of perceived mental demand based on the pupil dilation data. These models assess the extent to which variations in pupil size correspond to variations in subjective reports of cognitive workload. In Lecture 1, the learning phase (illustrated in Figure 3a) displayed no correlation ($R^2 = 0.18$, $p = .1940$). The problem-solving phase (Figure 2b) revealed a slightly stronger, still non-significant correlation ($R^2 = 0.25$, $p = .0807$). This might hint at a subtle increase in cognitive engagement but not enough to signify a profound connection in mental workload.

In contrast, Lecture 2 offered a more dynamic picture. The learning phase (depicted in Figure 3c) demonstrated a statistically significant correlation ($R^2 = .31$, $p = .0378$, RMSE $= 14.908$). In this phase of Lecture 2, pupil dilation demonstrated a statistically significant relationship with reported cognitive effort during the learning phase, suggesting its potential as a physiological indicator of mental workload. The problem-solving phase of Lecture 2 (illustrated in Figure 3d) presented a moderate yet non-significant correlation ($R^2 = .23$, $p = .2298$).

These findings compellingly illustrate the variable nature of cognitive workload in AR contexts, showing clear distinctions based on task phases and types of knowledge. Notably, the learning material in Lecture 2 elicited a more consistent connection between pupil dilation signals of mental demand, underscoring the interplay between content type and cognitive engagement. These results advocate for applying pupillometry as a sensitive, phase-dependent measure of mental workload, emphasizing the critical need to develop tailored AR instructional designs that adeptly navigate varying cognitive loads across diverse content types.

Discussion

This study offers valuable insights into measuring cognitive load during various phases of AR-based learning environments. By analysing normalized pupil dilation data, the research highlights how pupil responses can indicate the mental effort students face at different stages of their educational journey within AR settings. The findings underscore the importance of understanding these physiological markers to enhance the design and effectiveness of AR learning experiences. Contrary to earlier beliefs that problem-solving tasks might require greater mental effort, our data suggest a notable insight: a more pronounced relationship between pupil dilation and mental demand during the learning phase, particularly in the intensity of Lecture 2. This supports our idea that learning new concepts, particularly those involving the understanding and integration of complex ideas, shows a stronger correlation between pupil dilation and mental workload than problem-solving tasks.

Within Lecture 2, we observed elevated correlations between pupil dilation and mental demand during learning ($R^2 = .31$, $p = .0378$) compared to the problem-solving phase ($R^2 = .23$, $p = .2298$). This disparity suggests that the intricate nature or challenging complexity of the content presented in Lecture 2 likely intensified cognitive processing demands. In contrast, Lecture 1 exhibited no correlation throughout both phases. These variations highlight the critical importance of thoughtful content design and sequencing in shaping the mental workload experienced by students within AR contexts.

The absence of a statistically significant correlation such as the solving phase of Lecture 1 could be attributed to various factors, including individual differences in cognitive strategies or potential environmental distractions. Furthermore, while the NASA-TLX provides a validated subjective measure of workload, there are instances where it may not align precisely with physiological reactions, particularly when participants' self-perceptions fluctuate across different tasks or individuals.

Integrating subjective and objective metrics fosters a richer and better understanding of cognitive demand. The alignment between pupil dilation and perceived mental effort in specific contexts highlights the promise of pupillometry as a tool for real-time cognitive monitoring. These insights call for a deliberate approach to AR instructional design, one that prioritizes not just engagement but also cognitive efficiency by judiciously adjusting content density, pacing, and interactivity based on the mental workload thresholds of learners.

In this study, AR modules in Lecture 2 focused on procedural knowledge, guiding students through step-by-step applications such as posture analysis, while Lecture 1 emphasized declarative knowledge, including conceptual principles of biomechanics and anatomy. Procedural content often requires mental simulation during learning, leading to higher intrinsic cognitive load as learners attempt to internalize sequences. In contrast, declarative knowledge may appear less demanding initially, when learners must recall and apply abstract concepts. This distinction reflects the Cognitive Theory of Multimedia Learning and aligns with phases of the skill acquisition model, where early procedural learning activates different cognitive resources than conceptual problem-solving (Sweller et al., 1998). These findings can also be contextualized using the skill acquisition framework (Fitts & Posner, 1967), where learners transition from the cognitive stage (requiring significant mental effort) to associative and autonomous stages. During the initial stages of learning, particularly when grasping procedural tasks, pupils often experience significant dilation. This phenomenon reflects a heightened cognitive engagement as learners grapple with understanding complex sequences and mechanisms that demand intense mental effort. This aligns with the associative stage of learning, where the cognitive load is dominated by the need to recall information and apply it effectively in various contexts.

In future inquiries, it would be appropriate to explore the long-term effects of learning by analyzing how students respond over extended periods. Incorporating additional physiological indicators, such as heart rate variability or electroencephalography data, could provide deeper insights into the intricacies of cognitive load. Furthermore, testing AR contents across various disciplines would help to broaden the applicability of these findings. Optimizing the transitions between learning and problem-solving phases is crucial, as is offering adaptive feedback tailored to individual needs, both of which could significantly enhance the management of mental workload and boost overall learning effectiveness.

Conclusion

This study investigates the cognitive workload encountered during learning and problem-solving tasks within AR-enhanced lectures on biomechanics and ergonomics. We uncovered compelling insights by comparing subjective assessments through the mental demand of NASA-TLX and objective measures via pupil dilation.

The findings illuminated a compelling link between pupil dilation and perceived mental demand during the learning phase, particularly in Lecture 2, which concentrated on procedural knowledge. This indicates that the cognitive workload experienced in AR environments varies significantly across different types of tasks and knowledge domains.

These outcomes highlight the profound potential of using real-time physiological indicators, such as pupillometry, to gauge cognitive engagement and enhance the design of AR instructional experiences. Educators and developers can craft more adaptable and cognitively efficient AR learning environments by pinpointing the moments and areas where cognitive load surges.

Future research should focus on utilizing multimodal assessments to measure cognitive workload, explore individual differences in cognitive responses, and investigate how these findings can be applied across various fields and learner demographics. This exploration is crucial for validating and refining the role of eye-tracking technology in the rapidly evolving landscape of AR-based education.

Limitations

This study has several limitations. Firstly, while pupil dilation is widely recognized as a physiological indicator of mental workload, it is also influenced by various factors, including emotional states and fatigue. Although we took steps to mitigate these influences, it was impossible to eliminate them. Additionally, our sample was limited to students from a single discipline—Industrial Engineering—which may restrict the generalizability of our findings to broader populations. For future research, it would be beneficial to incorporate additional measurement techniques, such as EEG and heart rate variability, to more effectively monitor emotional states and fatigue. Moreover, involving a more diverse group of participants could lead to richer insights.

References

Buchner, J., Buntins, K., & Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, *38*(1), 285–303.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Brooks/Cole.

Gorin, H., Patel, J., Qiu, Q., Merians, A., Adamovich, S., & Fluet, G. (2024). A review of the use of gaze and pupil metrics to assess mental workload in gamified and simulated sensorimotor tasks. *Sensors*, *24*(6), 1759.

Guo, W., & Kim, J. H. (2021, July 25–29). Designing augmented reality learning systems with real-time tracking sensors. In *Advances in neuroergonomics and cognitive engineering: Proceedings of the AHFE 2021 virtual conferences on neuroergonomics and cognitive engineering, industrial cognitive ergonomics and engineering psychology, and cognitive computing and internet of things*. Springer.

Hopstaken, J. F., Van Der Linden, D., Bakker, A. B., & Kompier, M. A. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, *110*, 100–106.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585.

Kim, J. H., Monhanty, S., Pulipati, V., Wang, F., Mostowfi, S., Oprean, D., Wang, Y., & Seo, K. (2024). Assessing student performance through pupil dilation and problem-solving time in augmented reality. In *Proceedings of the human factors and ergonomics society annual meeting*. SAGE Publications.

Kim, J. H., & Yang, X. (2020). Measuring driver's perceived workload using fractal dimension of pupil dilation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 1900–1904.

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. In *Proceedings of the IEEE 7th conference on human factors and power plants* (pp. 7–5–7–9). IEEE.

Mohanty, S., Kim, J. H., Pulipati, V., Wang, F., Mostowfi, S., Oprean, D., Wang, Y., & Seo, K. (2024). Measuring cognitive workload in augmented reality learning environments through pupil area analysis. In *Human-computer interaction: Theories, methods, and tools* (Lecture Notes in Computer Science). Springer.

Nazareth, R. P., & Kim, J. H. (2020, July 16–20). The impact of eye tracking technology. In T. Ahram (Ed.), *Advances in usability, user experience, wearable and assistive technology:*

Proceedings of the AHFE 2020 virtual conferences on usability and user experience, human factors and assistive technology, human factors and wearable technologies, and virtual environments and game design (Vol. 121, pp. 3–9). Springer.

Othman, N., & Romli, F. I. (2016). Mental workload evaluation of pilots using pupil dilation. *International Review of Aerospace Engineering*, *9*(3), 80–84.

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). . Elsevier.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. *Work & Stress*, *14*(1), 74–99.

Yang, X., & Kim, J. H. (2019). Measuring workload in a multi-tasking environment using fractal dimension of pupil dilation. *International Journal of Human–Computer Interaction*, *35*(15), 1352–1361.

Yu, C.-Y., Kim, J. H., Mostowfi, S., Wang, F., Oprean, D., & Seo, K. (2023). Developing an augmented reality-based interactive learning system with real-time location and motion tracking. In M. F. Costabile & F. Paternò (Eds.), *Human-computer interaction—INTERACT 2023: Proceedings of the international conference on human-computer interaction*. Springer.