



DOI: <https://doi.org/10.57125/ELIJ.2026.09.25.02>

How to cite: Mistry, K., Dafoulas, G., Nalli, G., Aria, F. T., Langari, B., & Tsiakara, A. (2026). A multimodal investigation of learner attention and cognitive load in linear and non-linear educational content. *E-Learning Innovations Journal*, 4(2). 25–47.
<https://doi.org/10.57125/ELIJ.2026.09.25.02>

A Multimodal Investigation of Learner Attention and Cognitive Load in Linear and Non-Linear Educational Content

Kajal Mistry*

Academic Assistant in Computer Science, Computer Science, Science and Technology, Middlesex University London, The Burroughs, NW4 4BT, United Kingdom, <https://orcid.org/0000-0001-6472-599X>

Georgios Dafoulas

Professor of Information Systems & Business Computing, Director of Programmes, Computer Science, Science and Technology, Middlesex University London, The Burroughs, NW4 4BT, United Kingdom, <https://orcid.org/0000-0003-2638-8771>

Giacomo Nalli

Doctor and Lecturer in Computer Science, Computer Science, Science and Technology, Middlesex University London, The Burroughs, NW4 4BT, United Kingdom, <https://orcid.org/0000-0002-5667-3429>

Farzad Tahmasebi Aria

Doctor and Academic Assistant in Visual Analytics, Computer Science, Science and Technology, Middlesex University London, The Burroughs, NW4 4BT, United Kingdom, <https://orcid.org/0009-0009-3948-8991>

Bahareh Langari

Doctor and Academic Assistant in Visual Analytics, Computer Science, Science and Technology, Middlesex University London, The Burroughs, NW4 4BT, United Kingdom, <https://orcid.org/0009-0005-6385-063X>



Ariadni Tsiakara

*Associate Lecturer in Visual Analytics and Business Intelligence, Computer Science, Science and Technology, Middlesex University London, The Burroughs, NW4 4BT, United Kingdom,
<https://orcid.org/0000-0001-8732-1429>*

***Corresponding author:** *k.mistry@mdx.ac.uk*.

Received: April 14, 2026 | **Accepted:** August 28, 2026 | **Published:** September 25, 2026

Abstract: The measurement of real-time cognitive load remains challenging in e-learning environments, while self-report measures provide limited insight into learners' moment-to-moment cognitive responses. This study examines whether combining eye-tracking and physiological monitoring can provide such an objective account across different instructional content formats. A Tobii Pro Fusion eye tracker and Polar heart rate sensor were used in a within-subjects study where 52 participants engaged with four types of educational content: text-heavy linear, text-heavy non-linear, graph-heavy linear, and graph-heavy non-linear. Based on gaze patterns, heat maps, and heart rate data, two main reading strategies emerged: F-pattern readers ($n = 23$, 44%) and Z-pattern readers ($n = 29$, 56%). Graph-heavy content was associated with longer completion times and higher physiological markers of cognitive effort. Heart rate increases of more than 30% over individual baseline were consistently associated with periods of higher cognitive strain, suggesting a possible threshold for adaptive instructional intervention. Mean heart rate rose from 92.08 bpm for text-heavy linear material to 122.75 bpm for graph-heavy non-linear content, indicating greater physiological engagement as instructional complexity increased. Although these results suggest practical directions for adaptive e-learning systems, learning analytics dashboards, and content delivery in LMS platforms like Moodle, Blackboard, and Canvas, they are a foundation rather than definitive design rules because this was an exploratory, descriptive study. Based on this, three outputs are proposed: a real-time intervention protocol framework, a personalisation model based on objective learner profiling, and evidence-based content creation guidelines.

Keywords: adaptive learning, cognitive load, digital learning environments, e-learning design, eye-tracking, multimodal learning analytics, physiological sensing.

Introduction

Digital interfaces for learning have expanded rapidly since the pandemic, making the optimisation of online content format a progressively important area for research and design. This



paper proposes three outputs: a real-time intervention protocol framework, a personalisation model based on objective learner profiling, and evidence-based content creation guidelines (Tyng et al., 2017). The deployment of digital interfaces has surged since the pandemic, making online content format optimisation a critical area for study and design (Limniou et al., 2021). Despite significant advancements in educational technology, a large gap remains between current sensing capabilities and practical instructional design. Although multimodal data streams, particularly physiological measures and visual attention data, offer a potential way to close this gap (Bustos-López et al., 2022; Hwang et al., 2018), these streams have not been widely translated into specific, useful recommendations for e-learning designers. We present a multimodal investigation of how four distinct instructional content formats affect learner attention, cognitive load, and engagement, and we translate those findings into practical frameworks for adaptive e-learning design. Specifically, we contribute:

1. Evidence-based identification of cognitive load differences across four instructional content formats commonly encountered in LMS platforms.
2. Multimodal indicators of learner engagement that can serve as proxies for adaptive instructional decision-making.
3. Practical design recommendations for adaptive e-learning systems, including real-time intervention protocols and personalisation strategies.

The study is positioned at the intersection of multimodal learning analytics, cognitive load theory, and adaptive instructional design, three fields whose integration holds significant promise for the future of digital education.

Research Problem

The increased use of digital learning environments has made it essential to explore how learners interact with different types of educational content. Questionnaires and performance assessments are traditional methods for measuring engagement and cognitive load, but they are subjective and do not reflect learners' experiences during implementation. The growing shift toward technology-based learning makes this issue timely and requires a scientifically grounded approach to supporting learners and developing educational materials. The solution to this issue may benefit society by providing a more efficient, accessible and personalised digital learning experience through a deeper understanding of how learners interact with material and process information. From a scientific point of view, this research will contribute to the developing field of multimodal learning analytics and combine eye tracking and psychophysiological measurements. Although previous studies have examined eye-tracking or physiological measures independently, limited research has integrated both approaches across different educational interface designs. Consequently, the problem remains insufficiently explored. This study advances current knowledge by investigating how learners interact with text-heavy and graph-heavy content presented in linear and non-linear formats, providing new insights into attention patterns, cognitive load, and evidence-based educational design.



Research Focus

This study uses eye-tracking and physiological monitoring technology to examine how various instructional content designs affect learners' attention, cognitive load, and engagement.

To determine how visual attention patterns and physiological reactions differ across formats, the study examines how students interact with four instructional interfaces: text-heavy linear, text-heavy non-linear, graph-heavy linear, and graph-heavy non-linear material. The study pays particular attention to relationships between fixation behaviour, reading preferences, and heart rate variations as markers of cognitive processing and mental strain.

The groundwork of this scientific approach of the authors is based on the concept that through the application of multimodal learning analytics, one can get a scientifically valuable insight into the behaviour of the learners, something which is not possible through traditional assessment techniques. This study is designed to enhance flexible learning environments, develop evidence-based guidelines for designing educational content and create personalised learning experiences.

Research Aim

This study aims to generate meaningful correlations among eye movement measures, physiological responses, and learner engagement profiles through four types of educational content interfaces: (1) text-heavy linear, (2) text-heavy non-linear, (3) graph-heavy linear, and (4) graph-heavy non-linear. These four interface types cover the broad spectrum of learning materials used in digital learning environments, such as lecture slides, research reference pages, process diagrams, and concept maps.

The study's multimodal approach is intended to generate outputs directly relevant to developing adaptive learning systems and LMS platforms, such as recommendations for content format selection, real-time intervention protocols, and personalisation strategies based on objective learner profiling.

Research Questions

Three formal research questions guide this study:

RQ1: How do linear and non-linear educational interfaces influence learner visual attention patterns as measured by eye-tracking?

RQ2: How do text-heavy and graph-heavy educational interfaces affect cognitive load as indexed by physiological and attentional indicators?

RQ3: Can multimodal learning analytics provide actionable indicators to inform adaptive instructional design decisions for digital learning environments?



Literature Review

One of the key challenges in education is accurately assessing cognitive load and engagement in real time. Traditional methods, such as self-reporting, testing, and observation, often fall short of capturing a student's cognitive and affective state in real time (Fuentes et al., 2021). Cognitive load is also shaped by the delivery medium, with recent reviews reporting systematic differences between computer- and paper-based learning and assessment (Pengelley et al., 2025). Combining physiological data (heart rate) and eye-tracking data (heat map, gaze plot) makes it possible to better observe how students interact with different learning resources and identify cases of overload and disengagement.

Cognitive Load Theory and Multimedia Learning

As working memory capacity is limited, Cognitive Load Theory (CLT) distinguishes three categories of load that instructional design must manage: intrinsic load, arising from the inherent complexity of the learning material; extraneous load, arising from poorly designed instructional elements such as confusing navigation or disorganised information; and germane load, the mental effort devoted to schema construction and meaningful learning (Sweller, 2011).

Mayer's Cognitive Theory of Multimedia Learning (CTML) extends CLT to digital settings, proposing that learning is most effective when verbal and visual channels are synchronised, and unnecessary processing is minimised (Mayer, 2009). This study empirically tests the CTML prediction that graph-heavy or non-linear interfaces may impose additional, unnecessary load when visual complexity and navigation demands compete with processing instructional content. Self-Regulated Learning (SRL) theory further contextualises these findings: learners differ in their capacity to monitor and regulate their own cognitive states, and multimodal analytics can offer an external scaffold that compensates for gaps in metacognitive awareness (Banihashem et al., 2025; Zimmerman, 2002). Taken together, CLT, CTML, and SRL provide the theoretical basis for treating eye-tracking and physiological data as proxies for cognitive and affective engagement. In this study, we link eye-tracking and physiological indicators to CLT dimensions as follows: prolonged fixation durations or repeated regressions may signal elevated intrinsic load; excessive scanning between areas of interest (AOIs) is interpreted as a marker of extraneous load; and systematic attention to key instructional cues, combined with a moderate and stable heart rate, suggests germane load engagement.

Eye-Tracking in Educational Research

Eye-tracking equipment allows researchers to examine attentional allocation and cognitive processing across visual learning environments by recording eye movements and gaze points (Sola & Qureshi, 2024). In educational research, heat maps and gaze patterns are frequently used to characterise reading behaviour, identify areas of focus, and detect attention shifts. Two patterns are particularly relevant to this study: Z-pattern reading, involving diagonal movement from the top-left toward the top-right and typically associated with visually structured materials, and F-pattern



reading, where attention is concentrated across upper sections before moving vertically down the left side, commonly observed in text-heavy interfaces.

Recent research highlights the growing value of eye tracking in education. Ke et al. (2024) demonstrate its potential to link visual attention with learning objectives and evaluate digital interfaces, while Chytrý et al. (2025) highlight its usefulness for examining reading behaviour, attention allocation, and cognitive processing in digital learning environments. Witt et al. (2024) further demonstrate how gaze-based evidence can examine differences in attention allocation and decision-making between expert and novice teachers.

Eye tracking has also increasingly been applied to cognitive-load assessment and adaptive learning. Sola and Qureshi (2024) demonstrate how real-time gaze analysis could support dynamic content adaptation and cognitive-load management in online learning. Jamil et al. (2023) combined eye tracking with brain activity measures to investigate information processing, while Turčáni et al. (2024) used scalable eye-tracking technology to examine reading comprehension of multimedia-enhanced instructional texts among computer science students. Collectively, these studies support the value of gaze-based analytics for understanding learner attention and cognitive processing.

Physiological Measures in Learning Contexts

Heart rate (HR), heart rate variability (HRV), and galvanic skin response (GSR) index the emotional and cognitive state of learners: elevated HR is associated with increased cognitive load during demanding tasks, while reduced HRV reflects heightened attention and stress-management demands (Becerra et al., 2025; Francisti et al., 2023). Multimodal physiological monitoring is increasingly used to assess learner states in real time. Zangeneh Soroush and Zeng (2023) show how HRV and GSR can identify stress in learning environments, and Becerra et al. (2025) note that physiological signals capture learner states that behavioural data alone may not reveal. Al-Nafjan and Aldayel (2024) further support GSR as a marker of autonomic arousal, and Iarlori et al. (2024) demonstrate the potential of wearable sensors to monitor cognitive workload and support adaptive decision-making.

These signals are also shaped by factors beyond cognitive load, including individual physiological variability, movement, temperature, and emotions unrelated to the task. HR, HRV, and GSR findings therefore require cautious interpretation, particularly where thresholds are researcher-defined rather than statistically validated on larger samples; the present findings are treated as suggestive patterns warranting further investigation rather than definitive evidence of cognitive or emotional state.

Integration of Multimodal Data in E-Learning

Combining eye tracking with physiological data offers a more complete picture of learner behaviour than either modality alone (Giannakos et al., 2019; Wang et al., 2025). Wang et al. (2025) combined portable eye tracking with facial emotion detection in digital game-based vocabulary



learning and identified distinct fixation sequences and attention patterns linked to language ability. Such multimodal approaches are increasingly being extended into adaptive and intelligent learning systems that fuse multiple data streams to model learner states and optimise instructional delivery: Alkhatib et al. (2025) propose an AI-driven semantic fusion framework for adaptive e-learning that links heterogeneous learner data to dynamic system responsiveness and tailored content delivery, while Ramesh and Jayashree (2024) present a framework for incorporating multi-layered learning styles into adaptive e-learning environments to customise instructional routes and improve engagement.

Yürüm (2025) provides a broader methodological assessment of technology-enhanced multimodal learning analytics in higher education, underscoring the value of combining behavioural, physiological, and interactional data in next-generation educational analytics, while Tirumanadham et al. (2025) demonstrate that integrating multimodal data with machine learning can improve student performance prediction. Multimodal techniques of this kind have now been applied across problem-solving tasks, multimedia learning environments, and adaptive learning systems (Ramesh & Jayashree, 2024; Tirumanadham et al., 2025; Yürüm, 2025). Despite this progress, three gaps remain, and the present study addresses them. First, most prior research examines a single content type. In contrast, this study systematically compares four formats across two structural dimensions (text/graph $\times$ linear/non-linear), covering most material types found in typical LMS-based learning. Second, non-linear exploration in digital environments has received comparatively little attention relative to linear information flows. Third, translating multimodal data into concrete e-learning design recommendations, particularly for adaptive systems, remains limited.

Materials and Methods

This study examines the relationship between cognitive load, reading behaviour, and physiological response across four types of educational material, combining eye-tracking and heart-rate measurement to assess participant engagement. The design is within-subjects: each participant completed all four instructional conditions (text-heavy linear, text-heavy non-linear, graph-heavy linear, graph-heavy non-linear), with presentation order counterbalanced across participants to control for order and fatigue effects. Ethical approval was obtained prior to data collection, and all procedures followed the University's research ethics guidelines.

A total of 120 participants remained for the study after six datasets were eliminated because calibration accuracy did not meet the predetermined level. To improve the ecological validity of the results across educational contexts, we purposefully selected a wide range of ages (18-65) and educational backgrounds, including university students, adult learners, professionals in development programs, and members of the online learning community. Initially, G*Power 3.1 was used for an a priori power analysis with a statistical power of 80% and an alpha level of .05. However, a sensitivity power analysis was carried out using the final sample size of 52 people because participant recruitment had already been completed. While smaller effects might not have been consistently detected, the analysis showed that the study was appropriately powered to find

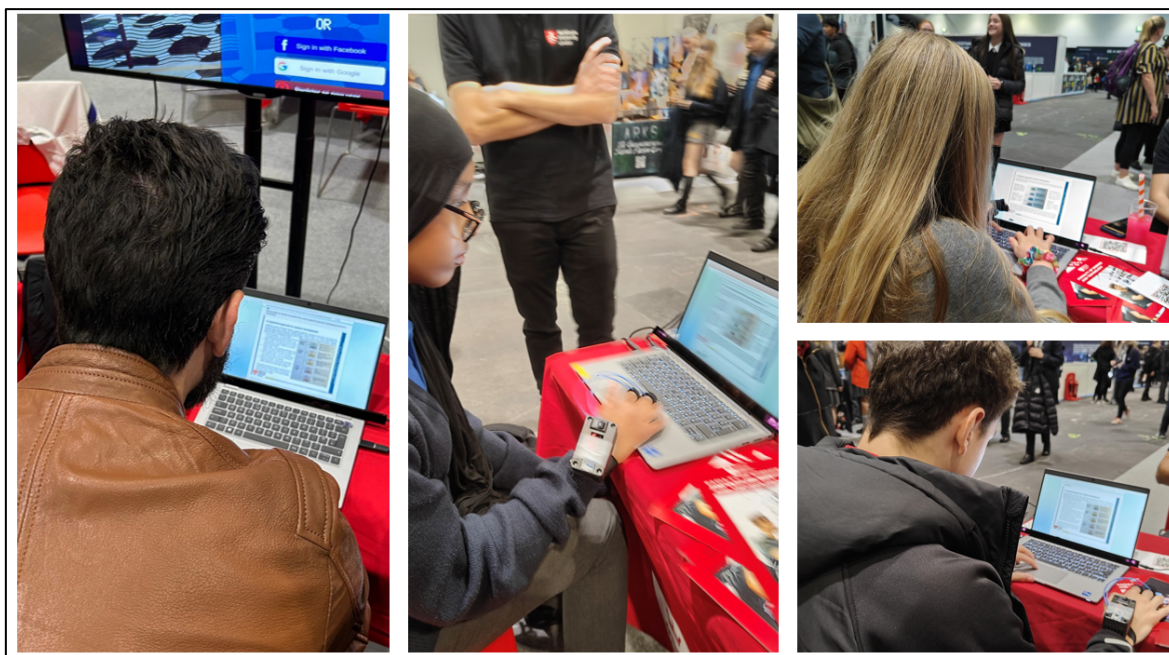


medium-to-large effects (e.g., correlations of around $r = .39$ or bigger). In educational research, where comprehensive multimodal data collection frequently yields moderate sample sizes, the final sample of 52 participants is congruent with exploratory eye-tracking and physiological studies.

Participants wore Polar OH1 wrist-mounted heart rate sensors (as shown in Figure 1). Baseline physiological measures were recorded before task commencement, providing individual reference values critical for interpreting proportional heart rate change. Participants were seated in front of a Dell laptop equipped with a Tobii Pro Fusion eye tracker at a standardised distance of 50–95 cm. Following a calibration procedure using Tobii Pro Lab, the presentation order of the four instructional panels was counterbalanced across participants to minimise order effects. Areas of Interest (AOIs) were defined by logical content divisions (headings, diagrams, body text). Eye-tracking data were recorded using a fixation threshold of 100 ms and a saccade velocity exceeding $30^\circ/\text{s}$.

Figure 1

Experimental setup showing participant positioning, Tobii Pro Fusion eye tracker, laptop, and Polar OH1 heart-rate sensor used for multimodal data collection.



Participants completed tasks involving four types of instructional material: (A) text-heavy with linear flow, (B) text-heavy with non-linear flow, (C) graph-heavy with linear flow, and (D) graph-heavy with non-linear flow. A comprehension question of increasing complexity followed each material. The comprehension questions were developed to assess participants' understanding of the instructional material immediately after each content panel. Each question directly reflected the learning objectives presented in its respective interface and was reviewed before data collection to ensure consistent scope and difficulty across the four instructional conditions. As each instructional panel contained a single comprehension question, internal consistency measures such as



Cronbach's alpha were not applicable. Participants completed one comprehension question following each instructional panel. Responses were collected as part of the experimental dataset to confirm task completion and engagement with the instructional material.

Although these responses were recorded, the present study focused on eye-tracking and physiological measures; therefore, comprehension scores were not included as outcome variables in the statistical analyses reported here. Participants were assigned up to ten minutes, with early completion permitted. This time constraint was deliberately chosen to mirror typical online learning modules, ensuring ecological relevance. Eye-tracking data were recorded continuously throughout, and participants were shown a playback of their gaze data upon completion.

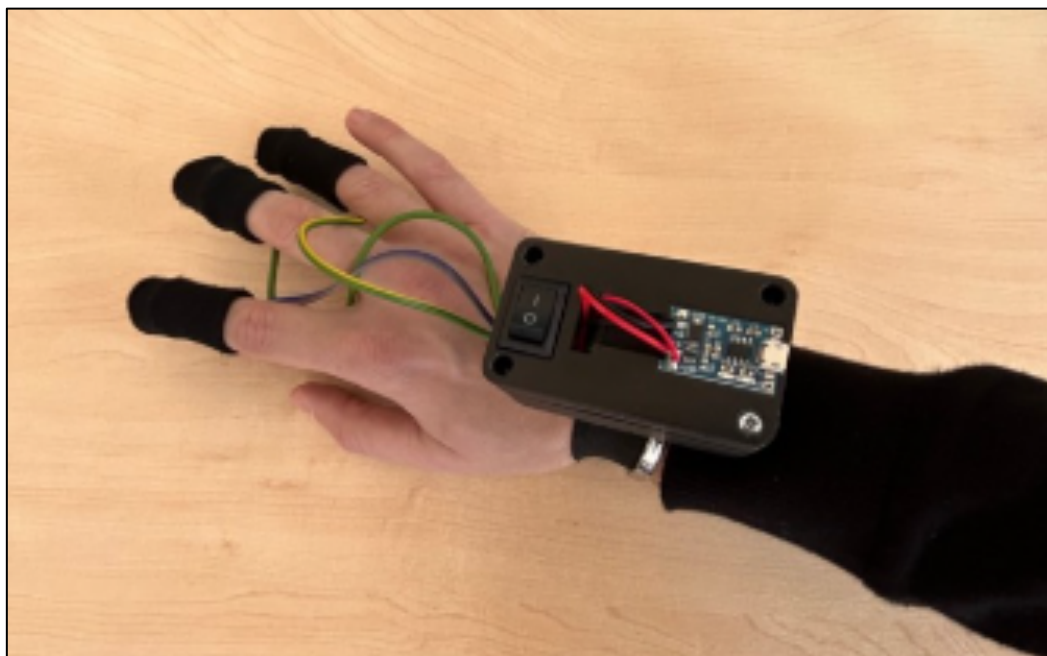
The Tobii Fusion Eye Tracker captures gaze data at a sampling rate of up to 250 Hz, supporting detailed analysis of fixation patterns, saccades, and scan paths. In this study, it measured: (1) fixation duration, as an indicator of attention and local cognitive load; (2) gaze heatmaps, to visualise AOI engagement; and (3) scan paths, to characterise information processing sequences across content formats.

In Figure 2, the IoT biometric sensor, developed and validated in-house (Nalli et al., 2023), included an ESP32 microcontroller with a Wi-Fi module, a pulse sensor for HR detection, and GSR sensors for sweat level detection, housed in a 3D-printed wearable casing with an independent power supply. The device collected approximately 200 records per minute and logged them to a MySQL database via a PHP connection. The Polar OH1 wrist sensor provided cross-validated BPM readings and enabled live HR monitoring during task completion. To minimise physiological confounding factors, participants completed the study individually in a controlled laboratory environment under consistent environmental conditions. Participants remained seated throughout the experiment and were instructed to avoid unnecessary body movement during data collection. Baseline heart-rate values were recorded before the experimental tasks, allowing subsequent interpretation of changes relative to each participant's resting state rather than relying solely on absolute heart-rate values. Although individual physiological differences and external influences cannot be eliminated, these procedures reduced non-task-related variability in the recorded measurements. Continuous participant-level GSR data were retained throughout data collection, enabling the participant-level inferential analyses reported in the Results section below; heart-rate data were recorded and analysed at the panel level only, as per-panel segmentation was not captured at the point of collection.



Figure 2

Wearable IoT physiological monitoring system incorporating heart rate (HR) and galvanic skin response (GSR) sensors used for multimodal data collection during the experimental sessions.



Results and Discussion

Gaze Patterns and Educational Applications

Figure 3 depicts participants' gaze interactions with the four content panels, revealing distinct attentional patterns for each format.

Participants who looked at Panel A (text-rich, linear) showed clusters of fixations near important textual elements, indicating a structured reading approach and an F-pattern, as described in Table 1. This format had the lowest average completion time (2-4 minutes) and the lowest cognitive load measures, suggesting it is suitable for teaching basic concepts and processes. Participants looking at Panel B (text-rich, non-linear) showed scattered fixation points with consistent shifts between elements, suggesting exploratory reading and use of bold text and central images as navigation points. This pattern can be associated with non-linear digital spaces where people look for information, such as wiki pages or research references. According to CLT theory, extraneous load can result in scattered fixations.

Panel C (graph-heavy, linear) showed participants engaging primarily with graphical elements in a sequential left-to-right flow, largely bypassing accompanying text. This structured visual processing pattern is well suited to explaining processes and illustrating cause-effect. Panel D (graph-heavy, non-linear) produced the most widely dispersed fixations, reflecting self-directed exploration consistent with discovery-based learning approaches. This format was associated with



the highest average completion time (8.5-10 minutes), indicating the greatest attentional and processing demands.

Linear layouts (Panels A and C) consistently promoted structured, predictable engagement patterns; non-linear layouts (Panels B and D) elicited exploratory behaviour requiring greater navigational support. These findings address RQ1 and suggest that layout structure may influence visual attention patterns.

Figure 3

Overall gaze plots illustrating participants' visual attention patterns across the four instructional content formats (text-heavy linear, text-heavy non-linear, graph-heavy linear, and graph-heavy non-linear).

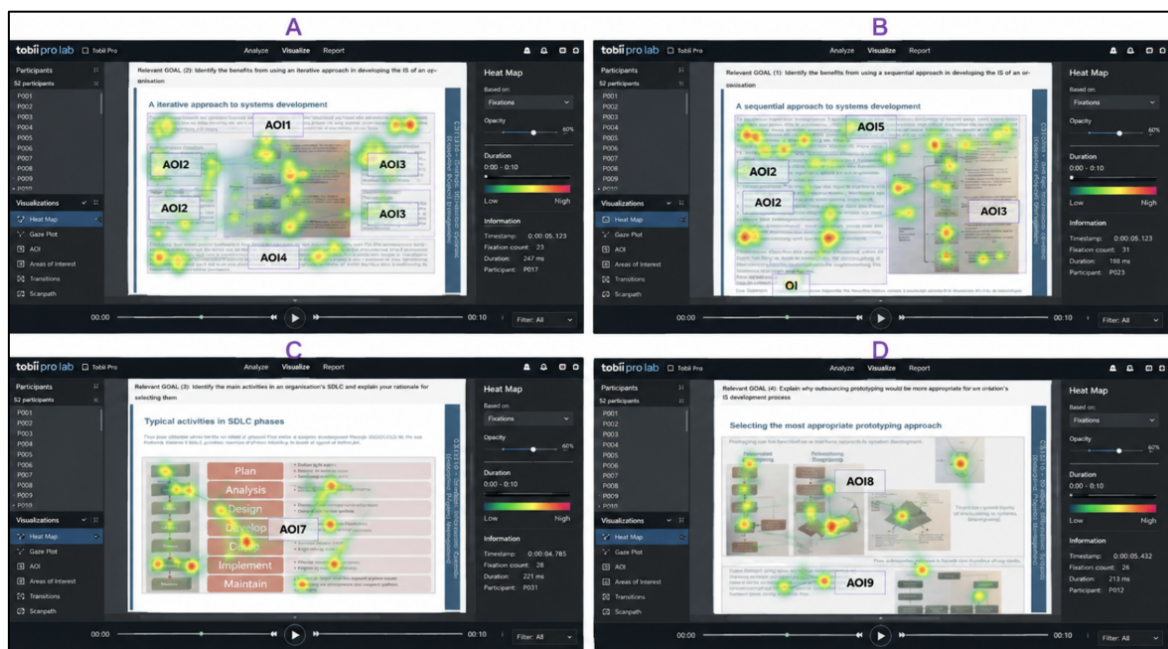


Table 1

Average participant time and educational applications for each layout type

Layout	Content Type	Avg. Time	Cognitive Level	Load	Educational Application
A	Text-heavy, linear	2-4 min	Low-Moderate		Lectures, procedures, foundational concepts
B	Text-heavy, non-linear	5-6.5 min	Moderate		Wiki resources, research materials, reference docs
C	Graph-heavy, linear	7-8 min	High		Process diagrams, SDLC flows, cause-effect visuals



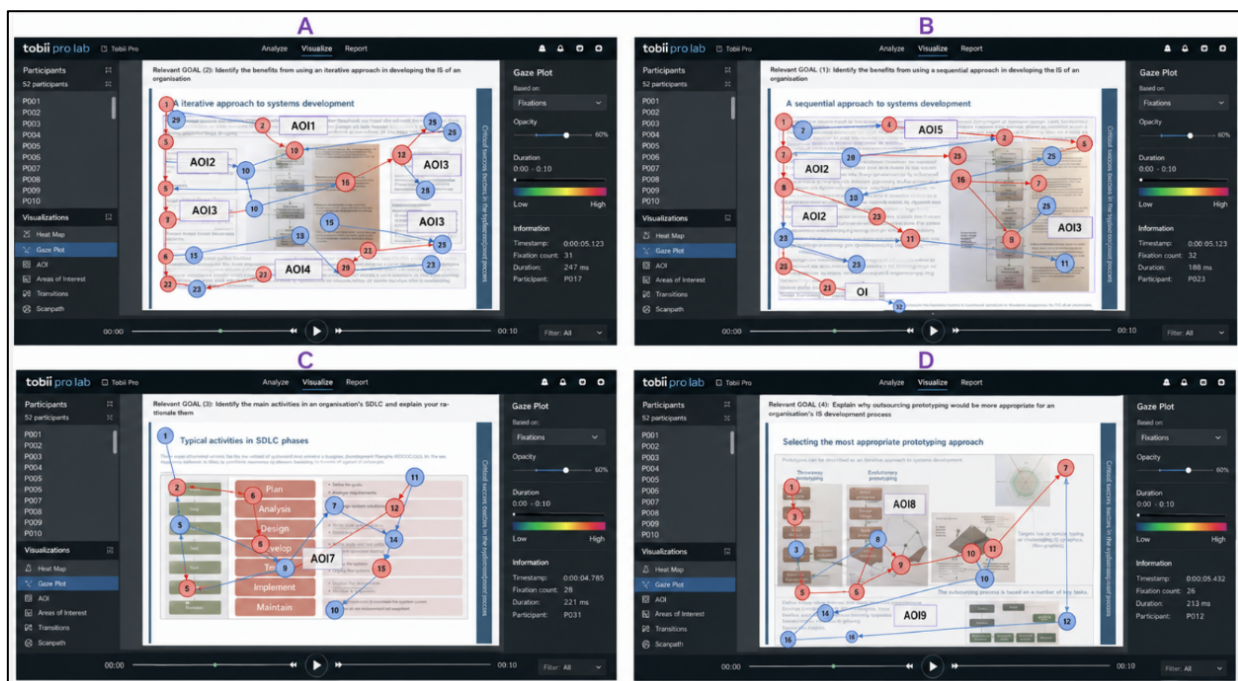
D	Graph-heavy, non-linear	8.5-10 min	Very High	Interactive simulations, concept maps, advanced analytics
----------	-------------------------	------------	-----------	---

Reading Style Identification: Personalising Educational Delivery

The heat map analysis specifies that linear content supports F-Type learners by preserving the sequential processing pathway they rely on, whereas non-linear content risks disrupting that pathway; graphical, non-linear content, by contrast, suits the exploratory processing pattern of Z-Type learners. These two processing preferences have direct implications for e-learning content design: hierarchical structure, headings, numbered lists, and sequential organisation best suit F-Type learners, while multiple entry points, graphical representation, and visual anchors better serve Z-Type learners (Figure 4). A dual-path approach that presents the same content both sequentially and as a visually networked structure offers one way to accommodate both reader types without real-time monitoring. Beyond answering RQ1, these findings provide practical design guidance for LMS-based content development.

Figure 4

Gaze maps illustrating the two dominant reading strategies (F-type and Z-type) identified from participants' eye-tracking data, demonstrating differences in visual attention distribution and navigation behaviour.



Heat Map Analysis

Figure 5's heat map visualisations supported the reading-style differences visible in the gaze plots. F-Type readers ($n = 23$) showed dense fixation on titles, introductory text, and sequential AOIs



within linear panels, but more dispersed attention in non-linear formats, consistent with additional extraneous load when their preferred sequential pathway is disrupted. Z-Type readers ($n = 29$) showed more evenly distributed fixation across AOIs in all content types, reflecting an exploratory style that aligns particularly well with non-linear formats and supports flexible navigation and selective information retrieval consistent with SRL. The visual complexity of graph-heavy non-linear content appears to suit Z-Type readers' natural processing style, potentially encouraging germane load through active model formation.

All four content panels show a clear trend toward higher average heart rates. The panel with the most graphical information (D) shows the highest average heart rate levels. The observed patterns suggest an association between graph-heavy, non-linear materials and increased cognitive demand. However, without replication in larger samples and inferential statistics, the heart rate levels discussed below should not be considered universal intervention requirements for learners. Despite the limited number of data points (four in total), the strong relationship ($r=0.90$) between peak heart rate levels and completion time should be interpreted with caution.

For this dataset, three threshold bands were identified as particularly significant for instructional purposes. For this dataset, three researcher-defined threshold bands were explored. Increases of approximately 20% above individual baseline were treated as an indicative lower-demand range, while increases approaching 30% and 50% were explored as potential markers of increasing cognitive demand. These thresholds are exploratory and require support in larger, prospectively structured datasets. Higher heart rate increases, close to 30% from individual baseline values, consistently appeared alongside longer task completion times and increased dispersion in gaze distribution, pointing to developing cognitive load. Heart rate increases of 50% or higher from individual baseline values indicated the highest cognitive load in the content presentation conditions (Panel D). These observations, if confirmed by further research, can form the basis of real-time adaptive interventions within the LMS systems with physiological data monitoring capabilities. Such an application would need to be calibrated to individual baseline values and integrated into content delivery systems that can dynamically adjust task difficulty.

Table 2

Heart Rate Summary by Content Panel

Panel	Content Type	Mean (bpm)	HR Range	Indicative Load Level	Implications for E-Learning
A	Text-heavy, linear	92.08	53-150	Low-Moderate	Suitable for foundational content; low scaffolding required



B	Text-heavy, non-linear	99.35	48-180	Moderate	Navigation recommended; progress indicators	support consider
C	Graph-heavy, linear	114.06	46-180	High	Scaffolding and examples recommended	worked
D	Graph-heavy, non-linear	122.75	53-180	Very High	Reserve for advanced learners; extensive support needed	adaptive

Table 2 summarises mean heart rate by panel: 92.08 bpm (Panel A, text-heavy linear), 99.35 bpm (Panel B, text-heavy non-linear), 114.06 bpm (Panel C, graph-heavy linear), and 122.75 bpm (Panel D, graph-heavy non-linear), indicating that graph-heavy materials were associated with higher physiological demand than text-heavy materials in this sample. Analysis of Figure 6 (fixation intensity vs mean HR scatter plot) demonstrated that high fixation intensity corresponded with elevated mean heart rates, providing a descriptive indication of convergence between the eye-tracking and physiological measures as a combined approach to cognitive load assessment in e-learning contexts.

The panel-level heart rate comparisons above remain descriptive: per-participant physiological data were not segmented by panel at the point of collection, so a repeated-measures inferential test across Panels A-D was not possible. However, we conducted participant-level inferential analyses on the concurrently collected GSR (skin conductance) data ($N = 50$), a physiological signal separable from panel-specific effects because it was recorded continuously across the task. Paired-samples tests comparing early-task and later-task GSR ($t(49) = -0.42$, $p = .677$), a linear trend test across the task ($t(49) = -1.16$, $p = .252$), and a baseline-versus-task comparison ($t(49) = 0.16$, $p = .873$) all returned non-significant results, with small effect sizes ($|d| < 0.06$ throughout). This null result suggests that GSR alone may not be a sufficiently sensitive marker of cognitive load escalation in this paradigm, and it means the panel-level HR pattern reported above should be interpreted as a descriptive trend rather than a statistically confirmed effect pending duplication with prospectively panel-segmented physiological data.



Figure 5

Heat maps illustrating fixation density for F-type readers across the four instructional content formats



The findings provide exploratory implications for e-learning design, adaptive learning systems, and learning analytics practice. This section discusses the implications of the findings for digital educational settings in the context of the theoretical framework introduced in the Literature Review.

Interface Structure, Visual Attention, and Cognitive Load Theory

The eye-tracking data align closely with CLT and CTML predictions about the distinct cognitive demands imposed by structural and modal differences between content types. Graph-heavy non-linear content (Panel D) produced the most scattered fixation data, the highest average heart rate, and the slowest completion times. This convergent pattern is consistent with elevated extraneous load resulting from the dual demands of extracting and processing instructional content and navigating complex visual structures. According to Mayer (2009), learning efficiency is decreased when unnecessary processing uses up cognitive resources that could be used for generative and necessary processing. In contrast, text-heavy linear material produced structured, predictable fixation sequences and the lowest physiological signs of demand (Panel A). This aligns with CLT expectations about the advantages of logical, in-order-arranged teaching materials for handling intrinsic load, especially for students encountering new material. These findings suggest that text-heavy, sequentially structured formats may be appropriate for foundational and procedural content, while graph-heavy or non-linear presentations may require additional support.

The increased fixation density and elevated heart rate observed in graph-heavy, non-linear content suggest that visual complexity may increase extraneous cognitive load when learners must



simultaneously navigate and interpret multiple information sources. This aligns with Sweller's (2011) redundancy effect and the split-attention effect in CLT: when textual and graphical elements are poorly integrated or require simultaneous cross-referencing, cognitive demand escalates disproportionately to instructional benefit (Sweller, 2011).

Reading Style Differences and Personalised E-Learning Design

Identifying two main reading patterns, F-pattern ($n = 23$, 44%) and Z-pattern ($n = 29$, 56%), has practical implications for individualised content design in e-learning platforms. This distribution is broadly consistent with prior research on reading behaviour on information-heavy websites, lending ecological validity to the LMS-based results reported here. Accounting for such individual differences matters because learner styles have been shown to influence performance in self-instructional online courses (Oliveira et al., 2023). F-Type learners' preference for structure and order may reflect self-regulation and prior-knowledge factors, though this interpretation remains speculative at this stage; Z-Type readers, by contrast, appear to rely on visual support, multiple content pathways, and graphic integration to support their exploratory approach. For LMS systems serving heterogeneous user groups, a dual-path design that presents the same content both sequentially and as a visually networked structure offers a practical way to accommodate both reader types without real-time monitoring. Where real-time eye tracking is available, F/Z classification could be explored as a learner-profiling approach, enabling adaptive content presentation based on the profile revealed during calibration.

Proposed Adaptive E-Learning Intervention Framework

Based on the exploratory patterns observed in this study, a conceptual adaptive intervention framework can be proposed for future validation. Patterns of heart rate threshold identified in this study, though requiring larger samples and inferential confirmation, suggest a practical model for an adaptive intervention system with three responsive stages relative to each participant's baseline: at HR up to 20% over baseline, content delivery proceeds unchanged, indicating appropriate engagement; at HR approaching 30% over baseline, the system could insert a formative check, worked example, simplified content, or progress indicator; at HR around 50% over baseline, more direct intervention is warranted, such as pausing progression, offering direct explanation, or redirecting to simpler content. Eye-tracking data could complement this system by using scattered fixations, excessive AOI transitions, or sustained regressions, which may signal emerging cognitive overload before physiological thresholds are reached, enabling earlier adaptive responses. This architecture is technically compatible with existing LMS platforms such as Moodle, Blackboard, and Canvas via plugin-based extension, browser-accessible eye-tracking APIs (e.g., WebGazer.js), and low-cost wearable heart rate monitors; the main barriers to deployment are organisational (data governance, instructor training, consent protocols) rather than technological.

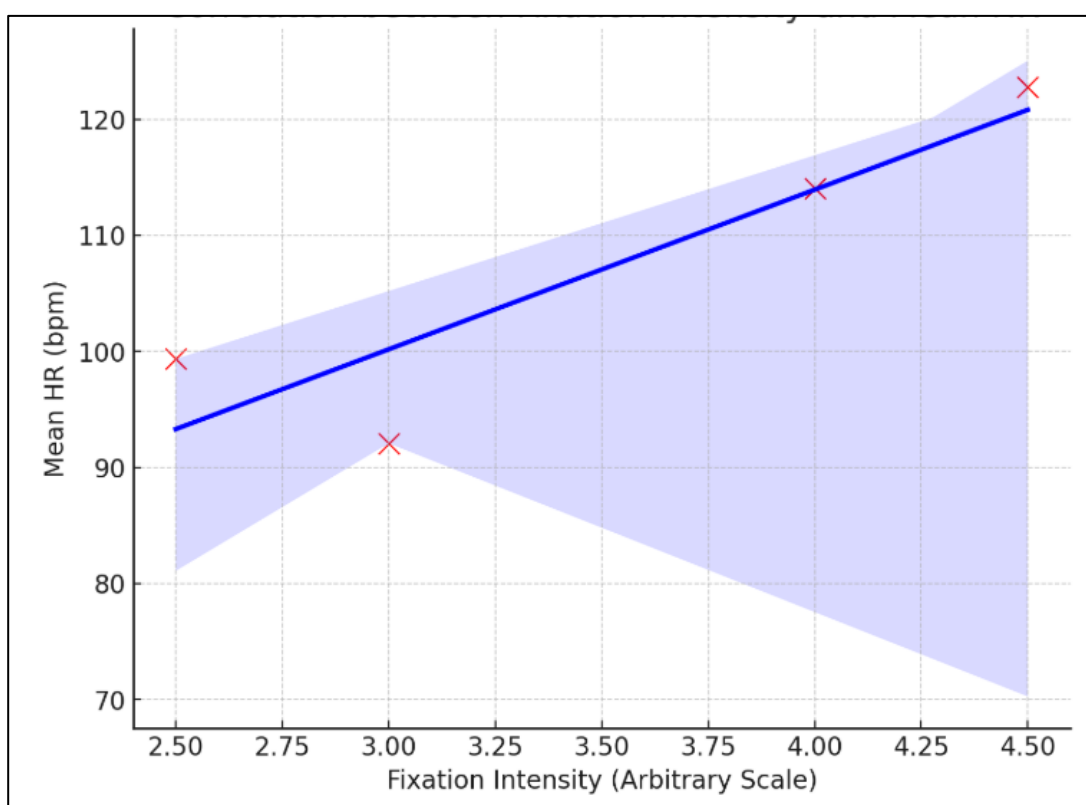
Figure 6 shows a positive correlation between fixation intensity and mean heart rate, a widely used indicator of cognitive effort, engagement, and physiological arousal. The regression line rises



as fixation intensity increases. Participants with fixation intensity scores of 4.0-4.5 showed higher mean heart rates (approximately 114-123 bpm) than those with scores of 2.5-3.0 (approximately 92-100 bpm). The red markers represent the four panel-level observations, and the shaded band represents the 95% confidence interval for the regression line; its width reflects the limited number of data points and signals a need for further statistical analysis before the strength of this relationship can be established with confidence.

Figure 6.

Relationship between fixation intensity and mean heart rate, illustrating the association between visual attention and physiological indicators of cognitive demand



Learning Analytics Dashboards for Educators

Beyond real-time adaptive systems, the multimodal data streams investigated in this study have strong potential as inputs to educator-facing learning analytics dashboards. Visualisations of aggregate fixation patterns, cohort-level heat maps by content type, and alert flags for learners whose physiological indicators suggest sustained high demand could help instructors make evidence-based decisions about content revision, targeted re-teaching, and formative intervention timing. The usability of such educator-facing tools is a key design consideration, since the learnability of authoring and analytics interfaces directly affects teacher adoption (Roldán-Álvarez et al., 2016). Such dashboards could support a shift from mostly retrospective analysis toward more prospective monitoring: instructors would have access to real-time engagement signals during content



interaction rather than reviewing performance data after assessments, allowing for earlier and more targeted intervention. This is particularly relevant for asynchronous online learning contexts, where instructors have no real-time visibility of learner states.

Alignment with Self-Regulated Learning Theory

The reading style differences and adaptive response architecture described above also have implications for Self-Regulated Learning (SRL) theory. These reading patterns may reflect differences in self-regulatory behaviour; however, SRL capacity was not directly measured in the present study: they rely on external content structure to guide their processing and may lack the metacognitive strategies to navigate non-linear environments productively. Z-Type readers' exploratory behaviour, by contrast, may reflect higher SRL capacity, enabling them to construct their own learning pathways through visually complex environments. Multimodal analytics could serve a scaffolding function for lower-SRL learners by externalising the monitoring and evaluation components of self-regulation: the system monitors cognitive load and provides adaptive prompts that substitute for or supplement the learner's own metacognitive awareness. This framing extends the present findings beyond instructional design and into the broader domain of learner support in digital environments.

Conclusions and Implications

This study demonstrates the potential of multimodal learning analytics by combining eye-tracking and physiological monitoring to support evidence-based instructional design decisions in digital learning environments. Across four educational content formats, consistent differences emerged in visual attention patterns and physiological indicators of cognitive demand, with graph-heavy non-linear content associated with the highest attentional and physiological load and text-heavy linear content with the lowest. Two dominant reading strategies were identified, with direct implications for personalised LMS content design.

These findings should be read as preliminary evidence consistent with Cognitive Load Theory and Mayer's Cognitive Theory of Multimedia Learning, rather than as definitive causal claims. The heart rate threshold patterns identified here require replication with larger samples and inferential statistical validation before they could serve as universal intervention criteria. Even so, the convergent patterns across the eye-tracking and physiological data streams provide a coherent foundation for an adaptive e-learning intervention architecture, with practical relevance for LMS design, learning analytics dashboard development, and personalised content delivery.

Future work should evaluate these approaches in authentic online and classroom learning environments, incorporate direct measures of learning performance, and investigate integrating multimodal indicators into adaptive learning systems compatible with existing LMS platforms. The methodology established here by combining Tobii Pro eye-tracking with wearable IoT physiological sensing across multiple instructional content types provides a replicable framework for this next phase of research.



Several limitations should be considered when interpreting the findings and their generalisability.

- **Laboratory environment:** Although the controlled laboratory setting was necessary to ensure equipment reliability and data quality, it may not fully reflect real-world learning environments, where factors such as motivation, distractions, and social interaction can influence learner responses. Future research should therefore validate these findings in authentic classroom and online settings to assess their ecological validity.
- **Sample size and physiological subsample:** The sample of 52 participants provides a useful basis for preliminary analysis but limits statistical power and generalisability. Future studies should therefore use larger and more diverse samples to strengthen the reliability of the findings.
- **Threshold:** The heart rate threshold levels (20%, 30%, 50% above baseline) mentioned in this study are researcher-defined. The criteria were not validated by other methods (NASA-TLX ratings, etc.), so they cannot be used as a general criterion.
- **Single institution:** All participants were students at a single university (UK).
- **Heart rate confounds:** Heart rate is influenced by factors beyond cognitive load, including physical movement, ambient temperature and emotional arousal unrelated to the learning task. Without concurrent measurement of these confounders, heart rate cannot be treated as a pure index of cognitive demand.
- **Physiological individual variability:** Individual differences in baseline heart rate and physiological reactivity mean that population-level thresholds may not be appropriate for all learners. Personalised baselines, as used in this study, partially address this, but real-world deployment will require further individualisation strategies.

Suggestions for Future Research

The present study establishes a foundation for a multimodal approach to evidence-based e-learning design, yet several important directions remain for future investigation.

First, longitudinal tracking of how multimodal engagement indicators evolve as learners develop domain expertise would help distinguish between cognitive overload and productive cognitive challenge, thereby refining the interpretation of physiological thresholds. Understanding these developmental trajectories over time is essential for building adaptive systems that respond appropriately to a learner's changing needs as their proficiency grows.

Second, future work with sufficient sample sizes should incorporate inferential statistical analysis to strengthen empirical validation. This could include ANOVA or Kruskal-Wallis tests on eye-tracking metrics such as fixation count, dwell time, and area-of-interest transitions by content format,



as well as repeated measures ANOVA on heart rate data across panels, accompanied by effect size reporting such as Cohen's *d* and eta-squared.

Third, controlled studies incorporating direct learning performance measures, including retention tests, transfer tasks, and competency assessments, are needed to establish the relationship between multimodal indicators and actual learning gains. Linking physiological and behavioural indicators to measurable learning outcomes would move the field beyond correlational observation toward causal understanding of how attention and cognitive load shape learning.

Fourth, research into scalable and low-cost implementation would broaden the practical reach of this approach. Investigation of browser-accessible eye-tracking solutions, such as WebGazer.js, alongside consumer-grade wearable devices, would support the development of low-resource implementation protocols suitable for deployment across diverse institutional contexts, particularly those with limited technical infrastructure.

Finally, automated integration with learning management systems represents a critical translational step from research to practice. Developing and pilot-testing plugin-based adaptive modules for platforms such as Moodle, Blackboard, or Canvas, which would operationalise the intervention protocol described in the Methodology and Results sections, would bridge the gap between experimental findings and everyday educational application.

Acknowledgements

None.

Conflict of Interest

The authors have no conflict of interest. This research received no external funding.

Funding

This research received no external funding.

References

- Al-Nafjan, A., & Aldayel, M. (2024). Anxiety detection system based on galvanic skin response signals. *Applied Sciences*, 14(23), Article 10788. <https://doi.org/10.3390/app142310788>
- Alkhatib, O., Latif, M., Safi, A., & Qamar, S. (2025). An intelligent AI-driven semantic fusion framework for adaptive and personalised learning in next-generation smart e-learning systems. *Annual Methodological Archive Research Review*, 1–33. <https://doi.org/10.63075/yabnn609>
- Banihashem, S. K., Bond, M., Bergdahl, N., Khosravi, H., & Noroozi, O. (2025). A systematic mapping review at the intersection of artificial intelligence and self-regulated learning. *International*



Journal of Educational Technology in Higher Education, 22, Article 50.
<https://doi.org/10.1186/s41239-025-00548-8>

Becerra, A., Cobos, R., & Lang, C. (2025). Enhancing online learning by integrating biosensors and multimodal learning analytics for detecting and predicting student behaviour: A review. *Behaviour & Information Technology*, 45(9), 1885–1910.
<https://doi.org/10.1080/0144929X.2025.256232>

Bustos-López, M., Cruz-Ramírez, N., Guerra-Hernández, A., Sánchez-Morales, L., Cruz-Ramos, N., & Alor-Hernández, G. (2022). Wearables for engagement detection in learning environments: A review. *Biosensors*, 12(3), 209–215. <https://doi.org/10.3390/bios12030209>

Chytrý, V., Mundoková, N., & Kubiátko, M. (2025). Using eye-tracking in education—A review study. *Education Sciences*, 15(7), Article 853. <https://doi.org/10.3390/educsci15070853>

Francisti, J., Balogh, Z., Reichel, J., Benko, L., Fodor, K., & Turčáni, M. (2023). Identification of heart rate change during the teaching process. *Scientific Reports*, 13, Article 16674. <https://doi.org/10.1038/s41598-023-43763-x>

Fuentes, S., González Viejo, C., Torrico, D. D., & Dunshea, F. R. (2021). Digital integration and automated assessment of eye-tracking and emotional response data using the BioSensory App to maximise packaging label analysis. *Sensors*, 21(22), Article 7641. <https://doi.org/10.3390/s21227641>

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, 48, 108–119. <https://doi.org/10.1016/j.ijinfomgt.2019.02.003>

Hwang, W.-Y., Li, Y.-H., & Shadiev, R. (2018). Exploring effects of discussion on visual attention, learning performance, and perceptions of students learning with STR-supported systems. *Computers & Education*, 123, 225–236. <https://doi.org/10.1016/j.compedu.2018.04.002>

Iarlori, S., Perpetuini, D., Tritto, M., Cardone, D., Tiberio, A., Chinthakindi, M., Filippini, C., Pozzi, S., Chiarelli, A. M., Merla, A., & Monteriù, A. (2024). An overview of approaches and methods for cognitive workload estimation in human-machine interaction scenarios through wearable sensors. *BioMedInformatics*, 4(2), 1155–1173. <https://doi.org/10.3390/biomedinformatics4020064>

Jamil, N., Belkacem, A. N., & Lakas, A. (2023). On enhancing students' cognitive abilities in online learning using brain activity and eye movements. *Education and Information Technologies*, 28(4), 4363–4397. <https://doi.org/10.1007/s10639-022-11372-2>



- Ke, F., Liu, R., Sokolikj, Z., Dahlstrom-Hakki, I., & Israel, M. (2024). Using eye-tracking in education: Review of empirical research and technology. *Educational Technology Research and Development*, 72, 1383–1418. <https://doi.org/10.1007/s11423-024-10327-z>
- Limniou, M., Varga-Atkins, T., Hands, C., & Elshamaa, M. (2021). Learning, student digital capabilities and academic performance during the COVID-19 pandemic. *Education Sciences*, 11(6), 361–372. <https://doi.org/10.3390/educsci11060361>
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Nalli, G., Dafoulas, G., Tsiakara, A., Langari, B., Mistry, K., & Tahmasebi Aria, F. (2023). Hybrid educational environments: Using IoT to detect emotion changes during student interactions. *Interaction Design and Architecture(s)*, 58(1), 39–52. <https://doi.org/10.55612/s-5002-058-001>
- Oliveira, A., Spinola, V., Garrido, D., Teixeira, M. M., Salles, C., & Haddad, A. E. (2023). Influence of learning styles on student performance in self-instructional courses. *PLOS ONE*, 18(7), Article e0289036. <https://doi.org/10.1371/journal.pone.0289036>
- Pengelley, J., Whipp, P. R., & Malpique, A. (2025). A testing load: A review of cognitive load in computer and paper-based learning and assessment. *Technology, Pedagogy and Education*, 34(1), 1–17. <https://doi.org/10.1080/1475939X.2024.2367517>
- Ramesh, M., & Jayashree, R. (2024). Adaptive e-learning environments: A methodological approach to identifying and integrating multi-layered learning styles. *SN Computer Science*, 5(6), Article 772. <https://doi.org/10.1007/s42979-024-03114-7>
- Roldán-Álvarez, D., Martín, E., García-Herranz, M., & Haya, P. A. (2016). Mind the gap: Impact on learnability of user interface design of authoring tools for teachers. *International Journal of Human-Computer Studies*, 94, 18–34. <https://doi.org/10.1016/j.ijhcs.2016.04.011>
- Sola, H. M., & Qureshi, F. H. (2024). AI eye-tracking technology: A new era in managing cognitive load for online learners. *Education Sciences*, 14(9), Article 933. <https://doi.org/10.3390/educsci14090933>
- Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Tirumanadham, N. K. M. K., Thaiyalnayaki, S., & Ganesan, V. (2025). Enhancing student performance prediction using e-learning through multimodal data integration and machine learning techniques. In *Proceedings of the 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)* (pp. 933–940). IEEE. <https://doi.org/10.1109/ICSADL65848.2025.10933211>



- Turčáni, M., Balogh, Z., & Kohútek, M. (2024). Evaluating computer science students reading comprehension of educational multimedia-enhanced text using scalable eye-tracking methodology. *Smart Learning Environments*, 11, Article 29. <https://doi.org/10.1186/s40561-024-00318-5>
- Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, 8, Article 1454. <https://doi.org/10.3389/fpsyg.2017.01454>
- Wang, C.-C., Chen, H.-C., & Hung, J.-C. (2025). Examining learners' cognitive processes and learning emotions in digital game-based vocabulary learning: Evidence from eye-tracking and facial emotion recognition. *Interactive Technology and Smart Education*, 22(2), 304-331. <https://doi.org/10.1108/ITSE-01-2024-0015>
- Witt, J., Schorer, J., Loffing, F., & Roden, I. (2024). Eye-tracking research on teachers' professional vision: A scoping review. *Teaching and Teacher Education*, 144, Article 104568. <https://doi.org/10.1016/j.tate.2024.104568>
- Yürüm, O. R. (2025). Technology-enhanced multimodal learning analytics in higher education: A systematic literature review. *IEEE Access*, 13, 92057-92073. <https://doi.org/10.1109/ACCESS.2025.3572467>
- Zangeneh Soroush, M., & Zeng, Y. (2023). Physiological approaches to students' stress detection. In *Proceedings of the Canadian Engineering Education Association (CEEAA)*. <https://doi.org/10.24908/pceea.2023.17155>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2