

Towards a Mapping for AI-Based Eye-Tracking Analytics in Multimedia Learning

Konstantinos Tsiakas¹^{*}, Francesca Zermiani²^{*}, Deniz Iren¹, Nghia Duong-Trung³, Michael Raschke⁴, Roland Klemke¹, Milos Kravcik³, Ladislao Salmerón², Halszka Jarodzka¹, and Leen Catrysse¹

¹ Open Universiteit, The Netherlands

² University of Valencia, Spain

³ German Research Center for Artificial Intelligence (DFKI), Germany

⁴ Blickshift GmbH, Germany

konstantinos.tsiakas@ou.nl, francesca.zermiani@uv.es

Abstract. Supporting students’ reading comprehension is demanding for teachers, as they have limited insight into the reading processes of each student while they are reading. Although eye-tracking and artificial intelligence (AI) approaches offer strong potential to generate such insights, they remain difficult to deploy and apply in actual classroom practice. A key limitation lies in the lack of a structured approach to translate gaze data into pedagogically sound, classroom-ready, actionable insights for teachers. In this paper, we address this gap with a proof-of-concept mapping that connects the Cognitive Theory of Multimedia Learning (CTML) processes—selecting, organizing, and integrating—and their associated eye-tracking metrics with specific AI methods and HCI-driven visualizations. In this ongoing work, we present an initial mapping derived from a literature review: we identify eye-tracking metrics associated with CTML processes and propose preliminary connections to AI methods and HCI-driven visualizations. Drawing on this mapping, we illustrate our approach with a concrete example from a PISA task. We argue that this structured approach is a necessary step toward designing more targeted applications, moving from gaze data to feedback that teachers can use. By grounding technical and design choices in CTML, this mapping aims to support the development of AI-enhanced tools that are technically sound and pedagogically relevant.

Keywords: Eye-tracking · Artificial Intelligence · Human-Computer Interaction · Teacher Dashboards · Multimedia Learning

1 Introduction

Reading comprehension is a central component of education and a cognitively demanding learning process [18]. Educational practice places heavy demands on teachers, limiting their ability to monitor individual learners’ needs during learning. Eye-tracking offers real-time access to otherwise unobservable data about

^{*} These authors contributed equally to this work.

how learners attend to, process, and integrate instructional material. Yet, eye-tracking data are inherently rich and complex, requiring substantial processing and interpretation before they can yield meaningful educational insights [10].

Despite technical progress, no structured approach exists to translate raw eye-tracking data into pedagogically sound, classroom-ready insights that teachers can act upon. In this ongoing work, we address this gap with a proof-of-concept mapping that links multimedia learning processes to established eye-tracking metrics, AI methods for automated inference, and HCI visualizations for teacher dashboards. Based on an overview of existing research, we present an initial mapping and illustrate it with a concrete example from a Programme for International Student Assessment (PISA) [14] reading task. We argue that this structured approach is a necessary step toward designing targeted applications that move from eye-tracking data to actionable teacher feedback. By grounding technical and design choices in CTML, this mapping aims to support the development of AI tools that are both technically sound and pedagogically relevant.

2 Background and Related Work

2.1 Multimedia learning and eye-tracking

The Cognitive Load Theory (CLT) posits that learning is constrained by the limited capacity of working memory [16], which acts as a filter of how much information can be attended and processed simultaneously. Instructional design that aligns with these constraints can support learning, whereas ineffective design can hinder it by overloading cognitive processing [16]. Building on CLT, the Cognitive Theory of Multimedia Learning (CTML) emphasizes the dual-channel architecture of working memory and its interaction with both instructional materials and long-term memory. It explains how visual (i.e., images in a text) and verbal information (i.e., text) are selected and organized into coherent mental representations, and integrated with prior knowledge. Learning is understood to occur when learners successfully identify relevant information, construct appropriate mental models, and connect these models to existing knowledge structures. Successful learning thus depends on three active processes: selecting, organizing, and integrating [12]. Although these three processes are crucial for successful learning, they remain highly invisible for teachers. Grounded in the eye-mind assumption that what we look at reflects ongoing cognitive processing [9], eye-tracking offers a window into these otherwise invisible processes. Eye movements consist of fixations (periods of relatively stable gaze on a stimulus) and saccades (rapid shifts between stimuli) [19].

Importantly, prior research has linked specific eye-tracking measures to each CTML process [15,3]. For selecting, the focus is on whether learners attend to task-relevant information, typically captured through metrics such as time to first fixation, early fixation patterns, and fixation proportions on task-relevant vs. irrelevant areas. For organizing, the focus is on how learners process and structure information into mental representations, reflecting the experienced cognitive load. This process is measured through fixation duration, total fixation

count on structural elements, pupil size, and dwell time on task-relevant sections. For integrating, the focus is on how learners connect information with prior knowledge, examined through transitions between areas of interest, rereading patterns, and scanpaths. Although the associations between eye-tracking measures and CTML processes offer a common framework, their operationalization is not always straightforward, as the same measure can be interpreted differently depending on the specific task and learning material [3].

2.2 AI and HCI for Eye-Tracking in Multimedia Learning

Recent advancements in AI have enabled the automated classification of specific reading behaviors and reader profiles [2]. Furthermore, AI models can estimate real-time cognitive effort and task engagement [6]. These capabilities provide a foundation for structuring eye-tracking metrics around CTML. In this section, we provide an overview of how AI and HCI methods can be structured around the three CTML processes to translate eye-tracking metrics into pedagogically actionable feedback. The methods described are illustrative examples, each relevant to a particular scenario associated with a given process, rather than an exhaustive or universal mapping.

For the *selecting* process, readers must distinguish between relevant and irrelevant information, Discriminative models, such as *RoBERTEye-Fixations* [7], utilize word-level features to decode information-seeking goals by distinguishing deep reading from scanning. Regarding the *organizing* process, advanced architectures like Time-aware Long Short-Term Memory (T-LSTM) autoencoders [6] can extract latent features from fixation sequences to identify behavioral reader profiles. Simultaneously, the Real-Time Index of Pupillary Activity (RIPA) [1] can provide a continuous indicator of cognitive load and mental effort during reading. To support the *integrating* process, AI models can analyze the connections between relevant areas of interest through scanpath analysis. Algorithms like *MultiMatch* [5,8] can provide quantitative scores for scanpath similarity, while Generative AI [4] can synthesize these into reports for teacher.

From an HCI perspective, these outputs can be rendered accessible via dashboards (e.g., VAAD [13]) which provide interactive attention heatmaps. Frameworks like *MetaDash* [17] employ layered storytelling, allowing teachers to overlay pedagogical layers (e.g., a “mistakes” or “arousal” layer) over a student’s activity timeline to pinpoint cognitive breakdowns [11]. This can further be enhanced by Generative AI agents that synthesize complex multimodal traces into structured narrative reports, linking behavioral reader profiles directly to educational standards to provide actionable instructional recommendations while reducing the teacher’s burden of manual data analysis [4]. Despite the emergence of such advanced AI and HCI methods, there remains a lack of a systematic mapping that aligns these computational techniques with core multimedia learning processes towards transforming complex metrics into structured, narrative insights for teachers and educators.

3 Proposed Approach

Building on previous research which consolidates empirical links between CTML processes and eye-tracking metrics [3,15], we extend this framework by adding candidate AI methods for automated analysis and HCI visualizations designed to make cognitive processes interpretable for teachers. This serves as an initial demonstration of how such a mapping can bridge the interpretative gap between eye-tracking data and instructional action. We illustrate the feasibility of this approach through the Galapagos Islands PISA task (Table 1), a multi-page digital reading task where students navigate a simulated website to locate, organize, and evaluate information. This requires learners to dynamically select relevant facts across four tabs ("*About*", "*Animals*", "*Conservation*" and "*Volunteer*"), organize them into a coherent understanding of the ecosystem, and integrate information across multiple sources. For teachers, this mapping aims to transform eye-tracking data into explanatory pedagogical insights that pinpoint exactly where a cognitive breakdown occurs—be it a selection failure (missing key information), organization difficulty (struggling to structure content), or an integration breakdown (failing to connect related representations).

Table 1. Proof-of-concept mapping of CTML processes to task requirements, scenario descriptions, eye-tracking metrics, candidate AI methods, and possible HCI solutions.

Process	PISA-Task Requirements	Scenario	ET Metrics	Candidate Method	AI Possible HCI Solution
Selecting	Search for and select relevant information from the "Animals" and "Conservation" pages	Student misses key information	Time to first fixation, fixation duration proportion	RoBERTEye-Fixations to decode information-seeking goals [7]	Attention Heatmaps: highlights zones of attention on relevant text [13]
Organizing	Construct and activate the mental representations of the selected texts and images within and across both pages	High cognitive load detected	Total dwell time, pupil dilation, fixation sequences	T-LSTM to extract latent features from temporal sequences [6], RIPA as indicator of cognitive load and mental effort [1]	Layered Storytelling: overlay pedagogical information over activity timelines [11]
Integrating	Connect the mental representations to make inferences and reflect on the content	Student does not connect information from different pages	Number of transitions, scanpath	MultiMatch and KLD for scanpath comparison [5,8]	Narrative Assessment: LLM agents to synthesize traces and translate into structured reports [4]

We demonstrate the application of this mapping with a specific question item from the Galapagos Islands PISA task: *"The 'Animals' and 'Conservation' pages present facts and details about animals and conservation efforts in the Galapagos Islands. Why does the Galapagos Reservation Society focus on this information on two pages?"*. To correctly answer this item, students must locate information on separate pages (selecting), structure the facts about conservation programs

into a coherent understanding (organizing), and reflect on how the information shape their purpose (integrating).

4 Discussion and Conclusion

This paper presented a proof-of-concept mapping that connects the CTML processes with eye-tracking metrics, candidate AI methods, and possible HCI visualizations. Its main contribution is a structured conceptual bridge between multimedia learning theory and the technical design of AI-enhanced teacher support tools. Rather than treating gaze data, AI-based analysis, and dashboard design as separate components, the mapping illustrates how these can be aligned to generate pedagogically meaningful insights into students' learning processes. The presented approach aims to offer an interdisciplinary mapping that fosters further collaboration across multimedia learning, eye-tracking, AI, and HCI research. By bridging these fields, this mapping can support an integrated investigation of how learning processes can be captured, interpreted, and supported by digital technologies. Future versions of the mapping should incorporate task demands, learning characteristics, and the interaction among multiple eye-tracking indicators, as these factors strongly influence the interpretation of eye-tracking measures and the related learning processes. To address this, future work will extend our mapping by conducting a systematic mapping across multimedia learning, eye-tracking, AI, and HCI research, followed by expert validation with developers, researchers and teachers in authentic educational contexts.

Funding. This work is funded by the European Union through the HORIZON EYE-TEACH project (101177543). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

References

1. Abeysinghe, Y., Mahanama, B., Jayawardena, G., Jayawardana, Y., Sunkara, M., Duchowski, A.T., Ashok, V., Jayarathna, S.: A-DisETrac advanced analytic dashboard for distributed eye tracking. *International Journal of Multimedia Data Engineering and Management (IJMDEM)* **15**(1), 1–20 (2024)
2. Arnold, L., Aryal, S., Hong, B., Nitharsan, M., Shah, A., Ahmed, W., Lilani, Z., Su, W., Piaggio, D.: A systematic literature review of eye-tracking and machine learning methods for improving productivity and reading abilities. *Applied Sciences* **15**(6) (2025)
3. Coskun, A., Cagiltay, K.: A systematic review of eye-tracking-based research on animated multimedia learning. *Journal of Computer Assisted Learning* **38**(2), 581–598 (2022)
4. Davalos, E., Zhang, Y., Srivastava, N., Salas, J.A., McFadden, S., Cho, S.J., Biswas, G., Goodwin, A.: LLMs as educational analysts: Transforming multimodal data

- traces into actionable reading assessment reports. In: International Conference on Artificial Intelligence in Education. pp. 191–204 (2025)
5. Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., Holmqvist, K.: It depends on how you look at it: Scanpath comparison in multiple dimensions with multimatch, a vector-based approach. *Behavior Research Methods* **44**(4), 1079–1100 (2012)
 6. Fang, L., Wang, S., Chen, Y., Zhang, S., Liu, Z., Zhong, W.: Analyzing complex educational data: A data analytic framework for integrating structured and unstructured eye-tracking data. *Psychometrika* pp. 1–29 (2026)
 7. Hadar, C.A., Shubi, O., Meiri, Y., Heshes, A., Berzak, Y.: Decoding open-ended information seeking goals from eye movements in reading. *arXiv preprint arXiv:2505.02872* (2025)
 8. Jarodzka, H., Holmqvist, K., Nyström, M.: A vector-based, multidimensional scanpath similarity measure. In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. p. 211–218. ETRA '10, Association for Computing Machinery, New York, NY, USA (2010)
 9. Just, M.A., Carpenter, P.A.: Eye fixations and cognitive processes. *Cognitive Psychology* **8**(4), 441–480 (1976)
 10. Lai, M.L., Tsai, M., Yang, F., Hsu, C.Y., Liu, T.C., Lee, S.W., Lee, M.H., Chiou, G.L., Liang, J., Tsai, C.C.: A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review* **10**, 90–115 (2013)
 11. Martinez-Maldonado, R., Echeverria, V., Fernandez Nieto, G., Buckingham Shum, S.: From data to insights: A layered storytelling approach for multimodal learning analytics. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. p. 1–15. CHI '20, Association for Computing Machinery, New York, NY, USA (2020)
 12. Mayer, R.E.: *Cognitive Theory of Multimedia Learning*, p. 43–71. Cambridge Handbooks in Psychology, Cambridge University Press (2014)
 13. Navarro, M., Becerra, A., Daza, R., Cobos, R., Morales, A., Fierrez, J.: VAAD: Visual attention analysis dashboard applied to e-learning. In: *2024 International Symposium on Computers in Education (SIIE)*. pp. 1–6 (2024)
 14. OECD: *PISA 2018 Results (Volume I): What Students Know and Can Do*. PISA, OECD Publishing, Paris (2019)
 15. Scheiter, K., Schubert, C., Schüler, A., Schmidt, H., Zimmermann, G., Wassermann, B., Krebs, M.C., Eder, T.: Adaptive multimedia: Using gaze-contingent instructional guidance to provide personalized processing support. *Computers & Education* **139**, 31–47 (2019)
 16. Sweller, J., Ayres, P., Kalyuga, S.: *Cognitive Load Theory. Explorations in the Learning Sciences, Instructional Systems and Performance Technologies*, Springer, New York, NY, 1 edn. (2011)
 17. Wiedbusch, M.D., Kite, V., Yang, X., Park, S., Chi, M., Taub, M., Azevedo, R.: A theoretical and evidence-based conceptual design of metadash: An intelligent teacher dashboard to support teachers' decision making and students' self-regulated learning. *Frontiers in Education* **6**, 570229 (2021)
 18. Wijekumar, K.K., Meyer, B.J., Lei, P.: Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development* **60**, 987–1013 (2012)
 19. Yarbus, A.L.: *Eye movements and vision*. Plenum (1967)