

Capturing Learning Experiences in Digital Education: A Multimodal Methodological Perspective

Ilona Pezenka^{}, Larissa Neuburger^{}, Evelyn Fränzl, Melanie Dejnega

Department of Communication, University of Applied Sciences for Management & Communication
Vienna, Austria.

How to cite: Pezenka, I.; Neuburger, L.; Fränzl, E.; Dejnega, M. (2026). Capturing Learning Experiences in Digital Education: A Multimodal Methodological Perspective. In: 12th International Conference on Higher Education Advances (HEAd'26). Valencia, 15-18 June 2026. <https://doi.org/10.4995/HEAd26.2026.21610>

Abstract

Learning analytics derived from learning management systems provide valuable insights into learners' interactions with digital learning environments, yet they remain limited in capturing subjective dimensions of learning such as cognitive load, motivation, emotional responses, and perceived self-efficacy. To address this gap, this paper proposes a multimodal methodological framework for analyzing learning experiences in a digital learning context. The framework integrates objective learning analytics data, psychophysiological measures (eye-tracking and facial expression analysis), qualitative real-time methods (think-aloud protocols), interaction-based conversation analysis of AI-supported tutoring systems, and quantitative self-report measures. By aligning and triangulating these complementary data sources, the approach aims to link observable learning behavior with learners' cognitive, emotional, and experiential processes. The paper outlines the methodological design, discusses its analytical potential through illustrative scenarios, and reflects on practical implementation considerations. In doing so, it contributes to the development of human-centered learning analytics and the evidence-based design of digital learning environments.

Keywords: learning analytics; multimodal methods; learning experience; eye-tracking; human-centered AI; digital learning.

1. Introduction

At a time when digital technologies – especially Artificial Intelligence – are increasingly shaping higher education and changing how students learn (Brandhofer et al., 2024), a deeper understanding how students experience digital learning environments has become a central research concern. While learning analytics have substantially advanced the analysis of students' behavior in digital contexts, much of this work continues to rely mainly on data derived from

learning management systems, such as clickstreams, navigation sequences, drop-off points, test results, number of attempts, and response behavior. As a result, important learning dimensions that shape how students engage with, process, and benefit from digital learning environments remain underexplored.

A growing body of research has highlighted the limitations of purely behavior-based approaches and emphasized the importance of subjective learning experiences for explaining learning processes and outcomes (Chettaoui et al., 2023; D'Mello & Jensen, 2022; Liu & Cui, 2025; Ouhaichi et al., 2023). Constructs such as cognitive load, motivation, emotional responses, and perceived self-efficacy are widely recognized as critical determinants of learning success (Pekrun, 2006; Richardson et al., 2012; Sweller, 1988), yet they are rarely integrated in learning analytics frameworks. Addressing this gap requires methodological approaches that go beyond isolated data sources and enable the integration of objective behavioral indicators with psychological and qualitative perspectives on learning.

Against this background, this paper proposes a multimodal methodological framework for analyzing learning experiences in digital e-learning environments. The framework combines objective learning analytics data (clickstreams, navigation sequences, drop-off points, test results, number of attempts, and response behavior) with psychophysiological measures, qualitative real-time methods, interaction-based analyses, and standardized self-reports. By triangulating these data sources, the approach of this paper aims to capture not only *what* learners do in digital environments, but also *how* they experience learning and *why* certain interaction patterns emerge. Based on this methodological focus, the paper addresses the following research questions:

RQ1: What characterizes students' learning experiences when interacting with digital e-learning content?

RQ2: How do cognitive, emotional, and motivational factors shape these experiences?

RQ3: How can insights into these learning experiences inform further improvements of digital e-learning content?

2. Background

An emerging body of literature in learning analytics and educational data mining demonstrates that behavioral data alone are insufficient to capture key dimensions of learning (Chettaoui et al., 2023; D'Mello & Jensen, 2022; Liu & Cui, 2025; Ouhaichi et al., 2023). While clickstreams, log data, and interaction traces provide valuable insights into observable learner behavior, they offer only limited explanatory power with regard to learners' cognitive states, emotional reactions, or motivational dynamics. As a result, critical aspects of the learning process often remain implicit or are inferred only indirectly.

Educational and psychological research has already established the relevance of cognitive load, motivation, emotional responses, and perceived self-efficacy for learning processes and learning outcomes (Pekrun, 2006; Richardson et al., 2012; Sweller, 1988). However, these constructs are rarely operationalized directly within classical learning analytics setups. Recent review studies indicate that multimodal approaches combining behavioral data with psychophysiological, qualitative, or quantitative measures remain scarce in higher education research and are typically limited to small-scale pilot studies rather than being systematically integrated into broader analytical frameworks (Mu et al., 2020; Ouhachi et al., 2023).

Prior research has already explored multimodal approaches to understanding learning processes, particularly within the framework of self-regulated learning. Azevedo (2005) has demonstrated how digital learning environments, such as hypermedia, can be conceptualized as metacognitive tools that support learners' cognitive and metacognitive processes. Building on this perspective, subsequent research has combined process data (e.g., log data), eye-tracking, and think-aloud protocols to capture learning processes in real time and to model the dynamic nature of self-regulated learning in digital environments (Blikstein & Worsley, 2016; Järvelä et al., 2021). In addition, studies have shown that metacognitive and motivational regulation prompts can effectively support learners' engagement and performance in digital learning contexts (Daumiller & Dresel, 2018). While these approaches provide important methodological and theoretical foundations for multimodal learning research, they are often conducted in controlled settings and primarily focus on specific aspects such as self-regulation or affective processes. Building on this line of work, the present paper extends the perspective by proposing a structured framework that integrates behavioral, psychophysiological, qualitative, and interaction-based data to capture broader dimensions of learning experiences in digital learning environments.

3. Methods

Psychophysiological data (eye-tracking and facial expression) will be collected under controlled laboratory conditions using iMotions software for the integrated analysis. Eye-tracking data will be used to capture gaze behavior, providing insights into attentional distributions and information processing, while facial expression analysis will be employed to identify indicators of emotional responses such as cognitive overload, frustration, or engagement during the use of selected digital learning content. Data collection will take place while participants work through the digital content. These methods are particularly suitable for the analysis of digital learning environments, as they allow for continuous and non-invasive measurement. Video recordings will subsequently be annotated, and visual and auditory data will be analyzed in conjunction with the sensor data. All data will be processed exclusively in anonymized or pseudonymized form and in compliance with applicable data protection regulations. Table 1 provides an overview of the methods and the expected data for each method.

Psychophysiological measurements will be complemented by think-aloud protocols. Think-aloud protocols constitute a qualitative data collection method in which participants verbalize their thoughts, considerations, and decision-making processes while completing a task (Charters, 2003). The aim is to make cognitive processes, comprehension strategies, uncertainties, and perceptions visible as they occur during the learning or problem-solving process. In contrast to retrospective surveys, think-aloud protocols capture cognitive processes in real time and thus provide direct access to subjective learning and understanding processes.

A further methodological approach will consist of qualitative conversation analysis of dialog sequences with AI-based tutoring systems (chatbots) that are used as part of the digital learning environments. Qualitative conversation analysis is an interpretative method for the systematic examination of dialogical interactions (Monteiro et al., 2023). In the context of AI-based tutoring systems, it will be used to analyze the structure, dynamics, and content-related quality of dialogues between learners and AI tutors. The analysis will focus on how students formulate questions, request support, respond to follow-up prompts, and how AI tutors pedagogically structure these interactions.

Table 1. Data type and constructs per method.

Method	Data Type	Learning Experience Constructs Captured
Learning Analytics (LMS log data, clickstreams)	Objective, behavioral	navigation patterns, engagement indicators, drop-off points, number of attempts, quiz results, interaction duration;
Eye-Tracking	Psychophysiological, objective	visual attention, information processing, attentional distribution, indicators of cognitive load (indirect), gaze;
Facial Expression Analysis (FEA)	Psychophysiological, objective	emotional responses (basic emotions: joy, anger, sadness, fear, disgust, surprise, frustration, engagement), cognitive overload signals;
Think-Aloud Protocols	Qualitative, real-time	user experience, cognitive processes, comprehension strategies, uncertainties, decision-making processes, subjective perceptions during learning;
Conversation Analysis (AI-supported tutoring dialogues)	Qualitative, interaction-based	support needs, interaction dynamics, question formulation, pedagogical structure of AI–learner dialogue, quality of tutoring interactions;
Quantitative Self-Report Questionnaires	Subjective, standardized	perceived usefulness, subjective learning success, self-efficacy, learner satisfaction, learning experience;

In addition, selected dimensions of learners' experiences will be assessed quantitatively using a questionnaire administered immediately after participants complete the digital learning content. Constructs such as usability, perceived usefulness, and subjective learning success will be measured using standardized scales. This quantitative component will enable structured comparison and will support the interpretation and validation of the qualitative findings.

To enable meaningful integration, all data sources are temporally aligned using synchronized time stamps, allowing behavioral, psychophysiological, and qualitative data to be linked to specific moments within the learning process. This alignment makes it possible to relate interaction events to concurrent gaze behavior, facial expression indicators, think-aloud verbalizations, and corresponding AI-supported dialogue sequences. Rather than analyzing each data source in isolation, the approach follows a triangulation logic, with particular emphasis on convergence. In this context, convergence refers to cases in which multiple modalities consistently indicate the same underlying learning phenomenon, thereby strengthening the interpretability of observed patterns.

In addition to convergence, the analysis also considers complementarity and divergence across data sources. Complementarity refers to cases in which different modalities provide distinct but mutually informative perspectives on the same learning episode, thereby enriching the interpretation of observed behavior. Divergence captures instances in which data sources indicate inconsistent patterns, which may point to latent processes that are not directly observable and therefore require further qualitative interpretation.

4. Expected Results

As this paper is part of an ongoing project, it discusses the potential of a multimodal methodological approach to capture and interpret learning experiences in digital learning environments. By systematically integrating behavioral learning analytics with psychophysiological measures, qualitative data, and self-reported learning experiences, the proposed approach enables an in-depth analysis of learning processes that goes beyond observable patterns of interaction. In particular, the multimodal integration is expected to uncover learning obstacles, phases of cognitive overload, and motivationally critical moments that cannot be adequately derived from behavioral data alone.

The analysis of these multimodal data is anticipated to generate actionable insights for the continuous enhancement of digital learning content. Expected results include empirically grounded indications for adjusting content structure, task formats, feedback mechanisms, and interaction design also to better align digital learning environments with learners' cognitive and experiential needs (Figure 1). From a methodological perspective, it will be demonstrated how multimodal data can be triangulated to produce robust interpretations of learning processes and to link behavioral traces with subjective learning experiences in a systematic manner.

To illustrate the analytical potential of the proposed approach, the integration of multimodal data can be exemplified through typical learning situations. In one scenario, eye-tracking data may indicate irregular or dispersed gaze patterns on specific interface elements, suggesting difficulties in orienting within the digital learning content. This can be further supported by think-aloud verbalizations in which learners express uncertainty or confusion regarding the function or meaning of these elements. Building on these initial observations, additional data sources are analyzed to deepen the interpretation: LMS data may reveal repeated interactions with these elements, or increased task completion times, while facial expression indicators may point to surprise, frustration or cognitive strain. In combination, these converging and complementary signals suggest that the interaction design or instructional guidance is unclear to learners. Such insights can inform targeted improvements in interface design, content structure, and guidance mechanisms to enhance usability and support more effective learning processes.

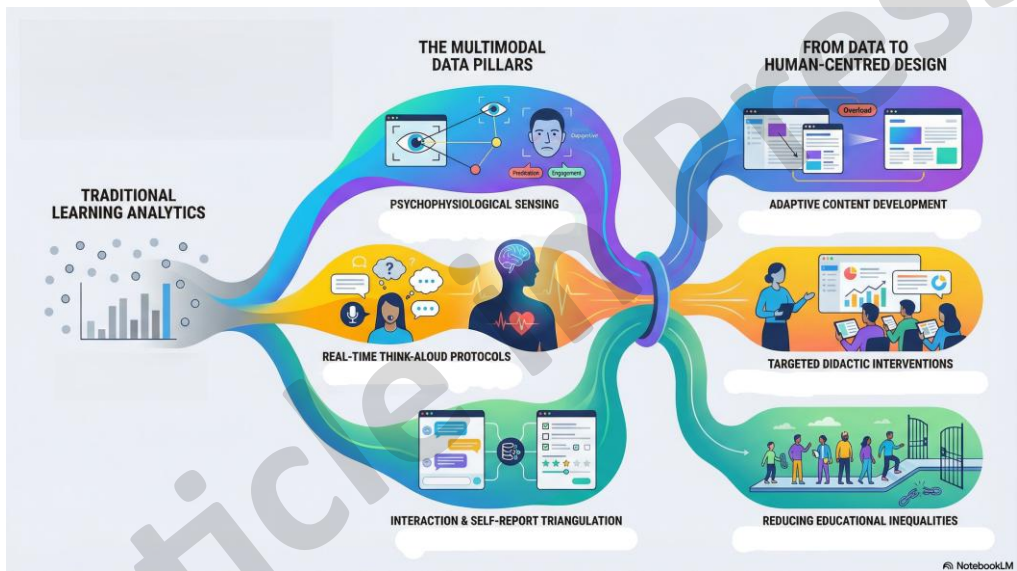


Figure 1. Overview of methods and results. Source: Own source (developed with NotebookLM)

However, the results should not only help to optimize learning designs but also provide instructors with valuable information. The expected results include more differentiated insights into learners' comprehension difficulties and cognitive-emotional states, which can lead to meaningful didactic interventions and support more responsive instructional design (Figure 1).

Despite its analytical potential, the implementation of multimodal approaches in authentic educational settings remains challenging. Methods such as eye-tracking, facial expression analysis, and think-aloud protocols can be intrusive and may affect natural learning processes, which often limits their use to controlled settings. To address this, the proposed framework is

conceptualized as a flexible and staged approach. While scalable data sources such as learning analytics and AI-supported interaction logs can be applied in real-world contexts, more intensive methods can be used selectively to investigate specific learning situations in greater depth. In this way, multimodal analyses provide targeted insights that can inform the design and improvement of digital learning environments.

Finally, the proposed approach contributes to the discourse on human-centered digital education by illustrating how learning experience data can be used to identify group-specific burdens and barriers in digital learning environments. By making subjective learning experiences empirically accessible, this approach supports the evidence-based design of inclusive and learner-centered digital learning content.

5. Conclusion

This paper proposes a multimodal methodological framework for the analysis of learning experiences in digital learning environments. By systematically combining objective learning analytics data, psychophysiological measures, qualitative real-time methods, interaction-based conversation analysis, and standardized self-report instruments, the approach extends existing learning analytics frameworks beyond behavior-based perspectives. The proposed combination of methods enables the triangulation of objective usage data with learners' subjective experiences, thereby providing richer insights into cognitive, emotional, and motivational processes that shape learning.

As a methods-focused contribution, the paper outlines how different data sources can be meaningfully integrated to interpret observable learning behavior. The proposed framework is designed to support both research-driven analyses of learning processes and practice-oriented improvements in the design of digital learning content. More broadly, the approach contributes to the development of human-centered learning analytics by offering a scalable and theoretically informed methodology for capturing learning experiences in increasingly AI-supported educational environments. Future research may build on this framework to refine analytical procedures, explore additional data sources, and examine its applicability across different educational contexts.

References

- Azevedo, R. (2005). Using Hypermedia as a Metacognitive Tool for Enhancing Student Learning? The Role of Self-Regulated Learning. *Educational Psychologist*, 40(4), 199–209. DOI: 10.1207/s15326985ep4004_2
- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. DOI: 10.18608/jla.2016.32.11

- Brandhofer, G., Gröblinger, O., Jadin, T., Raunig, M., & Schindler, J. (Hrsg.). (2024). Von KI lernen, mit KI lehren: Die Zukunft der Hochschulbildung. Projektbericht. Norderstedt: Books on Demand. <https://fnma.at/content/download/2990/19034>
- Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal*, 12(2).
- Daumiller, M., & Dresel, M. (2018). Supporting Self-Regulated Learning With Digital Media Using Motivational Regulation and Metacognitive Prompts. *The Journal of Experimental Education*, 87(1), 161–176. DOI: 10.1080/00220973.2018.1448744
- D'Mello, S., Jensen, E. (2022). Emotional learning analytics. In Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, Agathe Merceron (eds.). *Handbook of learning analytics*, 120-129. DOI: 10.18608/hla22
- Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, 72, 101203. DOI: 10.1016/j.learninstruc.2019.04.004
- Liu, X., & Cui, Y. (2025). Eye tracking technology for examining cognitive processes in education: A systematic review. *Computers & Education*, 105263. <https://doi.org/10.1016/j.compedu.2025.105263>
- Monteiro, M. D. S., Pereira, V. C., & Salgado, L. C. D. C. (2023, October). Investigating politeness strategies in chatbots through the lens of Conversation Analysis. In *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems* (pp. 1-12).
- Mu, S., Cui, M., & Huang, X. (2020). Multimodal data fusion in learning analytics: A systematic review. *Sensors*, 20(23), 6856. <https://doi.org/10.3390/s20236856>
- Chettaoui, N., Atia, A., & Bouhlel, M. S. (2023). Student performance prediction with eye-gaze data in embodied educational context. *Education and information technologies*, 28(1), 833-855. <https://doi.org/10.1007/s10639-022-11163-9>
- Ouhaichi, H., Spikol, D., & Vogel, B. (2023). Research trends in multimodal learning analytics: A systematic mapping study. *Computers and Education: Artificial Intelligence*, 4, 100136. <https://doi.org/10.1016/j.caeai.2023.100136>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational psychology review*, 18(4), 315-341. <https://doi.org/10.1007/s10648-006-9029-9>
- Richardson, M., Abraham, C., Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin* 138(2), 353-387. <https://doi.org/10.1037/a0026838>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2), 257-285. https://doi.org/10.1207/S15516709COG1202_4