Synthesizing Multiple Classroom Observer Logs with LLM: Comparing Approaches to Generate Learning Analytics Artifacts of Collaborative Circuit Building Task

Rwitajit Majumdar[1,*], Sunny Prakash Prajapati[2], Gayathri Pothancheri[3], Nobleson Kunjappy[1] and Takeshi Ohkawa[1]

[1]*Research and Education Institute for Semiconductors and Informatics, Kumamoto University, Japan*
[2]*Centre for Educational Technology, Indian Institute of Technology Bombay, India*
[3]*School of Social and Behavioral Sciences, Amrita Vishwa Vidyapeetham, Amritapuri, India*

Abstract

This research explores how large language models (LLMs) can be used to synthesize multiple classroom observer logs to better understand learners' cyber-physical reflective learning processes in an electronics experiment class. In hands-on engineering contexts, learning unfolds through interactions across physical components (e.g., breadboards, components, wires), digital applications, and social exchanges with peers and guidance from the instructor and the teaching assistants. While observer logs provide rich qualitative accounts of these processes, integrating multiple perspectives into a coherent analytical representation remains challenging. To address this gap, we conducted a pilot study in an undergraduate electronics laboratory where 4 observers recorded structured and unstructured logs of student activities, including interactions with equipment, peers, and instructional materials. We explored different LLM-based pipelines to aggregate, align, and synthesize these heterogeneous logs into temporally organized narratives and analytically meaningful segments. The paper compares 4 LLM interaction approaches that focus on providing context to create learning analytics artifacts aimed at enabling the instructor, as a user, to identify key phases in the learning process and take orchestration decisions. Preliminary findings suggest that changing prompting approaches synthesized different data ingestion methods within the model and produced artifacts that are difficult to capture in a single-observer analysis. In particular, the method highlights that defining context for learning analytics in the prompt lies in a spectrum from being implicit and offloading all the decisions to the LMM to being explicit about roles and rubrics that the LLM can then use to process the data. The study contributes a novel methodological approach for multimodal learning analytics. and offers insights into its utility for understanding the design of cyber-physical learning environments that better support reflective practice. Current limitations and future directions are discussed.

Keywords

Classroom Observation, Generative AI, Learning Analytics, Human-AI interactions, Engineering Education

1. Introduction

Laboratory classes are central to learning in many engineering domains, as they provide students with opportunities to engage directly with physical systems while applying conceptual knowledge. In electronics experiment classes in particular, students are required to identify components, assemble and test circuits, analyze connections, and iteratively troubleshoot and refine their designs. At the same time, these activities are increasingly mediated by digital tools, such as simulation environments, measurement interfaces, and data acquisition systems, which provide real-time feedback and support analysis. As a result, learning in such environments unfolds across intertwined physical and cyber spaces, requiring students to coordinate actions, interpretations, and decisions across multiple modalities. In addition,

HAI-Agency: Workshop on Orchestrating Human and AI Agency for Proactive and Reflective Learning, co-located with the 27th International Conference on Artificial Intelligence in Education (AIED 2026), Seoul, Korea, June 27 - July 3, 2026.
*Corresponding author
majumdar@kumamoto-u.ac.jp (R. Majumdar); 22d1201@iitb.ac.in (S. P. Prajapati); gayathri.p@ammachilabs.org (G. Pothancheri); nobleson@kumamoto-u.ac.jp (N. Kunjappy); ohkawa-takeshi@kumamoto-u.ac.jp (T. Ohkawa)
https://researchmap.jp/rwito (R. Majumdar)
0000-0003-4671-0238 (R. Majumdar); 0000-0002-2051-1866 (S. P. Prajapati); 0009-0005-2500-4512 (G. Pothancheri); 0000-0003-2715-4504 (N. Kunjappy); 0000-0002-6536-6439 (T. Ohkawa)

laboratory learning is inherently social: students collaborate with peers, consult instructors and teaching assistants, and engage in shared problem-solving processes. These layers of interaction—physical, digital, and social—collectively shape the learning experience and influence how students achieve the intended learning objectives in practical classes.

Capturing these complex, multimodal learning processes poses significant methodological challenges. While digital traces from learning management systems or specialized software can provide detailed logs of learner interactions, they are inherently limited to activities performed within those systems [1]. Conversely, physical actions can be recorded using video-based approaches, including multimodal learning analytics setups; however, such methods often require carefully controlled environments to avoid occlusion and ensure data quality [2]. Human observation remains a flexible and context-sensitive approach for capturing rich classroom activity, particularly in dynamic settings where learners move between tools, spaces, and interactions [3]. To mitigate observer bias and cognitive overload, structured observation protocols are often employed, enabling observers to focus on predefined aspects of activity while maintaining consistency across observations [4]. Nevertheless, human observation is inherently selective and constrained by the observer's attention, perspective, and interpretation.

These challenges are further amplified in multi-observer settings, where multiple individuals document classroom activity simultaneously. While such approaches can increase coverage and provide complementary perspectives, they introduce new complexities in data integration and analysis. Observer logs may vary in granularity, focus, and completeness, leading to sparsity and fragmentation in the data. Observers may also experience tension between adhering strictly to observation protocols and capturing emergent, unexpected phenomena that appear pedagogically significant. Moreover, the purpose of analysis can diverge: on one hand, researchers may seek to understand the underlying learning processes and interaction patterns; on the other, instructors may require synthesized insights to inform real-time or post-hoc orchestration of the class [5]. Consequently, transforming multiple, heterogeneous observer logs into a coherent and actionable representation of classroom activity remains a non-trivial challenge.

To address this gap, this study adopts a design-based research approach [6] to investigate how classroom interactions (amongst learner, peer, devices, teaching assistants) can be captured and synthesized in an electronics experiment class. As a pilot study, our primary aim is to explore the practical and analytical needs that arise when attempting to document and interpret cyber-physical reflective learning processes in authentic classroom settings. Specifically, we examine how multiple observer logs can be integrated using Large Language Models (LLMs) to provide a holistic understanding of learners' interactions across physical, digital, and social dimensions.

The research is guided by the following question: *What are the approaches to utilize LLMs to synthesize multiple classroom observer logs for analyzing learning processes in an electronics experiment class?*

By addressing this question, the study seeks to contribute methodological insights into multimodal learning analytics and to inform the design of LLM interaction approaches that support both research-oriented analysis and classroom orchestration.

2. Pilot Study

The pilot observation was conducted in a public university in East Asia, involving third-year undergraduate students enrolled in an interdisciplinary data science program. The study took place during an introductory electronics laboratory session, which is part of a newly established course. The session lasted three hours, with 19 out of the 20 registered students in attendance. The class was facilitated by one instructor, supported by one technical assistant and two teaching assistants. This session followed an initial lecture-based introduction, making it the second meeting in the course sequence.

During the laboratory session, students engaged in hands-on activities integrating both physical circuit construction and digital interaction through microcontroller programming. After successfully establishing a connection between the Arduino board and their personal computers via a USB interface, students were assigned three sequential tasks: 1. Constructing a voltage divider circuit to control the blinking of an LED light. 2. Building a variable resistor circuit to explore adjustable input values. 3. Developing an analog joystick interface to capture and interpret input signals.

Each task was accompanied by structured instructional scaffolds designed to guide students through the assembly and programming processes. Following these guided steps, students were required to complete exploratory exercises, interpret the results, and document their findings in a lab report. These activities required students to coordinate physical manipulation of components, interpretation of digital feedback, and iterative problem-solving, thereby exemplifying a cyber-physical learning environment.

As this was the first year the laboratory course was implemented, the instructor expressed a need to better understand classroom dynamics, including student engagement, collaboration patterns, and points of difficulty. Consequently, an observation-based data collection approach was adopted to capture rich, in-situ information about the learning process and to inform future instructional design and orchestration decisions.

An initial observation protocol was developed collaboratively by the four observers involved in the study. The protocol design drew on established frameworks from prior research. Specifically, a making-oriented design framework [7] was adapted to categorize observations related to electronic making practices, reflecting the embodied and iterative nature of the tasks. To capture collaborative interactions, dimensions for assessing the quality of collaboration in dyadic settings were incorporated, based on established work in computer-supported collaborative learning (CSCL) [8] .

The resulting observation protocol was implemented through an online platform to facilitate real-time data entry by observers. Two versions of the protocol were developed: a comprehensive version (P1), which included detailed categorical and descriptive fields, and a simplified version (P2), designed to reduce observer cognitive load while maintaining essential data capture. Figure 1 illustrates the interface elements included in both protocol versions and the structure of the sample log data. Table 1 provides the details of the observers and the data collected.

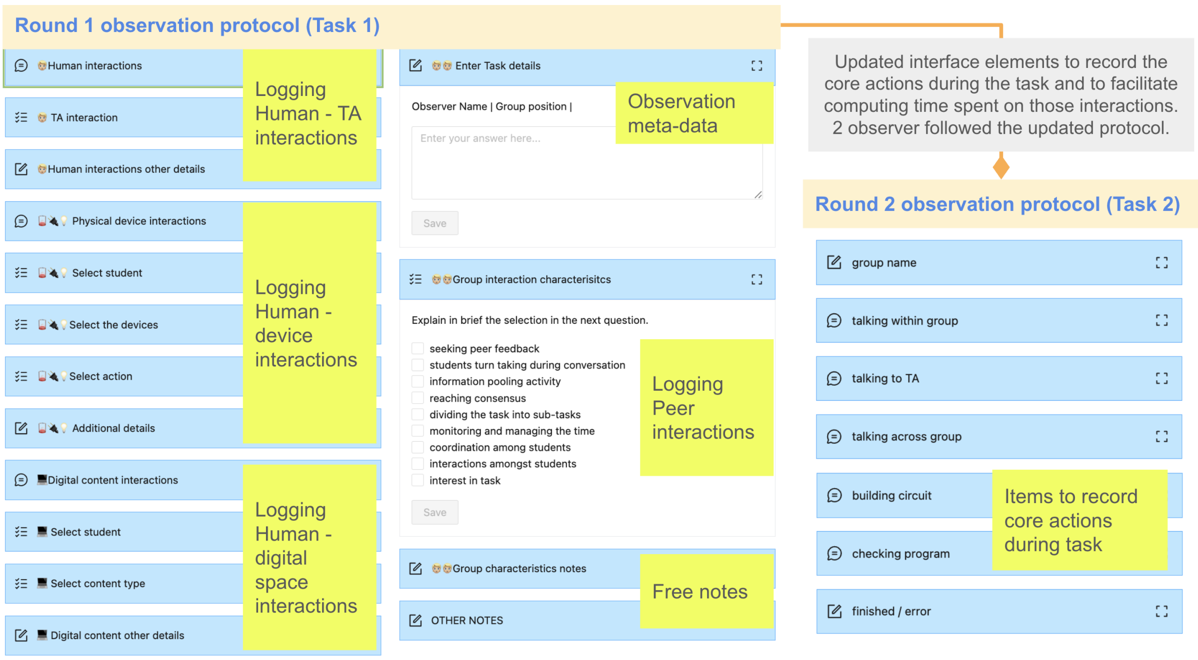

Figure 1: Observation recording interface

2.1. User Goals

Based on the instructor's need we formulated two goals to determine the usefulness of the artifacts created by LLM from the synthesis of the observation data:

Table 1
Details of Observers and Data Collected

Observer	Data Collection detail	Comments
Observer 1 (Ph.D. student)	Used Protocol 1 (P1) for all tasks (T1-3)	817 logs with P1
Observer 2 (Ph.D. student)	Used free notes for all tasks	77 event logs
Observer 3 (Post-doc Researcher)	Used P1 for T1 and P2 for T2,T3	618 logs with P1 189 logs with P2
Observer 4 (Faculty member)	P1 for T1 and P2 for T2,T3 and class timeline	831 logs

1. *Understanding the learning process during the collaborative circuit building task*: This goal include aspects like nature of collaboration, student engagement, topics that students struggled with, etc. relevant for determining how students are learning during the lecture.
2. *Assistance in orchestrating the lecture sessions with collaborative task*: This includes the information on the task components that assist the instructor in orchestrating lecture sessions, such as, time spent by students in different task phases, learner interaction dynamics, students' task progress, usage of task materials, etc.

3. LLM Interaction Approaches and Designed Analytics Artifacts

We define the LLM interactions for creating the analytics artifacts using the human-LLM interaction taxonomy suggested by Gao et al. [9], which consists of four phases, namely: *Planning, Facilitating, Iterating, and Testing*. Additionally, we scoped our interactions to the context-based interaction mode to align the LLMs output with the user goals. For our artifact creating approaches, these phases are operationalized as follows:

- Planning phase entails the information or instructions given to create the artifact, which can include log files, instructions on the 'look and feel' of the artifact to suit the instructor, etc.
- Facilitating phase involves the use of refined prompts and evaluating the suggestions given by the LLM in order to achieve the goals, which is to generate meaningful artifact for the instructor.
- Iteration phase includes the prompts that are oriented to adjust outputs, such as any hallucinated information that may interfere with the data interpretation, interface related changes, and so on
- Testing phase, includes generation and evaluation of multiple variations for the given prompts, such as changes in color scheme, composition, writing style, etc.

Context inputs, which range from simple textual information to complex data files, grounds the instructions given to the LLM [9]. For the present study, the context is composed of four parts, which are: *a) The manner in which the observation log files are input, b) Descriptors of (or related to) the uploaded files, c) Output constraints or specification, and d) Data interpretation rules for the synthesis.* We illustrate four approaches with varied formation of implicit contexts, summarized in Figure 2.

3.1. Offloading artifact creation with minimal specified constraints (Approach 1)

In approach 1, all observation log files were provided to the LLM at once, and the data descriptor is specified in the form of location information ("*the three attached files are of the same* ***classroom****...*"), and the output constraint is specified in terms of the output type ("*I want to produce a* ***visual artifact*** *from this*") in the prompt. No additional information related to how to interpret the data was provided. The text prompt input for the LLM was: "*All these three data from the three attached files are of the same classroom, I want to produce a visual artifact from this for the instructor such that they can use it to make some sense*", and Figure 3 shows the output.

Approach	Data	Data descriptors	Output contraints	Interpretation rules
1.	All at once	Minimal data description	Minimal/ Loosely specified output constraints (e.g., output type)	-
2.	All at once	Minimal data description	Output constraint using schematic image specification	-
3.	Progressive ingestion into LLM	Adequate data description (Data structure, variable descriptions, etc.)	Output constraint in file format (HTML)	Explicit mention of including learning theories in interpretation
4.	All at once	Adequate data description (Data structure, variable descripts, providing data labels, etc.)	Output interface, and file format (HTML) constraints	Explicit mention of including specific learning theories in interpretation

Figure 2: Approaches of prompting in this study

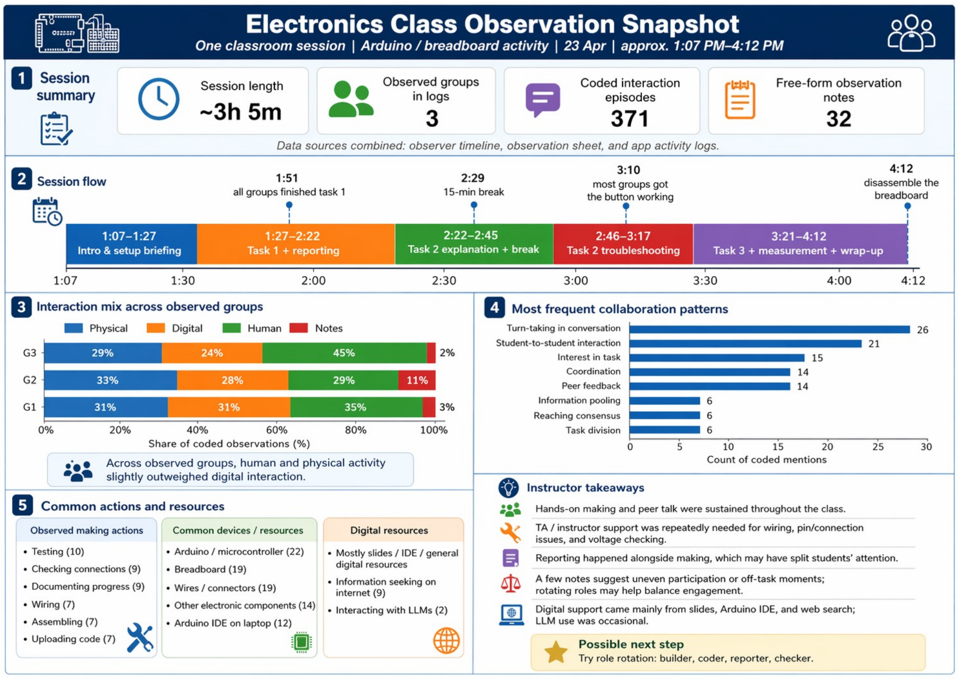

Figure 3: Sample English report generated by vision model ChatGPT v5.5, using approach 1.

3.2. Artifact creation with user specified constraints (Approach 2)

In this approach, observation log files and data descriptor was similar to the previous approach, with an explicit output constraint specified using an image (see Figure 4 (b)). The input image provides a schematic of the visual elements for depicting the lecture events and classroom interactions. The goal here was to support instructors in visually interpreting classroom activity data in a meaningful way, though rules related to interpreting the data was not provided. The final artifact is shown in Figure 4(b), which was obtained after multiple iterations required for fixing the visualization and the interactive behavior of the dashboard.

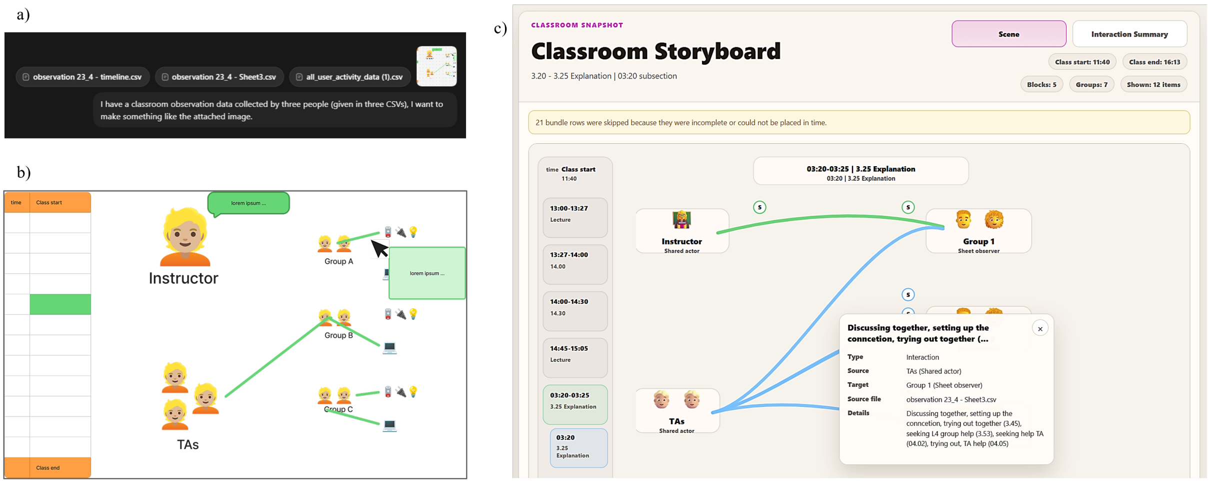

Figure 4: a) Cropped screenshot indicating the prompt text with uploaded log files and an image in the LLM platform, b) Schematic image to constraint the output by indicating structure of the data visualization, c) Dashboard visualization panel created by the LLM.

3.3. Artifact creation with progressive data input with descriptors, constraints, and interpretation rules (Approach 3)

Building on the previous approach, here the observation log files are input progressively for the LLM's context. The prompt corresponding to each file includes detailed data descriptors specifying the variable descriptions and logging procedure, e.g., "*...They used a web-based tool to log the data. clicked is when the element was clicked and timespend denotes the time when it was closed, giving us the total time for the activity described by the element. Answered is the text entry for a given element*". The output constraint was specified in the form of file format (HTML), additionally, instructions on including learning framework for data interpretation was specified in the prompt, e.g. "*...include a learning framework that best describes the data*". Figure 5 shows the output artifact.

3.4. Artifact creation with all data input with descriptors, constraints, and interpretation rules (Approach 4)

This approach resembles Approach 3, but differs in the data input method (which in this case is similar to approaches 1 and 2). Additionally, instructions on using specific learning theories (Vygotsky's Zone of Proximal Development (ZPD) [10] and Kolb's Experiential Learning Cycle (ELC) [11]) to interpret the data was the part of the interpretation rule. Figure 6 shows the data visualizations from the artifact that integrates the specified learning theories. During the generation of the artifact the context window for the agent was not exhausted during the run.

4. Discussion and Conclusion

4.1. User goals and methods for observing in-class activities

This study set out to explore how multiple classroom observer logs can be synthesized using large language models (LLMs) to better understand cyber-physical reflective learning processes in an electronics experiment class.

Prior work on engineering class observation focused on learning behaviors of students across phases of active learning in introductory computer science class [12], college students' learning with multiple devices [13] and electronics lab activities in the African socio-educational cultural context[14] . Observations in engineering classroom often had structure and protocols [15] for precise collection of

Figure 5: Artifact incorporating learning theory to interpret observation logs

evidence. Tools for classroom observation were also proposed earlier [16]. However with the collected data reanalysis to understand another construct is difficult. For an exploratory study to understand the dynamics observers need to be more flexible to capture various things happening during the activity, but compiling remains challenging. The findings from this pilot highlight both the promise and the complexity of leveraging LLMs within learning analytics workflows, particularly in contexts where data is heterogeneous, partial, and grounded in human interpretation.

4.2. Utility of LLM-generated learning analytics artifacts

Artifact creation with minimal context specifications resulted in creation of a report with multiple information segments that provided information on the session summary and task flow, aggregate patterns related to the collaboration, interactions with the artifact and common activities. Additionally, this approach also synthesized the 'instructor takeaways' that was a condensed qualitative summary of the session, and 'possible next step', which was a recommendation for trying out a suggested orchestration strategy (see Figure 3). This synthesis servers the user goal of assisting instructor with the orchestration, as the information such as the task flow, session summary, common student activities, etc. could help with the reflections of the current lecture, and suggestions (takeaways and next-steps) could provoke them to try new strategies in their future sessions. The assistance with the interpretation of classroom learning was minimal as the only synthesis directly relevant was related to the collaboration patterns and that too was present perhaps due to the nature of the observation data from protocol 1.

Providing output constraint in the context positively affects the interpretability, as the output follows the visualization structure suggested by the user. For instance in this approach the representation of the lecture as event segment blocks, and the interaction information using connectors between the entity symbols and the frequency using the connector line strength. Moreover, upon mouse hover over the links or the entities, the instructor can get additional information of the corresponding data items in the observation. This approach will allow the instructor to understand the multiple artifact usage

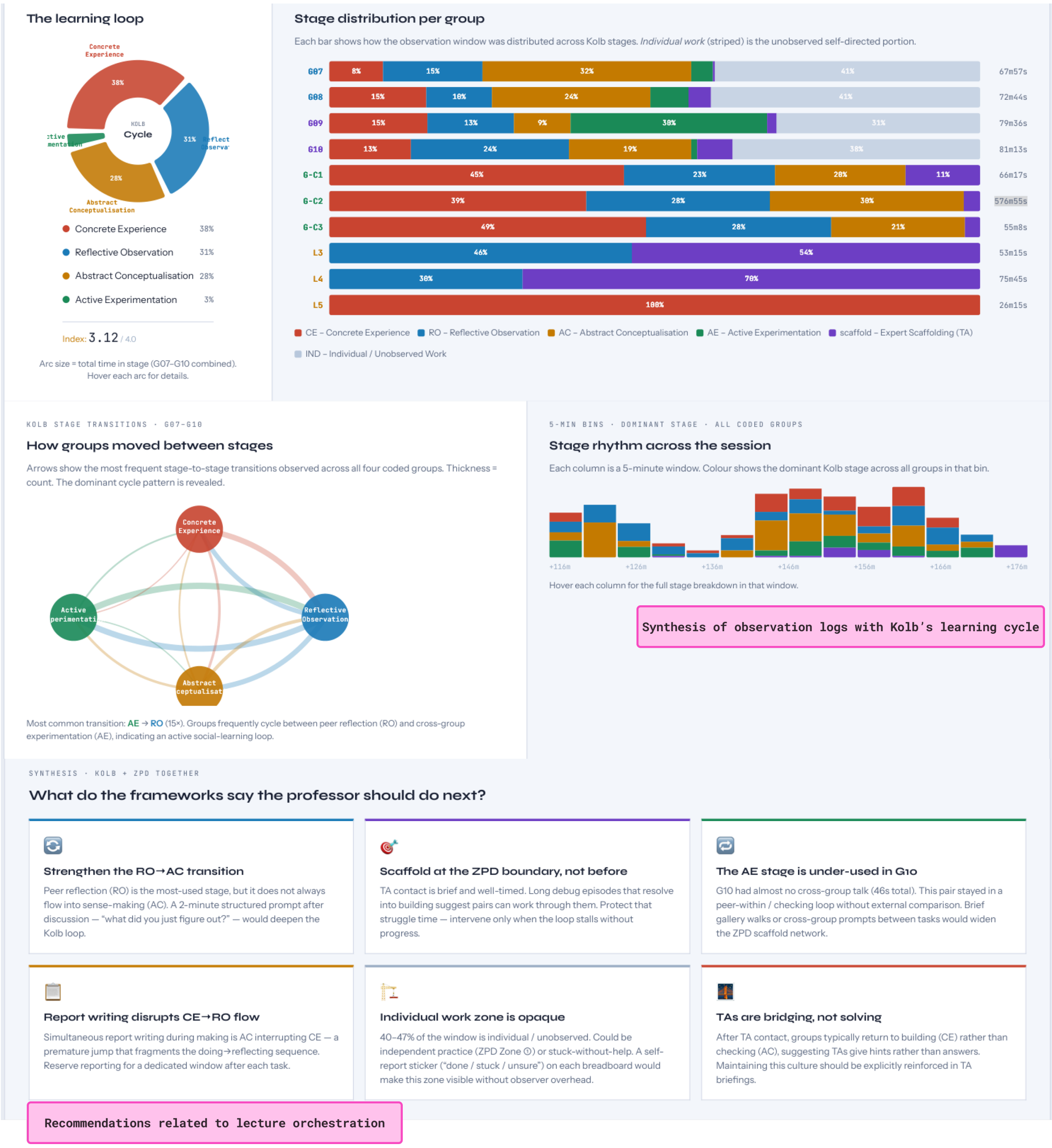

Figure 6: Artifact with data visualizations created based on Kolbe's cycle and recommendations for lecture orchestration.

pattern, device usage, and in general the learning artifact ecology [13] for the student groups that could assist in the learning design.

By explicitly stating to interpret the observation log data using theories of learning in approach 3 produces the artifact that accomplishes both user goals, i.e., of understanding the learning process, and assistance in lecture orchestration (to a certain extent). In this instance the LLM selects the learning theory that best fits the data based on self-reasoning, thus generating visualization grounded in theory.

Similarly, artifact created using approach 4 encapsulates all the elements from the previous approaches extended to all observation logs. It provides synthesis based on the specified learning theory (Kolb's learning cycle in this case), and recommendations related to the lecture orchestration that are grounded in the this theory thereby accomplishing both user goals. For both of these approaches (3 and 4), we noticed LLM's tendency to ignore the data which does not fit the given interpretation rules. Additionally, these approaches also revealed a limitation related to hallucination. For instance, as shown in Figure 5, the stacked bar chart (in box no. 3) of different interaction types. While the first three categories (Physical, Digital, Human) were grounded in the observation log data, the category 'Notes' was generated to represent information inferred from the observer notes.

4.3. Emerging spectrum of prompting learning context for LLMs

A central contribution of this work is the articulation of a context declaration spectrum when harnessing LLMs for learning analytics. Building on context-based prompting approaches [9], our study shows that providing raw observational data alone is insufficient to ensure meaningful synthesis. In exploratory classroom observations, even when structured protocols are used, the collected data varies in granularity, focus, and completeness. As a result, the context of the learning episode—including task structure, pedagogical intent, and situational nuances—may remain only partially represented in the input to the LLM.

We conceptualize this challenge as a spectrum ranging from explicit context declaration (where users clearly define task structure, roles, analytical goals and declare protocols to synthesise) to implicit context reliance (where the model is expected to infer context from fragmented inputs). The five approaches explored in this study largely reside toward the implicit end of this spectrum, demonstrating that while LLMs can generate coherent narratives and identify patterns, their outputs are shaped significantly by how context is framed—or left unspecified—by the user. This has important implications for the design of LLM-supported analytics pipelines: effective use requires not only data preparation but also deliberate articulation of context and analytical intent.

At the same time, the study reveals practical tensions between exploratory observation and structured analysis. Observers must balance adherence to protocols with responsiveness to emergent classroom phenomena, leading to variability in the captured data. LLMs offer a mechanism to bridge these fragmented perspectives, but their effectiveness depends on how well these variations are reconciled during prompting and synthesis. This reinforces the need for hybrid approaches that combine human sensemaking with machine-supported aggregation.

4.4. Current limitations and future directions

We acknowledge several limitations in the pilot study. First, it was conducted in a single classroom setting with a relatively small number of participants, which limits the generalizability of the findings. Second, the approaches explored relied on advanced reasoning-capable LLMs, which may not be readily accessible to many instructors due to cost, technical barriers, or institutional constraints. This raises concerns about the scalability and equity of such approaches in broader educational contexts. Third, the study focused primarily on methodological exploration rather than systematic evaluation of accuracy or reliability of the synthesized outputs.

Several avenues for future research emerge from this work. First, further investigation is needed into how different LLMs—and different configurations of the same model—affect the consistency and validity of synthesized learning analytics outputs, particularly in relation to hallucination and inference biases. Second, user-related factors such as prompting strategies, domain expertise, and familiarity with LLM interaction paradigms are likely to influence outcomes; systematic studies comparing these variations would provide valuable insights. Third, the role of interface affordances (e.g., differences between environments such as code-based interfaces and conversational agents) warrants closer examination, as these shape how users construct and refine prompts.

Finally, future work should explore methods for more structured context encoding, potentially through semi-automated pipelines that integrate observation protocols, metadata annotation, and LLM prompting. Such approaches could help move along the context declaration spectrum toward more explicit and reproducible analytical practices. By addressing these challenges, LLM-supported synthesis has the potential to become a powerful tool for understanding complex, multimodal learning processes and informing the design of cyber-physical learning environments.

Acknowledgments

This research was partially supported by the following grants: JST Z2025L8290089, JST Z2025L8290090 and JSPS 24KK0051.

Declaration on Generative AI

During the preparation of this work, the authors used OpenAI-GPT-5.5 and Claude for generating the artefacts, and Grammarly in order to: Grammar and spelling check. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] D. Gašević, S. Dawson, G. Siemens, Let's not forget: Learning analytics are about learning, TechTrends 59 (2015) 64–71.

[2] P. Blikstein, M. Worsley, Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks, Journal of learning analytics 3 (2016) 220–238.

[3] L. Cohen, L. Manion, K. Morrison, Research methods in education, routledge, 2002.

[4] R. Bakeman, J. M. Gottman, Observing interaction: An introduction to sequential analysis, Cambridge university press, 1997.

[5] P. Dillenbourg, Design for classroom orchestration, Computers & education 69 (2013) 485–492.

[6] Design-Based Research Collective, Design-based research: An emerging paradigm for educational inquiry, Educational researcher 32 (2003) 5–8.

[7] S. Davies, P. Seitamaa-Hakkarainen, K. Hakkarainen, Knowledge creation through maker practices and the role of teacher and peer support in collaborative invention projects, International Journal of Computer-Supported Collaborative Learning 19 (2024) 283–310. doi:10.1007/s11412-024-09427-2.

[8] A. Meier, H. Spada, N. Rummel, A rating scheme for assessing the quality of computer-supported collaboration processes, International Journal of Computer-Supported Collaborative Learning 2 (2007) 63–86. doi:10.1007/s11412-006-9005-x.

[9] J. Gao, S. A. Gebreegziabher, K. T. W. Choo, T. J.-J. Li, S. T. Perrault, T. W. Malone, A Taxonomy for Human-LLM Interaction Modes: An Initial Exploration, in: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, ACM, Honolulu HI USA, 2024, pp. 1–11. doi:10.1145/3613905.3650786.

[10] L. S. Vygotsky, M. Cole, Mind in society: Development of higher psychological processes, Harvard university press, 1978.

[11] D. A. Kolb, Experiential learning: Experience as the source of learning and development, FT press, 2014.

[12] A. Kothiyal, R. Majumdar, S. Murthy, S. Iyer, Effect of think-pair-share in a large cs1 class: 83% sustained engagement, in: Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research, ICER '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 137–144. doi:10.1145/2493394.2493408.

[13] Q. Li, Z. Xu, Y. Chen, Understanding college students' cross-device learning behavior in the wild: Device ecologies, physical configurations, usage patterns, and attention issues, Computers & Education 229 (2025) 105259. doi:10.1016/j.compedu.2025.105259.

[14] A. J. Kunnath, W. Botes, Transforming science education with artificial intelligence: Enhancing inquiry-based learning and critical thinking in south african science classrooms, Eurasia Journal of Mathematics, Science and Technology Education 21 (2025) em2655.

[15] L. M. Guimarães, R. D. S. Lima, A systematic literature review of classroom observation protocols and their adequacy for engineering education in active learning environments, European Journal of Engineering Education 46 (2021) 908–930.
[16] R. Majumdar, A. Kothiyal, Pulse: A framework for protocol based utility to log student engagement, in: 2013 IEEE Fifth International Conference on Technology for Education (t4e 2013), 2013, pp. 159–162. doi:10.1109/T4E.2013.46.