Interactive Workshop: Multimodal, Multiparty Learning Analytics (MMLA)

Peter W. Foltz[1], Gautam Biswas[2], Sidney D'Mello[1], Ekta Sood[1], T.S. Ashwin[2]

[1]University of Colorado Boulder, [2]Vanderbilt University

mmlaworkshop@gmail.com

ABSTRACT

Multimodal, multiparty learning analytics – combining multiple data streams from several participants- has great promise in providing richer insights into learning processes. Yet there are significant technical, measurement and ethical challenges. The objectives of the workshop are to bring together a diverse group of researchers who are researching and developing multimodal learning analytics. Through a series of short presentations and collaborative activities, the workshop participants will share their methodologies, illustrate techniques, demonstrate tools, and discuss challenges and solutions.

Keywords

Multimodal analytics, Multi-party analytics, Data fusion, Ethical data collection, Educational assessment frameworks.

1. INTRODUCTION

The field of multimodal, learning analytics – combining multiple data streams to provide richer insights into learning processes – continues to grow and evolve [1,2,3]. This growth can be attributed to multiple factors including novel computational modeling techniques, technologies for collecting rich streams of learning and interaction data, and the greater emphasis in education on assessing higher order thinking skills. Recently, research has moved beyond modeling individual users to multiparty settings including small groups of students and even entire classrooms (e.g., [4]). In addition, the growth of multimodal foundational large language models (e.g., [5]) has greatly increased the ability to process and fuse multiple modalities of data with reduced needs for training data. While there is great promise for the field, there are significant challenges with detecting and reducing bias in algorithms, maintaining privacy of data, validating performance of models of complex educational interactions, and scaling and implementation of these approaches in the wild. The purpose of this workshop is to bring together researchers investigating new approaches and applications in the nascent field of multimodal, multiparty learning analytics (MMLA). Key **advances** include:

Research on a wide range of modalities including eye-tracking to monitor attention and engagement. facial expression analysis to gauge emotional response, voice analysis to assess verbal participation and comprehension, clickstream data from digital learning environments, physical movement sensors in classroom settings to gauge gestures and pose.

Development of advanced AI analysis techniques. These include deep learning models that can process heterogeneous data streams simultaneously, computer vision algorithms that recognize classroom activities and interactions, natural language processing that analyzes both written and spoken student contributions, temporal analysis methods that track development over time, and multimodal foundational large language models.)

Development of learning environments that enable collection of multiple streams of aligned data and compute multimodal analytics. This can include smart classrooms with multiple sensors and recording capabilities, adaptive learning platforms that adjust content delivery based on multimodal feedback, augmented and virtual reality environments that track multiple physical and cognitive responses, and collaborative workspaces with embedded analytics.

While integrating multiple modalities offers a more complete understanding of behavior and educational outcomes, the field is still young, with a number of challenges to address (e.g., [2,6]). Indeed, while there is much research in the area, few fielded educational systems incorporate multimodal data analytics. Key **challenges** in the field include:

There are technical challenges to implement multimodal analytic systems. Expertise in sensor technologies, digital signal processing, machine learning, and AI is needed to design and implement MMLA systems. Issues in sensor operation, signal synchronization, feature extraction, and multimodal fusion can keep educational teams without access to experts from exploring MMLA solutions. Thus, cross-functional teams are needed to provide expertise in development, evaluation and implementation. Further, shareable MMLA pipelines and analysis tools can broaden participation in research and development while creating more consistency in approaches.

Integrating data from different sources can be difficult, with temporal data alignment and sampling rate issues arising frequently (e.g., [7]). Data collection of multimodal data for large studies can be effortful, often limiting the size of data sets and generalizability of the results (e.g., [6]). Further, human labeling of multimodal data such as videos and audio can be highly time consuming, limiting the signal needed to train models and/or validate performance.

Data sparsity. Recent advances in multimodal machine learning leverage deep multimodal transformers trained on vast volumes of data collected in non-educational settings [3]. However, it is unclear how these approaches can be modified for use with the much smaller volumes of educational data, especially data from youth that may not resemble the datasets used to pretrain the foundational deep models.

There are parallel **methodological challenges**. The absence of uniform methods for validating measurements, fusing multimodal information, and defining constructs and modalities

hinders the generalization, sharing, and reproduction of results (e.g., [1]). The field's novelty and intersection with diverse research traditions contribute to this issue. The lack of consistent methodological approaches and best practices can cause issues in generalizing, reproducing, and sharing results which limits the contribution of MMLA to a common theoretical body of knowledge.

Multiparty integration. In collaborative learning classrooms, multiple groups of individual learners work together on learning activities in coordination with the teacher [8]. Multimodal learning analytics must thereby be extended to encompass multiparty interactions among these students. This can be further extended beyond individual student groups to collections of groups and even to whole class discussions. However, modeling multiple modalities from multiple individuals simultaneously poses several conceptual and methodological challenges to current analytical methods [9,4] that have yet to be resolved.

The capture of information with digital tools raises **privacy concerns and highlights the need for ethical frameworks.** Multimodal data can be quite rich, encompassing video and audio of children as well as interpretations of learning and affective states (e.g.,[10]). Thus, there are strong needs for techniques for data anonymization, consent frameworks for data collection, as well transparent analytic systems that students and teachers can access and understand. Sharing and adhering guidelines for responsible use can help the field move ahead ethically.

2. WORKSHOP OBJECTIVES AND INTENDEND OUTCOMES

2.1 Objectives
The objectives of the workshop are to bring together a diverse group of researchers who are researching and developing multimodal analytics for education and to share their methodology and illustrate techniques on their own data. It is expected that the workshop will help the EDM community by improving researchers' approaches for incorporating multimodal, multiparty learning analytics (MMLA) to provide a more comprehensive understanding of learners working individually, in small groups, and in whole-class settings. The workshop will address key considerations including data collection of different modalities, techniques for integrating multiple data sources from multiple individuals, advances in AI-based analysis techniques, multimodal analytic pipelines, and approaches to addressing data privacy and ethical frameworks.

2.2 Dissemination of outcomes
The outcomes of the workshop will be a set of 2-4 page position and research papers which outline each participant's perspective on multimodal analytics, theoretical background, analytic techniques applied, and tasks and contexts to which the analytics is applied. The organizers will compile the position papers along with a larger summary document that outlines the full space of multimodal analytics, depicting the current state of the field, challenges, and opportunities forward. These papers will be made available through the workshop organizer's webpage at: https://sites.google.com/colorado.edu/edm2025-mmla-workshop

3. ORGANIZATIONAL DETAILS
The interactive workshop will be held as a full-day event. Authors will submit either a 4-page research paper or a 2-page position paper. A research paper could be a description of a particular analytic method, empirical results, or a data set. A position paper argues for a particular position on need in the field.

The workshop will be facilitated by the organizers who have extensive experience in multimodal analytics, each coming from different perspectives. They include backgrounds in development of multimodal AI-based assessment methods, measuring student learning and engagement, assessment frameworks for higher order skills, and modeling complex systems.

3.1 Workshop Schedule
The workshop will consist of an informal poster session, 1-2 keynote addresses, short talks as well as several facilitated discussions to identify 4-6 key topic areas that represent challenges for MMLA moving forward and "big questions" that still need to be addressed.

All participants will reconvene for a final guided discussion session. This session will cover insights that participants gained from the discussions, challenges identified, key techniques identified, and answers to the "big questions". Facilitators will guide this discussion to build a generalized map of challenges and solutions in MMLA.

4. WORKSHOP ORGANIZERS
Peter Foltz is a Research Professor in Cognitive Science at the University of Colorado Boulder. His work focuses on machine learning and natural language processing-based approaches for educational assessments, large-scale data analytics, team collaboration, and 21st Century skills learning. He has led R&D and deployment of large-scale educational technologies in both academia and industry.

Gautam Biswas is a Cornelius Vanderbilt Professor of Engineering, and Professor of Computer Science and Computer Engineering. He researches intelligent open-ended learning environments focused on learning and instruction in STEM domains. He has developed innovative learning analytics and data mining techniques for studying students' learning behaviors and linking them to their metacognitive and self-regulated learning strategies. He is also analyzing multi-modal data from augmented reality training environments to study individual and team performance.

Sidney D'Mello is a Professor in the Institute of Cognitive Science and Department of Computer Science at the University of Colorado Boulder. He is interested in the dynamic interplay between cognition and emotion while individuals and groups engage in complex real-world activities. D'Mello has co-edited seven books and has published more than 300 journal papers, book chapters, and conference proceedings.

Ekta Sood is an AI researcher combining cognitive science and deep learning to create interpretable and user-center AI systems, achieving state-of-art results across NLP and CV tasks. Currently a postdoctoral researcher at the University of Colorado Boulder, she focuses on advancing AI-driven educational technologies and human-AI collaboration.

T. S. Ashwin is a Research Scientist at the Institute for Software Integrated Systems, Vanderbilt University, Nashville, Tennessee. His research focuses on Affective Computing, Learning Technologies, Computer Vision Applications, and Human-Computer Interaction. Dr. Ashwin's work has significantly contributed to video affective content analysis, multi-modal learning analytics, and the application of deep learning in education. He also works in ethics, privacy, and bias in K-12 education vision data.

5. REFERENCES
[1] Ochoa, X., Lang, C., Siemens, G., Wise, A., Gasevic, D., & Merceron, A. (2022). Multimodal learning analytics-

Rationale, process, examples, and direction. In C. Lang, G. Siemens, A. F. Wise, D. Gasevic, & A. Merceron (Eds.), The handbook of learning analytics, Vancouver, BC: SoLAR, 54-65.

[2] Worsley, M., Abrahamson, D., Blikstein, P., Grover, S., Schneider, B., & Tissenbaum, M. (2016). Situating multimodal learning analytics.Worsley, M. and Blikstein, P., (2018). A Multimodal Analysis of Making. International Journal of Artificial Intelligence in Education 28, 3 (Sept. 2018), 385–419. https://doi.org/10.1007/s40593-017-0160-1

[3] Yuan, Y., Li, Z., & Zhao, B. (2025). A Survey of Multimodal Learning: Methods, Applications, and Future. ACM Computing Surveys.

[4] Subburaj, S. K., Stewart, A. E., Ramesh Rao, A., & D'Mello, S. K. (2020). Multimodal, Multiparty Modeling of Collaborative Problem Solving Performance. In Proceedings of the 2020 International Conference on Multimodal Interaction (pp. 423-432). ACM.

[5] Küchemann, S., Avila, K.E., Dinc, Y. et al. On opportunities and challenges of large multimodal foundation models in education. npj Sci. Learn. 10, 11 (2025).

[6] Cohn, C., Davalos, E., Vatral, C., Fonteles, J. H., Wang, H. D., Ma, M., & Biswas, G. (2024). Multimodal Methods for Analyzing Learning and Training Environments: A Systematic Literature Review. arXiv preprint arXiv:2408.14491.

[7] Liu, R. Stamper, J. Davenport, J. Crossley, S., McNamara, D., Nzinga, K., and Sherin, B., (2019). Learning linkages: Integrating data streams of multiple modalities and time-scales. Journal of Computer Assisted Learning 35, 1 (Feb. 2019), 99–109. https://doi.org/10.1111/jcal.12315

[8] Stewart, A. E., Keirn, Z., & D'Mello, S. K. (2021). Multimodal modeling of collaborative problem-solving facets in triads. User Modeling and User-Adapted Interaction, 31(4), 713-751.

[9] Moulder, R., & D'Mello, S. K. (in review). Quantifying Multimodal Dynamics in Groups Containing Members of Mixed-Distinguishability Using Dynamic Bayesian Network Models.

[10] D'Mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. ACM computing surveys (CSUR), 47(3), 1-36.