# Multimodal Learning Analytics Research in the Wild: Challenges and their Potential Solutions

Pankaj Chejara[1], Reet Kasepalu[2], Luis P. Prieto[3], María Jesús Rodríguez-Triana[1], Adolfo Ruiz-Calleja[1] and Shashi Kant Shankar[1]

[1]*School of Digital Technologies, Tallinn University, Tallinn, Estonia*
[2]*School of Education Science, Tallinn University, Tallinn, Estonia*
[3]*Universidad de Valladolid, Valladolid, Spain*

Abstract

Multimodal Learning Analytics (MMLA) has enabled researchers to address learning in physical settings which have long been either overlooked or studied using observational methods. With the use of sensors, researchers have been able to understand learning through an entirely new perspective (e.g., analyzing heart-rate variability to find collaboration indicators). Consequently, MMLA has grown significantly in the past few years, moving from a nascent stage towards a more mature field. It raises a question on how the MMLA researcher can move further, i.e., the transition towards practice which started getting researchers' attention. This paper discusses the challenges we faced while conducting MMLA studies in classroom settings over four years and potential solutions to realize the goal of transitioning MMLA research to educational practice. This paper aims to start a discussion in the field of MMLA over the transition of research to practice.

**Keywords**

Collaboration analytics, Multimodal learning analytics, Machine learning, Collaboration intervention, MMLA

## 1. Introduction

Traditional log-based Learning Analytics (LA) has utilized digital traces of students' activities to understand learning and its context, which only provides a partial picture of a learning situation that occurs across spaces (e.g., collaborative learning in face-to-face settings). To address this issue, researchers have started employing various types of data sources (e.g., audio, video) to complement log-based LA analysis [1]. This research field is known as Multimodal Learning Analytics (MMLA) [2]. MMLA enables researchers to harness the multimodal educational data collected from digital as well as physical spaces. MMLA was proposed for the first time in the workshop held at the ICMI conference in 2012 [3]. Since then the field has made significant progress, resulting in moving from a nascent research field towards a more mature field [4].

Researchers have employed MMLA for studying a range of learning scenarios (e.g., project-based learning [5], collaborative learning [6]) with the use of a wide variety of data sensors (e.g., audio, video, eye-gaze, physiological) [1, 7]. These data sensors have allowed researchers to investigate the physical space of learning, thus enabling a more holistic understanding

*CrossMMLA'23: Leveraging Multimodal Data for Generating Meaningful Feedback March 14, 2023, Arlington Texas, USA*

of learning behavior for researchers as well as teachers. Consequently, the field has gained substantial attention from researchers to understand how learners work together in Computer-Supported Collaborative Learning (CSCL) activities which often happen across spaces (e.g., digital and physical spaces). [5, 7, 8].

MMLA Researchers have explored the relationship between various data features and collaboration constructs (e.g., collaboration quality) [7] and have developed machine learning models to automate the estimation of collaboration constructs. The number of MMLA research studies in authentic settings has grown significantly in the past few years [9]. However, this number is still skewed toward laboratory settings when considering MMLA research in CSCL.

The majority of current MMLA research in CSCL research has been conducted in laboratory settings [10] which allow more controlled settings than classroom settings. However, students may not necessarily behave in a similar way when participating in a learning activity in laboratory and classroom settings [11]. Furthermore, in our previous vignette study where we investigated teachers' trust in the MMLA dashboard for CSCL, teachers indicated trusting the data from the dashboard and proposing a variety of interventions in the imagined classroom [12]. However, we were unable to replicate those findings in authentic settings. These concerns raise the need of conducting MMLA research in classroom settings. However, given its technological complexity, the use/implementation of MMLA in classroom settings is not straightforward.

In this paper, we share our experience of conducting MMLA research in authentic settings and building MMLA tools for supporting teachers [13, 14]. We conducted two iterations of the MMLA study in Estonian high school and vocational school. We collected multimodal educational data to build automated models for collaboration behavior to build a guiding tool to support teachers with intervention strategies during collaborative learning.

## 2. MMLA research studies on collaborative learning

The goal of our research was to build automated models which can estimate collaboration quality and its underlying dimensions (e.g., argumentation) as per Rummel et al. [15] framework, and integrate the developed models in intervention guiding tools to support teachers in the classrooms. We developed our prototype in two iterations (later explained). Table 1 presents the stakeholders involved in both iterations.

**Table 1**
Details on stakeholders involved

| iteration | teachers | students | researchers | schools |
|---|---|---|---|---|
| 1 | 1 | 9 | 2 | 1 |
| 2 | 9 | 174 | 2 | 4 |

### 2.1. Iteration-1: Using a Raspberry Pi-based prototype

In our first iteration (Figure 1), we developed a Raspberry Pi-based prototype to capture the speaking behavior of participants of a group working on a given problem in classroom settings.

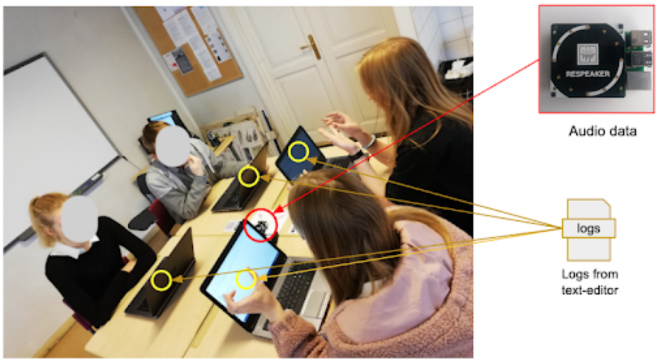


**Figure 1:** MMLA research iteration-1

The prototype was used along with Etherpad[1] which allowed the group's participants to draft solutions to the given problem. We collected multimodal data (audio and logs) from two classroom sessions in an Estonian High school during a collaborative learning activity with the same teacher [6, 8]. There was a total of nine students, 2 researchers, and 1 teacher involved in the study (Table 1). The learning activity was prepared in advance by a researcher and the subject teacher. The data collected from this study was used to study the feasibility of automating collaboration quality estimation [6] and to gain insights into the teacher's perception regarding the utility of MMLA for monitoring collaborative learning in the classroom [12].

### 2.2. Iteration-2: Using CoTrack

In our next iteration (Figure 2), we developed a web-based application called CoTrack[2] to simplify the use of multimodal data collection in f2f and to also make it possible in online learning settings [13, 16]. With the development of CoTrack, we automated all the steps which needed

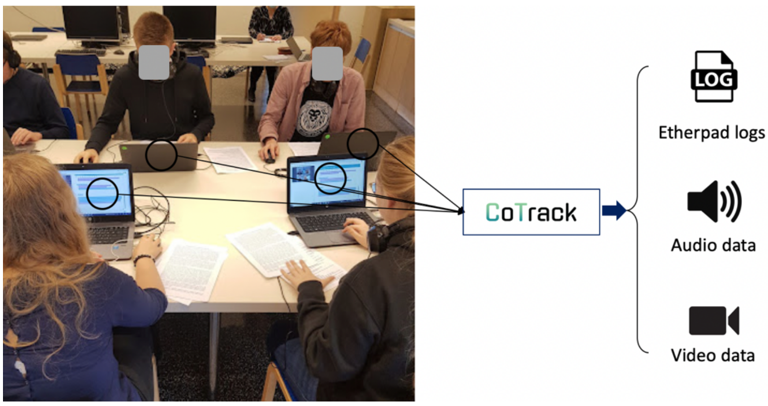


**Figure 2:** MMLA research iteration-2

[1]Etherpad is a real-time collaborative text editor.
[2]https://www.cotrack.website

manual setup in the first prototype (i.e., time synchronization, and feature pre-processing). In our second iteration, the study involved 174 students from 4 different schools, 9 different subject teachers, and 2 researchers (Table 1) [13]. The data from this study was used for developing automated collaboration estimation models for building intervention guiding tools.

# 3. Challenges faced while conducting MMLA research in classroom settings and potential solutions

Here, we outline the main challenges we faced during the enactment of our MMLA research in the classroom setting.

**#1. Technical expertise was needed to set up the first prototype to collect multimodal educational data.**

The first challenge we faced was associated with the technological complexities inherent in MMLA. The developed prototype in the first iteration required a particular setup with a specific set of commands (e.g., running an MQTT server, starting Raspberry Pi devices, setting up an ad-hoc network, and configuring time synchronization). Thus, there was a need for a person with technical expertise whenever the prototype was used, resulting in data collection from fewer classroom sessions.

**Solution:** We addressed this challenge **by automating most of the technological configuration steps**. It resulted in a prototype (iteration-2) with an easy-to-use interface without the need for technical expertise which was needed in the first version of the prototype. The prototype allowed teachers to create collaborative learning activities with multimodal data capturing and monitoring functionality without the need for a technical expert. Another potential solution could be to use a fixed MMLA setup that is already configured in a classroom.

**#2. The prototype used in classroom settings caused obtrusiveness in the learning process.**

The prototype's usage from the first iteration required a preliminary set-up phase which caused interruptions in the classroom. During the study, the setup took an additional time off (10-15 minutes) before starting the learning activity. Moreover, during the activity, the researchers present in the classroom were ensuring the device's functioning which most likely distracted students' attention.

**Solution:** We minimized the obtrusiveness by **using a minimal number of devices** for data collection purposes. We used the laptop's microphone and video camera for data collection purposes.

**#3. Covid-19 restrictions complicated the situation further.**

The arrival of the pandemic Covid-19 put restrictions on physical participation in Estonian schools which closed all possibilities of using the developed prototype (version-1) in classroom settings for multimodal data collection.

**Solution:** We tackled this challenge in our second iteration by transforming our raspberry pi-based prototype into a complete **web-based application**. This feature allowed us to collect more data in our second iteration of the study. Furthermore, this also enabled researchers without technical expertise to use multimodal data-capturing functionality in their research.

**#4. An increased number of requests sent for storing extracted multimodal features**

**exhausted the server's resources.**
As we had integrated Voice Activity Detection (VAD) and Speech-to-Text in CoTrack, it increased the number of requests received on the server. Furthermore, CoTrack also stores media files (e.g., audio and video) that need a larger space on the server. These posed the issue of scalability.
**Solution:** We are now exploring **cloud-based solution** to store our multimodal data and deal with the challenge of server resource exhaustion.

**#5. Students found it difficult to follow MMLA-related instructions along with learning activities.**
Though we simplified the use of CoTrack in comparison with the Raspberry Pi-based prototype, there were still a few instructions for students to follow while using CoTrack. For example, one instruction was to click on the finish button once the learning activity is done which saved the final audio (or video) file on the server. However, during our data collection, we observed many instances when students simply closed the application without clicking on the finish button which caused a loss of audio/video data.
**Solution:** This challenge made us realize the need for the MMLA tool with **minimal reliance on students** for any data collection-related controls. We are now aiming to eliminate this requirement by simply saving media files (audio/video) in the background when students leave the learning activity space in CoTrack.

**#6. Multimodal data collection had a lower quality of data due to the noisy settings of the classroom.**
Multimodal data collected from classroom settings are most likely to be of a lower quality than data collected from a controlled setting. For example, in the case of face-to-face group activity where participants sit close to each other, the chances of audio data getting corrupted become higher.
**Solution:** The data from classroom settings is most likely to have noise, thus, there should be a mechanism to deal with that noise or we should employ data analysis (or modeling) techniques which are noise resistance. In this direction, the next step could be to first understand how the presence of noise in data affects results from analysis or modeling. Another potential idea to deal with noisy data is to exploit cross-modal features. For example, speaking time extracted from audio data can be corrupted as a result of loud noises in the classroom; here, we can extract the same feature also from video using computer vision (e.g., lip-activity detection [17]) and then utilize it in a complementary way.

**#7. Issue of variation in learning contexts on multiple aspects.**
The learning contexts from where the datasets were collected varied on multiple aspects, e.g., schools, subjects, educational level, communication language, teacher, etc. This raised our concern over the applicability of analysis results to a broader context.
**Solution:** To deal with this challenge of context variance, broader data gathering in multiple learning contexts is needed which could further enable a stricter evaluation of the generalizability of analysis results, e.g., across contexts evaluation [8, 13].

**#8. Complexity of multimodal data visualization.**
During our initial phases of prototype development, we involved teachers in the design phase [18]. Teachers reported having freedom in selecting what analytics to see rather than having all the data plotted for them. In addition, they also suggested having some abstract representation for the group's participation or to show collaboration quality levels.

**Solution:** We addressed this challenge by incorporating teachers' suggestions in our current version of CoTrack. We now plan to investigate teachers' perspectives on the usefulness of CoTrack in monitoring and guiding collaborative learning activities in the classroom.

## 4. Discussion

This paper shares our experience of conducting MMLA research in classroom settings in terms of the challenges faced and how we handled them. We also presented some open challenges (e.g., noisy data from the classroom) with some ideas on their potential solution. We like to mention that our solutions are the result of general engineering and are not MMLA-specific. Some solutions (e.g., for #5) can be related to already existing design guidelines for edTech, e.g., minimalism principle for classroom orchestration [19].

In the workshop, how other MMLA researchers have solved challenges in authentic settings could be discussed. The knowledge from the such discussion could result in a manual or tutorial about applying MMLA in authentic settings which we hope can help researchers to conduct MMLA studies in authentic settings.

## References


[1] D. Di Mitri, J. Schneider, M. Specht, H. Drachsler, From signals to knowledge: A conceptual model for multimodal learning analytics, Journal of Computer Assisted Learning 34 (2018) 338–349.

[2] P. Blikstein, Multimodal Learning Analytics and Education Data Mining : Using computational technologies to measure complex learning tasks, Journal of Learning Analytics 3 (2016) 220–238.

[3] S. Scherer, M. Worsley, L.-P. Morency, 1st international workshop on multimodal learning analytics, in: Proceedings of the 14th ACM international conference on Multimodal interaction, 2012, pp. 609–610.

[4] X. Ochoa, C. Lang, G. Siemens, A. Wise, D. Gasevic, A. Merceron, Multimodal learning analytics-rationale, process, examples, and direction, in: The Handbook of Learning Analytics, Soc. Learn. Analytics Res., 2022, pp. 54–65.

[5] D. Spikol, E. Ruffaldi, G. Dabisias, M. Cukurova, Supervised machine learning in multimodal learning analytics for estimating success in project-based learning, Journal of Computer Assisted Learning 34 (2018) 366–377.

[6] P. Chejara, L. P. Prieto, A. Ruiz-Calleja, M. J. Rodríguez-Triana, S. K. Shankar, R. Kasepalu, Quantifying collaboration quality in face-to-face classroom settings using MMLA, in: A. Nolte, C. Alvarez, R. Hishiyama, I. Chounta, M. J. Rodríguez-Triana, T. Inoue (Eds.), Collaboration Technologies and Social Computing - Proceedings of the 26th International Conference, CollabTech 2020, volume 12324 of *Lecture Notes in Computer Science*, Springer, Tartu, Estonia, 2020, pp. 159–166. URL: https://doi.org/10.1007/978-3-030-58157-2_11. doi:10.1007/978-3-030-58157-2_11.

[7] B. Schneider, G. Sung, E. Chng, S. Yang, How can high-frequency sensors capture collab-

oration? A review of the empirical links between multimodal metrics and collaborative constructs, Sensors 21 (2021).
[8] P. Chejara, L. P. Prieto, A. Ruiz-Calleja, M. J. Rodríguez-Triana, S. K. Shankar, R. Kasepalu, Efar-mmla: An evaluation framework to assess and report generalizability of machine learning models in mmla, Sensors 21 (2021) 2863. doi:10.3390/s21082863.
[9] L. Yan, L. Zhao, D. Gasevic, R. Martinez-Maldonado, Scalability, sustainability, and ethicality of multimodal learning analytics, in: LAK22: 12th International Learning Analytics and Knowledge Conference, ACM, New York, NY, USA, 2022, pp. 13–23.
[10] M. Worsley, Multimodal learning analytics' past , present , and , potential futures, in: 2nd Multimodal Learning Analytics Across (Physical and Digital) Spaces, CrossMMLA, 2018, pp. 1–16.
[11] A. Kharrufa, M. Balaam, P. Heslop, D. Leat, P. Dolan, P. Olivier, Tables in the wild: Lessons learned from a large-scale multi-tabletop deployment, Conference on Human Factors in Computing Systems - Proceedings (2013) 1021–1030. doi:10.1145/2470654.2466130.
[12] R. Kasepalu, P. Chejara, L. P. Prieto, T. Ley, Do teachers find dashboards trustworthy, actionable and useful? a vignette study using a logs and audio dashboard, Technology, Knowledge and Learning 27 (2022) 971–989.
[13] P. Chejara, L. P. Prieto, M. J. Rodriguez-Triana, R. Kasepalu, A. Ruiz-Calleja, S. K. Shankar, How to build more generalizable models for collaboration quality? lessons learned from exploring multi-context audio-log datasets using multimodal learning analytics, in: LAK23: 13th International Learning Analytics and Knowledge Conference, LAK2023, ACM, New York, NY, USA, 2023, pp. 111–121.
[14] R. Kasepalu, L. P. Prieto, T. Ley, P. Chejara, Teacher artificial intelligence-supported pedagogical actions in collaborative learning coregulation: A wizard-of-oz study, Frontiers in Education 7 (2022). URL: https://www.frontiersin.org/articles/10.3389/feduc.2022.736194. doi:10.3389/feduc.2022.736194.
[15] N. Rummel, A. Deiglmayr, H. Spada, G. Kahrimanis, N. Avouris, Analyzing collaborative interactions across domains and settings: An adaptable rating scheme, Springer US, Boston, MA, 2011, pp. 367–390.
[16] P. Chejara, L. P. Prieto, M. J. Rodriguez-Triana, A. Ruiz-Calleja, M. Khalil, Impact of window size on the generalizability of collaboration quality estimation models developed using multimodal learning analytics, in: LAK23: 13th International Learning Analytics and Knowledge Conference, LAK2023, ACM, New York, NY, USA, 2023, pp. 559–565.
[17] S. Siatras, N. Nikolaidis, M. Krinidis, I. Pitas, Visual lip activity detection and speaker detection using mouth region intensities, IEEE Transactions on Circuits and Systems for Video Technology 19 (2009) 133–137. doi:10.1109/TCSVT.2008.2009262.
[18] P. Chejara, R. Kasepalu, A. Ruiz-Calleja, M. J. Rodriguez-Triana, L. P. Prieto, Co-designing a multimodal dashboard for collaborative analytics, in: A. Weinberger, W. Chen, D. Hernández-Leo, B. Chen (Eds.), 15th International Conference on Computer Supported Collaborative Learning, CSCL 2022, Hiroshima, Japan, June 6 –10, 2022, International Society of the Learning Sciences, 2022. URL: https://repository.isls.org/bitstream/1/7849/1/577-578.pdf.
[19] P. Dillenbourg, Design for classroom orchestration, Computers & education 69 (2013) 485–492.