

Bringing Collaborative Analytics using Multimodal Data to the Masses: Evaluation and Design Guidelines for Developing a MMLA System for Research and Teaching Practices in CSCL

Pankaj Chejara
pankajch@tlu.ee
Tallinn University
Tallinn, Estonia

Reet Kasepalu
reetkase@tlu.ee
Tallinn University
Tallinn, Estonia

Luis P. Prieto
luisp@tlu.ee
University of Valladolid
Valladolid, Spain

María Jesús Rodríguez-Triana
mjrt@tlu.ee
Tallinn University
Tallinn, Estonia

Adolfo Ruiz-Calleja
adolfo@tlu.ee
Tallinn University
Tallinn, Estonia

ABSTRACT

The Multimodal Learning Analytics (MMLA) research community has significantly grown in the past few years. Researchers in this field have harnessed diverse data collection devices such as eye-trackers, motion sensors, and microphones to capture rich multimodal data about learning. This data, when analyzed, has been proven highly valuable for understanding learning processes across a variety of educational settings. Notwithstanding this progress, an ubiquitous use of MMLA in education is still limited by challenges such as technological complexity, high costs, etc. In this paper, we introduce CoTrack, a MMLA system for capturing the multimodality of a group's interaction in terms of audio, video, and writing logs in online and co-located collaborative learning settings. The system offers a user-friendly interface, designed to cater to the needs of teachers and students without specialized technical expertise. Our usability evaluation with 2 researchers, 2 teachers and 24 students has yielded promising results regarding the system's ease of use. Furthermore, this paper offers design guidelines for the development of more user-friendly MMLA systems. These guidelines have significant implications for the broader aim of making MMLA tools accessible to a wider audience, particularly for *non-expert MMLA users*.

CCS CONCEPTS

• **Applied computing** → **Collaborative learning**; • **Human-centered computing** → **Collaborative and social computing design and evaluation methods**.

KEYWORDS

Multimodal Learning Analytics, MMLA, CSCL

ACM Reference Format:

Pankaj Chejara, Reet Kasepalu, Luis P. Prieto, María Jesús Rodríguez-Triana, and Adolfo Ruiz-Calleja. 2024. Bringing Collaborative Analytics using Multimodal Data to the Masses: Evaluation and Design Guidelines for Developing a MMLA System for Research and Teaching Practices in CSCL. In *LAK24: 14th International Learning Analytics and Knowledge Conference (LAK 2024)*, March 18–22, 2024, Kyoto, Japan. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUCTION

Traditionally, learning analytics (LA) researchers have utilised digital traces obtained from students' interactions with digital tools to gain a better understanding of learning processes and eventually, improve teaching and learning practices. This has resulted in the development of numerous tools and methods to support teachers, e.g., learning analytics dashboards for monitoring [14, 18]. However, this sole focus on digital traces overlooked the physical spaces of learning such as classroom learning happening in face-to-face settings.

Technological advancement has enabled the research community to utilise alternate data sources (e.g., audio or video) to address the aforementioned LA gap. This field has been coined as Multimodal Learning Analytics (MMLA) [3]. The term was introduced at the ICMI conference in 2012 where the first MMLA workshop was organized [24, 34]. The subsequent influential research studies by Blikstein and Worsley [3], Ochoa et al. [20], Schneider and Blikstein [25] have significantly contributed to the progress of this field, which has since experienced substantial growth.

MMLA is defined as the intersection of three ideas: "*multimodal teaching and learning, multimodal data, and computer-supported analysis. At its essence, MMLA utilizes and triangulates among non-traditional as well as traditional forms of data to characterise or model student learning in complex learning environments*" [35]. With this, MMLA extends LA's capabilities by going beyond digital traces of students' activities in order to improve teaching and learning. Researchers have applied MMLA in diverse pedagogical scenarios, including collaborative learning [15], project-based learning [29], and game-based learning [12] with the purpose of advancing our understanding of the learning process, monitoring learning, and providing feedback. A random controlled trial study by Ochoa and Dominguez [21] even demonstrated a positive impact of an MMLA

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions.acm.org.

© 2024 Association for Computing Machinery.
<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

system on students' oral presentation skills after using the system. Furthermore, Olsen et al. [22] showed that the use of multimodal data improves the classification of learning behaviors. Importantly, these benefits extend beyond students and researchers to educators. For instance, Kasepalu et al. [15] found a positive impact of an MMLA system on teaching practices for monitoring group work. These findings, combined with the current state-of-the-art research, provide a comprehensive overview of the potential MMLA holds for both teaching and learning practices as well as research.

Notwithstanding the numerous benefits, the broader utilisation of MMLA does face substantial challenges, in particular among researchers and educators lacking technical expertise. Di Mitri et al. [10] have identified MMLA challenges from a researcher perspective and have grouped them under categories, e.g., data collection, data preprocessing, feedback. Even though the field has grown significantly these core challenges still continue to persist [8].

Delving deeper into the constraints surrounding multimodal data collection, Hassan et al. [13] identified issues such as restricted access to MMLA due to technological complexity. Some of these constraints are also identified by Martinez-Maldonado et al. [19] in their recent study. Their study reported the challenges related to the use of MMLA for data collection, in particular the challenges originating from device installation and orchestration as they, *"increase the complexity of the learning situation from the teachers' perspective"*. As a potential solution, the same study underscored the critical need for an MMLA system designed to be *"easily used by non-experts users"*. The design and development of MMLA systems with a simple interface is an emerging need that has also been identified as a key aspect towards a potentially more ubiquitous MMLA adoption [28].

In this paper, we present *CoTrack*¹, an open source and open access MMLA system designed for teachers and researchers to facilitate the collection of multimodal data, along with real-time visualisation of collaborative learning activities. We present four authentic use cases of the system to illustrate its usefulness for different stakeholders. In addition to this, we share the results of our evaluation of the system's usability. The remainder of this paper is organized into six sections. Section 2 provides a comprehensive review of available MMLA systems. In Section 3, we provide the details of the methodology employed for the design and development of *CoTrack* for multimodal data collection and visualisation. We present the specifics of our system along with use cases in Section 4. Section 5 presents the results of our user evaluation, focusing on the system's usability and ease of use. Section 6 discusses the design guidelines and their implications as well as the limitations of research. Finally, in Section 7, we conclude the paper.

2 MULTIMODAL DATA COLLECTION AND VISUALIZATION SYSTEMS

Multimodal data collection often requires multiple data recording devices (or sensors) for capturing different types of data such as video, audio, or skin conductance response which further provides fine-grained features (e.g., body movement). The use of multiple recording devices often includes setting up a server, establishing communication networks between sensor devices and the server,

and time synchronization. These steps complicate multimodal data collection. To alleviate challenges related to multimodal data collection, several tools have been developed in the past. For example, iMotions Lab² facilitates recordings of multimodal data in a synchronized manner. The tool supports the integration of different streams of data such as eye-gaze, physiological, and affect measures. However, here we are dealing with a commercialised product. Multimodal Learning Hub (MLH) offers an open-access alternate [27]. MLH can be extended to work with any kind of data but requires a configuration setup. The tool is developed in .NET and only works on Windows systems. It has been extended with a visualisation tool (e.g., VIT) for annotating purposes to assist researchers [9]. VIT is reported to be scalable and supports different types of sensors. However, it is unclear how VIT can support data from tools other than MLH.

There are other tools that are not specific to the MMLA research field but could be used for multimodal data collection and processing. For example, SSI (Social Signal Framework) [33] and LSL (Lab Streaming Layer) [17] offer multimodal data collection from a wide range of sensors (including commercialised products) and support time synchronisation. SSI even performs high-level feature extraction from the collected data (e.g., hand gestures). These aforementioned tools fall under conventional multimodal data collection tools using sensors (e.g., eye-gaze tracker). The advancement of web technologies and the availability of browser-based machine learning libraries such as TensorFlow.js³ have enabled the collection of multimodal data on the web. In another case, contrary to the aforementioned tools, the EZ-MMLA toolkit, which is a web-based multimodal data collection tool, employed web technologies to offer capturing diverse data features with the use of only audio and video data [13]. This makes it an ideal candidate for authentic settings where the use of physical sensors might be too obtrusive and could complicate the learning orchestration process for the teacher. However, their system still required a prior setup phase. For example, the collection of speaking time requires the teacher/researcher to either manually start the corresponding tracking service using their website or to ask students to do that. Also, this only allows the capturing of a single type of feature from a particular data stream. Furthermore, as the processed features are stored on the client, it complicates the visualisation and aggregation of this data.

The majority of these aforementioned tools require technical expertise to some extent and often lack an easy-to-use interface. Moreover, these tools are limited to physical spaces and can not be extended to online spaces easily. The EZ-MMLA toolkit addresses this issue to some degree by allowing the use of web technologies. However, the use of this tool in online space still requires some technical expertise from the users to collect multimodal data. Therefore, we see a further need to have an easy-to-use system that: (1) allows simultaneous collection of multimodal data; (2) offers an easy-to-use interface; and (3) enables the adoption of MMLA by potential users without requiring MMLA expertise, such as teachers and researchers. With this in mind, we set up the following research question: **RQ:** How to design and develop an easy-to-use MMLA

¹<https://www.cotrack.website/en/>

²<https://imotions.com/products/imotions-lab/>

³<https://www.tensorflow.org/js>

system for multimodal data collection and visualisation of students' activities during CACL activities?

3 HUMAN-CENTERED DESIGN AND DEVELOPMENT METHODOLOGY

This section presents our methodology employed for the design and development of the proposed system. The presented system has been co-designed and developed over the duration of three years (2020-2022) in three major iterations involving a total of 58 teachers, 603 students and 6 researchers. The following subsections offer a detailed description of those three iterations.

3.0.1 Iteration-1. The first iteration involved interviewing Estonian teachers and the development of a paper [16] and a working prototype [6]. We interviewed 8 in-service Estonian teachers and asked them what kind of information could help them before, during, or after collaborative learning. This step provided us with ideas on the *teachers' preferences*, e.g., the teachers were interested in the individual contribution of the students in each group, their conversation topic, etc. In the next step, we first prepared a paper prototype which then developed into a working prototype using a Raspberry Pi board. The working prototype was equipped with a microphone array (Figure 1a) that detected the direction of the sound, e.g., whenever a student spoke, the prototype detected voice activity in a specific direction which was then used to identify the speaker. However, this version had several *limitations*. Firstly, it required a technical person to set up the devices and start the servers (Etherpad, MQTT) for data collection. Secondly, the use of this prototype was not possible for learning happening in online settings which became a norm during the COVID-19 pandemic.

3.0.2 Iteration-2. In the second iteration, we addressed the aforementioned issues we encountered with the use of the first working prototype. We developed a web-based version, which enabled easier access to the prototype. We also integrated a real-time dashboard for monitoring the groups' activities in the classroom. We used the developed prototype in a workshop to gain feedback from 50 Estonian teachers (32 teachers were in-service English language teachers and 18 were IT teachers). The *feedback* from teachers enabled us to identify new functionalities for the system, e.g., the need to indicate collaboration quality, a feature to duplicate already created learning activities, and automation of group formation (see [5] for more details). Figure 1b shows a group of students during collaborative learning activities using a web-based version. Students accessed the system on their laptops using the Google Chrome browser. There was no additional setup needed to use CoTrack. The web-based version allowed teachers to create group activities with a monitoring functionality without the need for a technical expert. This iteration resulted in the identification of other *limitations*, e.g., high-frequency multimodal data caused the server to perform a high number of disk-write operations to save data; there was a lack of support for stakeholders to download processed multimodal data.

3.0.3 Iteration-3. In the third iteration, we addressed the aforementioned technological issues in addition to integrating the teachers' suggestions with a prediction system to offer an indication of the

groups' collaboration quality. As a part of this iteration, we developed machine learning models with a larger dataset [4, 7], which were collected from two different Estonian schools with the purpose of classifying collaboration quality and its underlying dimensions, e.g., argumentation, knowledge exchange, and sustaining mutual understanding as per Rummel et al. [23]. This iteration helped us to identify the need for a guidance system that can offer some suggestions on the potential intervention strategies.

4 SYSTEM DESCRIPTION

This section explains the technical details of CoTrack and also presents authentic scenarios where the system has been employed.

4.1 Technical Description

This section provides implementation details of the presented MMLA system. *CoTrack* is a web-based application developed using Python's Django web framework. It consists of five main modules: Fetcher, Preprocessing, Storage, Visualisation, and Prediction. The Fetcher serves two primary purposes: firstly, it includes a REST API interface to interact with external applications; secondly, it offers an interface for user interaction. The current version of the system utilises Etherpad, which is a collaborative text editor, to provide a collaborative working space for users. Figure 2a shows the collaborative area in the system. The Fetcher module is also responsible for retrieving multimedia files such as audio and video. The Preprocessing module extracts features from multimodal data. Currently, the system supports speaking activity detection using the Voice Activity Detection algorithm, speech-to-text translation using the Google Speech-To-Text API, and processing of log features obtained from Etherpad. All these features are extracted in real time. The Storage module is responsible for saving multimedia files and preprocessed features. The last two modules, Visualisation and Prediction, leverage the extracted features to generate real-time dashboards of students' activities and to estimate collaboration quality. Figure 2b shows a dashboard generated from a dataset collected from an authentic classroom setting. The dashboard provides insights into group speaking behaviour, displays speech content in the form of a word cloud, and presents writing contributions from each group. In addition, the system also offers the downloading of synchronized and anonymized multimodal data in CSV format for research purposes as well as downloading the responses from all groups into a single file for the teacher.

4.2 Authentic Use Cases

The presented system has been utilised by 6 researchers, 58 teachers, and 603 students. Four researchers were from an educational science background and had no technical expertise. While the rest did have a technical background, they nevertheless did not have expertise and prior experience in multimodal data collection. The system has primarily been employed by researchers for two main purposes: collecting multimodal data for their research studies and studying the responses of stakeholders on the use of MMLA in classrooms. Teachers, in contrast, have made use of the system for either monitoring the students' activities or for demonstrating the potential of MMLA tools to students. The following subsections will introduce two typical use case scenarios where two researchers



Figure 1: Different versions of CoTrack



Figure 2: CoTrack (1: basic details of learning activity; 2: group's speaking behavior in terms of 'who is talking after whom'; 3: controls to check groups' response; 4: each group's contributions in terms of updates made in the editor)

and two teachers used CoTrack for research or teaching purposes within our *current* evaluation study.

4.2.1 Researchers using CoTrack for research purposes. The first researcher (R1) was a Ph.D. student having a computer science background from a Spanish university who used the system for collecting audio data from an authentic classroom setting [31]. The research study was conducted bi-weekly throughout two undergraduate courses on computer networks in 2021. There were a total of 33 students who participated in the study. The goal of the research was to explore students' socially shared regulation of learning in collaborative learning activities. The second researcher (R2) was an educational technology master student with a primary teaching background from an Estonian university. This means that contrary to R1, R2 had a non-technical background. R2 closely interacted with Estonian teachers and collected data from authentic classroom settings using our system in 2022. The goal of the study was to investigate the impact of using MMLA systems during collaborative learning on students' subject knowledge and collaboration skills.

4.2.2 Teachers using CoTrack in their teaching. The first teacher (T1) was from a vocational school in Estonia who used the system in her classrooms for enacting and monitoring group activities in autumn 2023. The teacher had a pedagogical non-technical background. The participants were students, mostly 18-20 years old, enrolled in a software development curriculum. Another teacher

(T2) was from an Austrian university who used the system in his classroom in autumn 2023. The students were enrolled in a master's program in e-education. The goal of using the system was to illustrate an example of utilising (MM)LA tools. Contrary to the first teacher (T1), T2 had a technical background.

5 USER FEEDBACK

To gain an understanding of the usability of CoTrack, we collected the responses from different stakeholders to the System Usability Scale (SUS) survey, which uses a 5-point Likert scale. SUS has been widely accepted and used for evaluating other educational technologies [32]. We collected responses from the teachers (1 female, 1 male) and the researchers (1 female, 1 male) from the aforementioned use cases. Additionally, we also collected responses from teachers and researchers about their overall experience of using the system and their suggestions for further improvement. In addition, we collected responses from 24 students (19 male, 5 female) from the use case where the teacher (T1) employed CoTrack for conducting and monitoring group activities during two lessons. In one lesson the students were asked to complete two activities: the first activity was to go over the class rules from the previous year and analyze them in regards to their functionality and whether there was a need to incorporate some changes into the agreements; the second activity was to plan a group hike day and brainstorm potential group activities for the upcoming school year. In the other lesson,

Table 1: Stakeholder responses to SUS survey (Statements in red color represent the negative statements.)

System usability scale items	Students	Researchers		Teachers	
	Average (std)	R1	R2	T1	T2
I think that I would like to use this system frequently.	3.1 (0.9)	3	4	5	2
I found the system unnecessarily complex.	2.0 (0.8)	1	1	1	3
I thought the system was easy to use.	4.1 (0.6)	4	4	3	3
I think that I would need the support of a technical person to be able to use this system.	1.2 (0.5)	2	1	2	4
I found the various functions in this system were well integrated.	3.7 (0.8)	4	4	5	2
I thought there was too much inconsistency in this system.	1.7 (0.7)	2	1	1	2
I would imagine that most people would learn to use this system very quickly.	4.2 (0.6)	4	3	4	3
I found the system very cumbersome to use.	2.1 (0.8)	2	2	1	3
I felt very confident using the system.	3.9 (0.8)	3	3	3	3
I needed to learn a lot of things before I could get going with this system.	1.3 (0.5)	1	1	5	2
Final SUS score	74.4 (10.8)	75	80	75	47.5

the students were asked to do a vocabulary solidification activity in small groups of three. The students first needed to decide which newly acquired vocabulary was to be used in a gapped sentence and then they needed to discuss how they could use the new vocabulary to write about their personal experiences. In both of the lessons, the students were asked for consent before using the system and T1 introduced the system and informed students about the data it collects. At the end of the activity, the students responded to the SUS survey. It is important to note that the students in the former lesson did already have prior experience using the system while the students from the latter lesson were first-time users of the system.

Table 1 presents the responses of the different stakeholders. We present the average score (standard deviation) for each item of the SUS survey for the students. Due to the small number of responses from teachers and researchers, we do provide responses from each stakeholder. The responses were originally on a 5-point Likert scale, which were used to compute the final SUS score which ranges from 0-100. Bangor et al. [2] specifies SUS scores above 70 as 'Acceptable' and below 50 as 'Unacceptable'. The score between 50-70 was considered as 'Marginally acceptable'. The average SUS score from students was 74.4, therefore it can be deemed as acceptable. Students and researchers strongly perceived the system as easy to use (average rating was above 4 for item "I thought the system was easy to use.") while teachers gave it a moderate rating of 3. In general, all stakeholders except T1 rated lower for the item "I needed to learn a lot of things before I could get going with this system" indicating the minimal amount of effort needed to start using the system. We further analyzed the SUS scores of students for any impact of prior experience of using the system. We noticed that the students' group who had prior experience of using the system, rated higher (SUS score=81) than the students using the system for the first time (SUS score=70). This suggests that as the students use the system more, their perceived usability of the system also increases.

The individual SUS scores from researchers were comparatively higher than the ones from teachers. Overall, the researchers responded positively to the usability and ease of use aspects of the

system and explicitly mentioned the potential of the system. For example, R1 mentioned that "The data collection was done correctly and the students could use the system without any problem." Researchers also highlighted the significance of iterative development of the system, e.g., R2 mentioned that "the system was constantly improved and became more and more convenient." The lowest SUS score representing the 'Unacceptable' level was from one of the teachers (T2). The teacher (T2) said, "It is cool as an example of LA tools, but it is hard to implement in a classroom!". T2 outlined the current limitations in comments and one of those was that the system "does not work with every browser". The current version only supports the Google Chrome browser and this might have caused students to install it on their devices if the browser was not installed before the lesson, which in turn affected the orchestration of the activity. Such feedback from teachers is valuable for further improvement of the system. The lower SUS score from T2 could also be explained with the focus of the system, i.e., addressing the teachers' need to monitor groups' activities and design an easy-to-use multimodal data collection for group work. We did not plan the use of CoTrack for it being an illustrative tool for future learning analytics researchers but interestingly, this case emerged during the development.

6 DISCUSSION

In this paper, we presented *CoTrack*, an MMLA system to collect and visualise multimodal data from CSCL activities. With the presented system, we aimed to address the gap in the lack of an easy-to-use multimodal data collection and visualisation tool that can be used by non-expert users (primarily teachers and students). The use cases presented in the paper illustrate the utility of the system and our usability evaluation provides promising results on the perception of various stakeholders such as students, teachers, and researchers towards usability and ease of use. The SUS evaluation of an MMLA system has only been attempted in a single study in prior research which assessed an MMLA tool (i.e., EZ-MMLA) with only students and reported a SUS score of 71.94 (14.84) [13, 26]. Our evaluation study extends this study by including teachers and researchers. Moreover, the remote accessibility of our system with the use of just

a built-in microphone and camera from the user's device minimizes efforts otherwise needed to set up sensors. Forthwith, this can help in the seamless integration of MMLA capabilities in classroom settings. We would like to remind the reader that the development of the current state of the system has not been an easy process, but rather the product of three major iterations working closely with 58 teachers and 6 researchers following the principles of participatory design. This process also provided a set of guidelines which are presented in the following subsections. For the sake of brevity, we only report MMLA-related guidelines but other good practices for education technology/CSSL also emerged from the process (e.g., automated group formation, reusing learning design). These guidelines can be useful for future MMLA technology designers aiming at designing tools to be used by non-MMLA experts in authentic learning settings.

- **Guideline #1: Minimal reliance on participants for data collection-related instructions** The MMLA solution should be developed with minimal reliance on participants regarding data collection. From our experience, we noticed that the participants found it difficult to follow MMLA instructions (e.g., save all audio/video recordings on the server by clicking a button at the end of the activity). This issue could be resolved by keeping in mind the minimalism principle of classroom technologies design as noted by Dillenbourg and others [11].
- **Guideline #2: Automation of technological configuration** The use of MMLA often involves a configuration stage which is unavoidable. This stage is where stakeholders select the kind of data they want to collect, whether they want to store the raw media files, and what features they want to extract in real-time for visualisation. This is why a simple configuration step needs to be integrated into the system. Admittedly, this poses a "design tension" [30] with guideline #1 which suggests minimal reliance on the user.
- **Guideline #3: Real-time feature computation** The collection of multimodal data raises ethical concerns and has been a topic of increasing interest [1]. This ethical aspect could be partly addressed by integrating a feature computation (and storage) mechanism that extracts features in real-time without storing the raw data, e.g., the video file (which often is more sensitive personal data). This eliminates the need for storing multimodal data. Notwithstanding, this poses threats to the transparency of the models (as we cannot go back to the raw data to see if the machine learning models got something wrong, or why).
- **Guideline #4: Multimodal visualization guidelines** Multiple guidelines emerged during the teachers' use of real-time multimodal data visualization. These guidelines included a *preference of abstract representation* such as sticky figures holding hands to represent equal speaking contribution rather than showing numerical values for speaking time. Another guideline was *to allow the teacher to choose which multimodal data measures* (e.g., speaking time, characters written) to show instead of displaying them all at once (Refer to Anonymized ref-3 for more details).

6.1 Limitations

Aside from the obvious limitations deriving from the system being a research prototype (e.g., device support limited to Chrome browsers), our development process and evaluation study have the following limitations. The first limitation is related to the use of fixed data sources (i.e., audio, video, logs) in the current implementation phase of the system. This limits the use of the system for cases where other types of data sources are needed. The second limitation relates to the scalability of the system. It currently supports a classroom of about 100 students due to its limited server resources. The third limitation is in regard to our usability evaluation study. To explain, we only had students' responses from a single use case and on the whole, had a small sample of teachers and researchers. It needs to be considered that the stakeholders were not MMLA experts, nevertheless, they were from a technical background (e.g., students from an IT curriculum). Therefore, our findings require further investigation with a larger and more varied sample of stakeholders.

7 CONCLUSION & FUTURE WORK

In this paper, we presented a multimodal learning analytics system for capturing multimodal data during group activities and visualising such multimodal data. The system addresses the technical challenge of multimodal data collection by offering an easy-to-use solution targeted at *non-expert MMLA users*. The design and development guidelines that emerged through our iterative design, development, and evaluation process with *authentic* use cases can potentially assist other MMLA researchers and developers. Our initial evaluation results provide evidence of the positive response by different stakeholders (teachers, researchers, students) on the perceived usability and ease of use of the system. Additionally, the aforementioned use cases also highlight the benefits in terms of research and teaching practices. With our system, we extend the capabilities of MMLA for audiences without MMLA expertise by integrating an easy-to-use interface without the need for complex configuration or setup phases (which are currently needed by sensor-based MMLA systems). In view of this extension, our study may promote a wider use of multimodal data by non-technical researchers from educational fields, and result in a greater adoption of MMLA in teaching practice. In our future work, we aim to gain more qualitative insights into the usability aspects of the system and plan to extend the list of data features by employing computer vision-based methods (e.g., lip-based voice activity detection, facial expression, head movement). In addition to this, we also plan to extend CoTrack with an API targeting to simplify the analysis process for researchers.

REFERENCES

- [1] Haifa Alwahaby, Mutlu Cukurova, Zacharoula Papamitsiou, and Michail Gianakos. 2022. The evidence of impact and ethical considerations of Multimodal Learning Analytics: A Systematic Literature Review. *The Multimodal Learning Analytics Handbook* (2022), 289–325.
- [2] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction* 24, 6 (2008), 574–594.
- [3] Paulo Blikstein and Marcelo Worsley. 2016. Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics* 3, 2 (Sep. 2016), 220–238. <https://doi.org/10.18608/jla.2016.32.11>

- [4] Pankaj Chejara, Reet Kasepalu, Luis P. Prieto, María Jesús Rodríguez-Triana, Adolfo Ruiz Calleja, and Bertrand Schneider. 2023. How well do collaboration quality estimation models generalize across authentic school contexts? *British Journal of Educational Technology* 00, 00 (2023), 1–23. <https://doi.org/10.1111/bjet.13402>
- [5] Pankaj Chejara, Reet Kasepalu, Adolfo Ruiz-Calleja, María Jesús Rodríguez-Triana, and Luis Prieto. 2022. Co-Designing a Multimodal Dashboard for Collaborative Analytics. In *Weinberger, A. Chen, W., Hernández-Leo, D., Chen, B. (Eds.), Proceedings of the 15th International Conference on Computer-Supported Collaborative Learning - CSCL 2022*. 577–578. https://repository.isls.org/bitstream/1/8367/1/CSCL2022_577-578.pdf
- [6] Pankaj Chejara, Reet Kasepalu, Shashi Kant Shankar, Luis P Prieto, María Jesús Rodríguez-Triana, and Adolfo Ruiz-Calleja. 2020. Mmla approach to track collaborative behavior in face-to-face blended settings. In *Companion Proceedings LAK20: International Conference on Learning Analytics & Knowledge. Association for Computing Machinery (ACM)*. 543–548. <https://ceur-ws.org/Vol-2610/paper3.pdf>
- [7] Pankaj Chejara, Luis P. Prieto, María Jesús Rodríguez-Triana, Reet Kasepalu, Adolfo Ruiz-Calleja, and Shashi Kant Shankar. 2023. 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* (Arlington, TX, USA) (LAK2023). Association for Computing Machinery, New York, NY, USA, 111–121. <https://doi.org/10.1145/3576050.3576144>
- [8] Mutlu Cukurova, Michail Giannakos, and Roberto Martínez-Maldonado. 2020. The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology* 51, 5 (2020), 1441–1449.
- [9] Daniele Di Mitri, Jan Schneider, Roland Klenke, Marcus Specht, and Hendrik Drachslers. 2019. Read Between the Lines: An Annotation Tool for Multimodal Data for Learning. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (Tempe, AZ, USA) (LAK19). Association for Computing Machinery, New York, NY, USA, 51–60. <https://doi.org/10.1145/3303772.3303776>
- [10] Daniele Di Mitri, Jan Schneider, Marcus Specht, and Hendrik Drachslers. 2018. The Big Five: Addressing Recurrent Multimodal Learning Data Challenges.. In *CrossMMLA@ LAK*.
- [11] Pierre Dillenbourg. 2013. Design for classroom orchestration. *Computers & education* 69 (2013), 485–492.
- [12] Andrew Emerson, Elizabeth B Cloude, Roger Azevedo, and James Lester. 2020. Multimodal learning analytics for game-based learning. *British Journal of Educational Technology* 51, 5 (2020), 1505–1526.
- [13] Javaria Hassan, Jovin Leong, and Bertrand Schneider. 2021. Multimodal Data Collection Made Easy: The EZ-MMLA Toolkit: A Data Collection Website That Provides Educators and Researchers with Easy Access to Multimodal Data Streams.. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (Irvine, CA, USA) (LAK21). Association for Computing Machinery, New York, NY, USA, 579–585.
- [14] Ioana Jivet, Maren Scheffel, Marcus Specht, and Hendrik Drachslers. 2018. License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the 8th international conference on learning analytics and knowledge*. 31–40.
- [15] Reet Kasepalu, Luis P. Prieto, Tobias Ley, and Pankaj Chejara. 2022. Teacher Artificial Intelligence-Supported Pedagogical Actions in Collaborative Learning Coregulation: A Wizard-of-Oz Study. *Frontiers in Education* 7, February (2022), 1–15. <https://doi.org/10.3389/educ.2022.736194>
- [16] R Kasepalu, PL Prieto, and T Ley. 2019. Providing teachers with individual and group-level collaboration analytics: A paper prototype. In *Proceedings of the International Workshop on Collaboration Analytics: Making Learning Visible in Collaborative Settings*. 1–4. <https://collaborationanalytics.files.wordpress.com/2019/06/submission-2-reet.pdf>
- [17] Christian Kothe, David Medine, and Matthew Grivich. 2018. Lab streaming layer (2014). URL: <https://github.com/scen/labstreaminglayer> (visited on 26/02/2020) (2018).
- [18] Katerina Mangaroska and Michail Giannakos. 2018. Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies* 12, 4 (2018), 516–534.
- [19] Roberto Martínez-Maldonado, Vanessa Echeverría, Gloria Fernández-Nieto, Lixiang Yan, Linxuan Zhao, Riordan Alfredo, Xinyu Li, Samantha Dix, Hollie Jaggard, Rosie Wotherspoon, et al. 2023. Lessons learnt from a multimodal learning analytics deployment in-the-wild. *ACM Transactions on Computer-Human Interaction* (2023).
- [20] Xavier Ochoa, Katherine Chiluiza, Gonzalo Méndez, Gonzalo Luzardo, Bruno Guamán, and James Castells. 2013. Expertise estimation based on simple multimodal features. In *Proceedings of the 15th ACM on International conference on multimodal interaction*. 583–590.
- [21] Xavier Ochoa and Federico Dominguez. 2020. Controlled evaluation of a multimodal system to improve oral presentation skills in a real learning setting. *British Journal of Educational Technology* 51, 5 (2020), 1615–1630.
- [22] Jennifer K. Olsen, Kshitij Sharma, Nikol Rummel, and Vincent Aleven. 2020. Temporal analysis of multimodal data to predict collaborative learning outcomes. *Br. J. Educ. Technol.* 51, 5 (2020), 1527–1547. <https://doi.org/10.1111/bjet.12982>
- [23] Nikol Rummel, Anne Deiglmayr, Hans Spada, George Kahrmanis, and Nikolaos Avouris. 2011. Analyzing collaborative interactions across domains and settings: An adaptable rating scheme. In *Analyzing interactions in CSCL*. Springer, 367–390.
- [24] Stefan Scherer, Marcelo Worsley, and Louis-Philippe Morency. 2012. 1st International Workshop on Multimodal Learning Analytics: Extended Abstract. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (Santa Monica, California, USA) (ICMI '12). Association for Computing Machinery, New York, NY, USA, 609–610. <https://doi.org/10.1145/2388676.2388803>
- [25] Bertrand Schneider and Paulo Blikstein. 2015. Unraveling Students' Interaction around a Tangible Interface Using Multimodal Learning Analytics. *Journal of Educational Data Mining* 7, 3 (2015), 89–116.
- [26] Bertrand Schneider, Javaria Hassan, and Gahyun Sung. 2022. Augmenting Social Science Research with Multimodal Data Collection: The EZ-MMLA Toolkit. *Sensors* 22, 2 (2022). <https://doi.org/10.3390/s22020568>
- [27] Jan Schneider, Daniele Di Mitri, Bibeg Limbu, and Hendrik Drachslers. 2018. Multimodal Learning Hub: A Tool for Capturing Customizable Multimodal Learning Experiences. In *Lecture Notes in Computer Science*, Vol. 11082 LNCS. Springer International Publishing, 45–58.
- [28] Shashi Kant Shankar, Luis P. Prieto, María Jesús Rodríguez-Triana, and Adolfo Ruiz-Calleja. 2018. A Review of Multimodal Learning Analytics Architectures. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*. 212–214.
- [29] Daniel Spikol, Emanuele Ruffaldi, Giacomo Dabisias, and Mutlu Cukurova. 2018. Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning* 34, 4 (2018), 366–377.
- [30] Deborah Tatar. 2007. The design tensions framework. *Human-Computer Interaction* 22, 4 (2007), 413–451.
- [31] Cristina Villa-Torrano, Pankaj Chejara, Juan I Asensio-Pérez, Yannis Dimitriadis, and Miguel L Bote-Lorenzo. 2021. Dataset on an online collaborative learning situation in a computer networks course. (2021). <http://ceur-ws.org/Vol-3051/>
- [32] Prokopia Vlachogianni and Nikolaos Tselios. 2021. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education* 0, 0 (2021), 1–18. <https://doi.org/10.1080/15391523.2020.1867938>
- [33] Johannes Wagner, Florian Lingensfelder, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. 2013. The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time. In *Proceedings of the 21st ACM international conference on Multimedia*. 831–834.
- [34] Marcelo Worsley. 2012. Multimodal Learning Analytics: Enabling the Future of Learning through Multimodal Data Analysis and Interfaces. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (Santa Monica, California, USA) (ICMI '12). Association for Computing Machinery, New York, NY, USA, 353–356. <https://doi.org/10.1145/2388676.2388755>
- [35] Marcelo Worsley, Dor Abrahamson, Paulo Blikstein, Shuchi Grover, Bertrand Schneider, and Mike Tissenbaum. 2016. Situating multimodal learning analytics. International Society of the Learning Sciences (ISLS), International Society of the Learning Sciences (ISLS), Singapore.