Multimodal Learning Analytics Device

Central Michigan University

CPS698 Capstone Project

December 10, 2024

Team members: Connor Wiltsie, Sakshyat Sharma, Riad Hossain

Mentor: Dr. Jesse Eickholt

Abstract

The Multimodal Learning Analytics Device is a web-based tool aimed at enhancing the learning experience by integrating audio recording, image capture, transcription, and pdf generation capabilities. Utilizing the Raspberry Pi, the project is targeted to providing students and learners with a user-friendly interface and portable handheld device, for capturing and analyzing multimodal data in an educational setting. The key functionalities include audio transcription, image capturing, and the ability to generate PDFs and email the files securely, enabling learners to document their learning process more effectively. The device's main focus is on privacy, where all the captured data is stored only for the duration of the session, preventing long-term data retention. All the transcriptions and file management occur within the session, and the interface provides clear options for users to manage the data during the session, including the deletion capabilities. The system offers a unique learning solution particularly suitable for sensitive learning environments for users prioritizing data security in their educational technology.

Table of Contents

1. Introduction

Technology today has been an integral part of the learning process. The way students learn and collaborate is driven by the need for personalized and effective learning experiences. This brings up the upfront multimodal learning concept with diverse data sources such as audio, pictures, and texts to help understand and optimize the learning process. Multimodal learning analytics captures these diverse data streams to gain valuable insights into individual learning styles, preferences, and areas needing improvement, holding the immense potential for personalized learning. However, the adoption of multimodal learning analytics faces several challenges. Many of such tools are expensive, and complex to use and also raise significant privacy concerns, regarding the collection and use of sensitive student learning data, which is the main research problem of this project.

This project addresses these challenges by introducing the Multimodal Learning Analytics Device, which is a privacy-centric tool designed for groups of students/learners in different learning scenarios such as in the library, study rooms, classrooms, etc. The device is built upon a Raspberry Pi platform and offers a portable and user-friendly solution for capturing multimodal data in various educational settings, whether it's a classroom or collaborative group study rooms. The device can seamlessly capture images, record audio, transcribe conversations, and generate PDF documentation of each learning session, facilitating a greater understanding of student learning processes.

The project's main commitment lies in student data privacy. The device is designed to store all the captured information only for the duration of the session, and only locally on the device's SD card. Once the session ends, the student can clear out all the data from the device permanently, ensuring that they have complete control over their information. This means students can either use the device and clear all the data after the session is over or bring their own SD card to put it in the device and take it back after the session is over.

Furthermore, the project's commitment to privacy is beyond just data storage. Unlike many other learning tools and devices that require logins or accounts, the Multimodal Learning Analytics Device operates without collecting any personally identifiable information of students such as names, IDs, or login credentials, ensuring complete anonymity. This not only ensures privacy but

also simplifies the user experience making the device much less complex to use. The device's user-friendly interface and portable nature make it adaptable to diverse learning environments, from traditional classrooms to many informal study spaces. This privacy-first and user-friendly approach encourages students to confidently engage with the device, knowing that their data is protected.

This report documents the comprehensive details of the Multimodal Learning Analytics Device project, including its design, components, features, implementation process, and broader impacts, and also explores the potential future developments to further enhance the device's capabilities addressing the emerging challenges in the field of educational technology.

2. Related Works

The integration of Multimodal Learning Analytics (MMLA) has revolutionized educational research. MMLA provides deeper insights into learners' cognitive processes and behaviors by analyzing diverse data modalities such as audio, video, and textual interaction logs. Unlike traditional teaching analytics, which often relies on singular data sources, MMLA fosters personalized learning by enriching the understanding of learner engagement and emotional responses [1].

MMLA's interdisciplinary approach combines technologies like machine learning, sensor-based data collection, and adaptive feedback systems. These techniques enhance classroom analytics by leveraging multiple modalities to create comprehensive and accurate interpretations of learning environments. By fusing text, audio, and video data, MMLA enables the development of tools that can adapt to diverse educational needs, providing instructors and students with actionable insights.

Sensor-based data collection is fundamental to MMLA, and devices like microphones, cameras, and physiological sensors are used to gather environmental and behavioral data. For instance, Crescenzi et al. [10] explored the potential of wearable sensors to assess children's engagement and emotions through non-invasive methodologies such as facial recognition and eye tracking. Similarly, physiological signals, including EEG recordings and heart rate monitoring, have enhanced the analysis of cognitive states and attention levels [11]. Integrated frameworks, such as those proposed by Sharma et al. [13], combine physiological signals with textual data to predict engagement levels accurately, enabling more tailored educational interventions.

Beyond sensor-based methods, interaction data such as clickstream logs, mouse movements, and student responses provide valuable insights into learning processes. Studies by Anjewierden et al. [14] and others have demonstrated how visualizing interaction behaviors can enhance awareness and self-regulation among learners. These methodologies extend MMLA's capabilities by offering a detailed understanding of student engagement with digital platforms.

The fusion of multimodal data is central to MMLA, enabling the integration of diverse inputs to create unified analytical models. Machine learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are frequently employed to extract features and patterns from complex datasets [5, 7]. Transfer learning has further improved the efficiency of MMLA, leveraging pre-trained models to analyze multimodal data even with limited labeled samples [8].

Improving learning outcomes is the ultimate goal of MMLA. Research demonstrates that combining multiple data sources enhances engagement tracking, supports instructors, and improves academic performance.

Engagement is a critical predictor of learning success, and MMLA has been shown to accurately measure engagement through non-verbal cues such as facial expressions, body posture, and hand gestures [19]. For instance, machine learning models have achieved high accuracy in predicting task performance by analyzing these cues, highlighting their potential to foster conducive learning environments.

Another key focus of MMLA is providing instructors with actionable insights. Tools such as IATracer automatically analyze classroom interactions, enabling educators to redesign their teaching strategies based on evidence-driven recommendations [18]. This feedback supports continuous improvement in instructional methods, ultimately benefiting student learning outcomes.

Several studies have linked MMLA to improved academic performance. For example, Purdue University's early warning system significantly improved grades by leveraging data-driven insights to identify at-risk students [20]. These findings underscore the importance of integrating MMLA into institutional frameworks to support student success.

Automation has become a cornerstone of modern MMLA systems, streamlining data collection, annotation, and analysis. Automatic Speech Recognition (ASR) technologies, for example, are widely used to transcribe classroom discussions, enabling scalable analysis of instructional activities [27, 29]. Machine learning algorithms further enhance these capabilities by classifying instructional segments, such as lectures and group discussions, based on audio recordings [32]. However, the over-reliance on audio-based approaches introduces limitations, such as the exclusion of non-verbal cues and potential biases in data interpretation.

Privacy is a critical concern in MMLA, given the sensitive nature of the data collected. Students often lack control over how their data is captured, stored, and analyzed, raising ethical questions about surveillance and consent [33]. Centralized storage of learner data increases the risk of breaches and unauthorized access, while the reliance on automated algorithms may introduce biases that disproportionately affect diverse student populations. Addressing these challenges requires robust safeguards, including data minimization, transparency, and fairness [34, 35].

Modern MMLA systems, including our project, emphasize localized data storage, secure access controls, and user autonomy to balance innovation and ethics. These principles protect student privacy and promote trust and acceptance of educational technologies.

While existing research underscores the transformative potential of MMLA, many solutions remain inaccessible due to their reliance on expensive hardware and proprietary technologies. Additionally, privacy concerns often deter their adoption in real-world educational settings. The Privacy-Centric Multimodal Learning Analytics Device directly addresses these gaps by offering a low-cost, privacy-first solution. By integrating affordable hardware with open-source software and user-friendly interfaces, the project democratizes access to MMLA, enabling its benefits to reach diverse educational contexts.

3. Project Details

3.1 Components and Resources

The Multimodal Learning Analytics Device comprises a selected set of hardware, software, and networking components integrated, each playing a significant role in capturing and processing multimodal data.

- **Hardware**
  - **Raspberry Pi 4:** The Raspberry Pi 4 is responsible for running the operating system, hosting the web application facilitating the user interaction and computational demands of multimodal data processing.
  - **Raspberry Pi Camera Module:** The Raspberry Pi Camera is used to capture high-resolution images of the learning environments. The compact size of the camera makes it compatible with the device, making it an ideal and portable choice for the project.
  - **Conference Microphone:** A high-quality conference microphone is connected to the device to capture the audio interactions during the learning process. The microphone is suitable for capturing clear audio in large conferences or group settings.
  - **Jumper Wires:** They are required for prototyping and connecting all the hardware components.
  - **Micro SD Card:** A micro-SD card is inserted into the device, which is used to store the required OS as well as all the captured media in the device.
  - **Power Supply/Battery:** A battery or a power supply is required to connect and start the device.

- **Software/Libraries/Languages**
  - **Raspbian Operating System:** It is an operating system designed for Raspberry Pi ARM architecture and provides a stable and lightweight foundation for running the device's software components.
  - **Nginx Web Server:** It is used for hosting the device's web application.
  - **Python (Django Framework):** Python is used as the primary programming language for implementing the device's software components. Django, which is the python web

framework is for building the web application and managing the functionalities of the application.

- **Torch Wheel:** It is the distribution of the PyTorch Machine Learning library optimized for Raspberry Pi's ARM architecture. It is used in this project for the efficient execution of the Whisper AI model which is used for transcription.
- **PiCamera2:** It is a Python library that provides an interface for controlling the Raspberry Pi camera module, allowing the device to configure camera settings and capture images.
- **PyAudio:** It is a Python library providing cross-platform audio input and output functionality. It is used by the device to record audio from the microphone data, manage audio streams, and process audio data.
- **Whisper AI:** It is an open-source speech-recognition model by OpenAI, which is used to transcribe the audio into written text. The library provides the python interface for interacting with the Whisper AI model to facilitate audio transcription.
- **ReportLab:** It is a Python library for generating PDFs. It is used by the device to combine the captured images and create a downloadable PDF report.
- **Frontend Development (HTML/CSS/JavaScript):** HTML, CSS, and JavaScript are used as the core technologies for frontend development of the system which includes, web interface, styling, and interactive functionalities for dynamic web content.

- **Network Components:**
  - **Wi-Fi/Ethernet/Hotspot:** Either Wi-fi, Hotspot, or ethernet connectivity could be used for enabling network access and file email transfer.

3.2 System Overview

The flowchart here demonstrates a high-level overview and workflow of the system, highlighting all the main components and features.

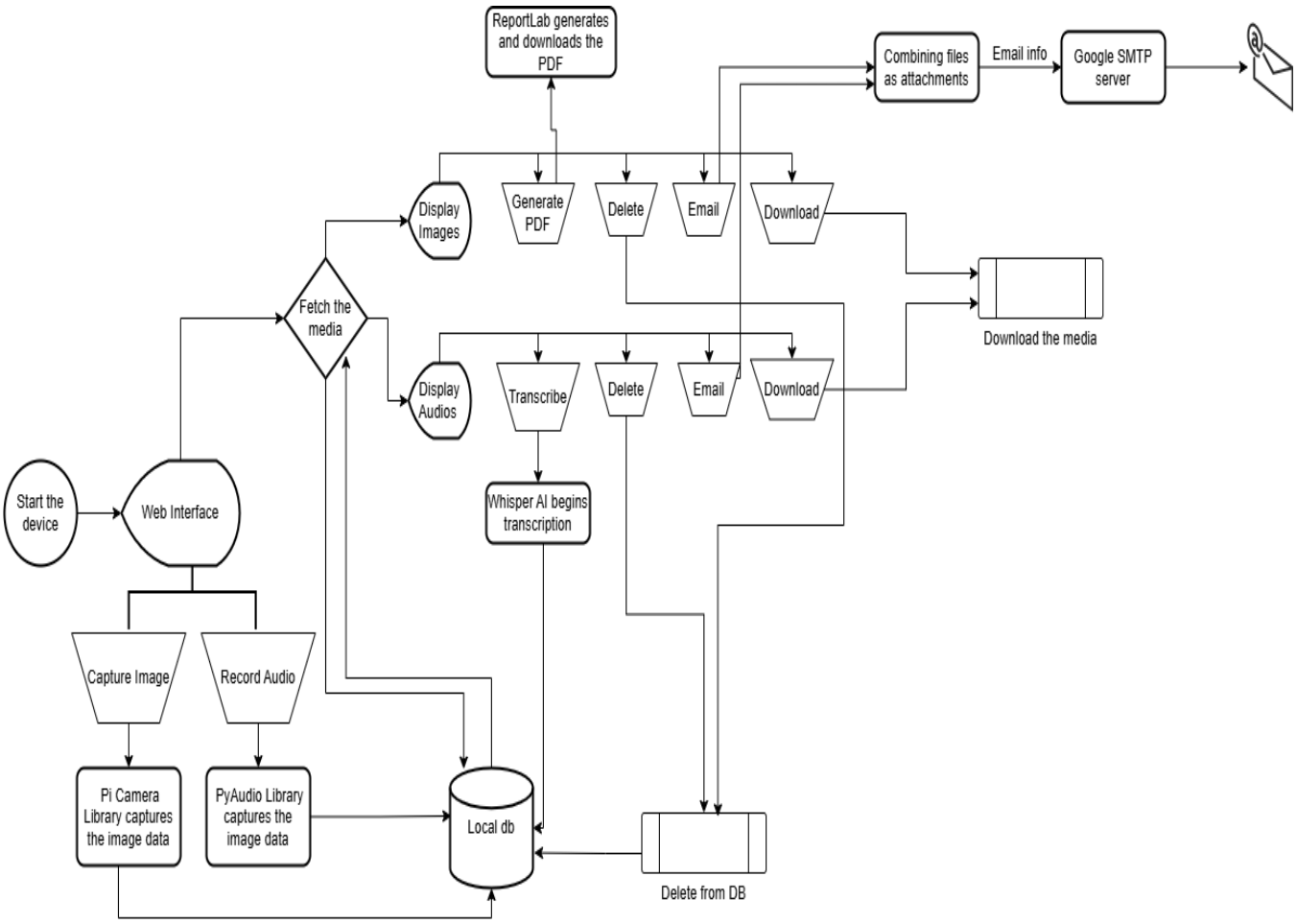

Figure 1 System Overview - Flowchart

3.3 Functionalities: Technical Implementation

The device has a range of features implemented to enhance the learning experience and meet the project's objectives.

- **Image Capturing**

  The feature can be used to capture notes, diagrams, or any other student work relevant to the learning process. The integrated camera is triggered in order to capture the image, using the '*Picamera2'* library. This involves initializing the camera, configuring the settings, and capturing images to a file. A timestamped filename is generated to ensure a unique name for each image captured. The captured images are then saved in *'jpg'* format in the designated media directory, which is in the local device.

- **Audio Recording**

  The function triggers the connected microphone in order to capture the audio interactions during the learning process. *'PyAudio'* library is used for recording the audio. The recording parameters such as *chunk size, format, channels, audio stream* are encapsulated in the *'AudioRecorder'* class. The *'start_recording'* function initiates the audio recording process in a separate thread in order to avoid blocking the main thread. Then, *'stop_recording'* function stops the recording, processes the audio data, and saves it as a timestamped '*WAV*' file.

- **Audio Transcription**

  The device utilizes '*Whisper'* library from OpenAI to transcribe the recorded audio. A pre-trained Whisper model is loaded, which is a general-purpose speech recognition model. Transcription is performed asynchronously in a separate thread allowing the application to remain responsive while the transcription process is underway. Once the transcription process is complete, the file is saved to the designated directory in *'.txt'* format.

- **Media Download and Deletion**

  This function handles the download and the deletion of captured images, audios and their corresponding transcription files. It takes the filename as an input, constructs the file path for that media, and deletes it permanently from the stored directory. For downloading, it checks if

the file exists, and if it exists, it returns it as an *HttpResponse* with an appropriate *Content-Disposition* header for downloading.

- **PDF Generation**

  The function allows the creation of a PDF document containing all the captured images using the *'ReportLab'* library. It iterates through the list of all the captured images and creates a page for each image. The generated PDF is then sent as a downloadable file in an *HTTP* response.

- **Secure File Emailing**

  This feature allows to send the captured media (images, audio recordings and transcriptions) to a specified email address using Django's email functionality, which utilizes Google SMTP for sending emails. It constructs an *'EmailMessage'* object with the recipient's email address, subject, and body, along with the attachments of the files. Then the email is sent to the recipient.

3.4 Project Lifecycle and Maintenance

The project followed an Agile Lifecycle model to ensure iterative development throughout the project, allowing the team to adapt to the changing requirements, incorporate feedback continuously, and deliver the product to meet all the requirements. To facilitate this, Trello was utilized as a project management tool, and boards were created to manage sprints, track progress, and monitor the overall project timeline.

The project involved the following key phases:

- **Project Planning:** This involved defining project goals, scope, and objectives, and identifying the user roles.
- **Sprint Creation/Planning:** This involved breaking down the project into smaller, manageable iterations, selecting functionalities to be completed in each sprint, and estimating and assigning the tasks to the team members along with the timeline for the task to be completed.
- **Development:** This phase involved developing and implementing the defined functionalities in each sprint.

- **Testing and Feedback:** This involved conducting regular code reviews, gathering feedback from the stakeholder(mentor) through demonstrations, incorporating feedback, and making necessary adjustments to ensure quality and identify the issues early.
- **Deployment:** This involved the final demonstration and closure of the project development after meeting all the proposed requirements of the system.

3.5 Acceptance Testing

Throughout the development of the project, testing was a continuous process. Rather than formal test scripts or a separate testing phase, the project went through an approach of iterative testing and continuous integration during the development of the system. This allowed immediate identification and resolution of the issues, ensuring that the components function correctly. The testing process mostly involved:

- Developer Testing: This involved rigorous testing during the development of the system, ensuring each function and module that are implemented work efficiently in isolation, and the interaction between functions and modules were verified during system development.
- User (Mentor) testing/feedback: Throughout the development, the project mentor/stakeholder was involved in testing the device's useability and providing feedback on its functionalities. This user-centered approach ensured that the device met the intended needs and expectations.

4. Conclusion

The Multimodal Learning Analytics Device represents a groundbreaking step in integrating technology into education by addressing key challenges such as affordability, ease of use, and data privacy. By leveraging Raspberry Pi, open-source software, and user-centric design, this system offers a seamless solution for capturing and analyzing multimodal data in diverse learning settings. The device's ability to record audio, capture images, transcribe text, generate PDFs, and securely email files provides a holistic tool for documenting and enhancing the learning process.

The project's privacy-first approach is one of its most significant contributions, ensuring that users maintain complete control over their data. By storing data locally and implementing session-based storage mechanisms, the device alleviates concerns about long-term data retention and unauthorized access. This design empowers students and instructors to engage with the system confidently, fostering trust and adoption in privacy-sensitive environments.

Functionality-wise, the device excels in its simplicity and effectiveness. The integration of Whisper AI for transcription, along with features such as PDF generation and secure emailing, enhances its utility in educational contexts. These features make the device versatile, catering to a wide range of applications, from traditional classrooms to collaborative study rooms and even remote learning setups.

Looking forward, the Multimodal Learning Analytics Device has the potential to evolve further. Future enhancements could include real-time analytics, the integration of additional data modalities such as motion tracking, and the adoption of advanced encryption methods to bolster data security. By continuing to refine its capabilities, the device can adapt to emerging educational needs and technological advancements.

In conclusion, the Multimodal Learning Analytics Device is more than just a tool—it is a catalyst for transforming the way learning processes are understood and improved. Its affordability, privacy-centric design, and comprehensive functionality position as a valuable asset in the educational technology landscape, with the potential to make a lasting impact on learners and educators worldwide.

5. References

[1] L. P. Prieto, K. Sharma, P. Dillenbourg, and M. Jes´us, “Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors,” in Proceedings of the sixth international conference on learning analytics & knowledge, 2016, pp. 148–157.

[2] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, “The faces of engagement: Automatic recognition of student engagementfrom facial expressions,” IEEE Transactions on Affective Computing, vol. 5, no. 1, pp. 86–98, 2014.

[3] H. Li, Y. Kang, W. Ding, S. Yang, S. Yang, G. Y. Huang, and Z. Liu, “Multimodal learning for classroom activity detection,” in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 9234–9238.

[4] T. Ashwin and R. M. R. Guddeti, “Unobtrusive behavioral analysis of students in classroom environment using non-verbal cues,” IEEE Access, vol. 7, pp. 150 693–150 709, 2019.

[5] R. Klein and T. Celik, “The wits intelligent teaching system: Detecting student engagement during lectures using convolutional neural networks,” in 2017 IEEE international conference on image processing (ICIP). IEEE, 2017, pp. 2856–2860.

[6] G. Heigold, I. Moreno, S. Bengio, and N. Shazeer, “End-to-end text-dependent speaker verification,” in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016, pp. 5115–5119.

[7] R. Cosbey, A. Wusterbarth, and B. Hutchinson, “Deep learning for classroom activity detection from audio,” in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 3727–3731.

[8] S. Wang, J. Qu, Y. Zhang, and Y. Zhang, “Multimodal emotion recognition from eeg signals and facial expressions,” IEEE Access, vol. 11, pp. 33 061–33 068, 2023.

[9] G. Varol, I. Laptev, and C. Schmid, “Long-term temporal convolutions for action recognition,” IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 6, pp. 1510–1517, 2017.

[10] L. Crescenzi-Lanna, “Multimodal learning analytics research with young children: A systematic review,” British Journal of Educational Technology, vol. 51, no. 5, pp. 1485–

1504, 2020.

[11] Y. Liu, T. Wang, K. Wang, and Y. Zhang, “Collaborative learning quality classification through physiological synchrony recorded by wearable biosensors,” Frontiers in Psychology, vol. 12, p. 674369, 2021.

[12] J. Gao, P. Li, Z. Chen, and J. Zhang, “A survey on deep learning for multimodal data fusion,” Neural Computation, vol. 32, no. 5, pp. 829–864, 2020.

[13] K. Sharma, Z. Papamitsiou, J. K. Olsen, and M. Giannakos, “Predicting learners’ effortful behaviour in adaptive assessment using multimodal data,” in Proceedings of the tenth international conference on learning analytics & knowledge, 2020, pp. 480–489.

[14] A. Anjewierden, B. Kolloffel, and C. Hulshof, “Towards educational data mining: Using data mining methods for automated chat analysis to understand and support inquiry learning processes,” in International Workshop on Applying Data Mining in e-Learning (ADML 2007), 2007.

[15] K. Huang, T. Bryant, and B. Schneider, “Identifying collaborative learning states using unsupervised machine learning on eye-tracking, physiological and motion sensor data.” International Educational Data Mining Society, 2019.

[16] T. T. Cai and R. Ma, “Theoretical foundations of t-sne for visualizing high-dimensional clustered data,” Journal of Machine Learning Research, vol. 23, no. 301, pp. 1–54, 2022.

[17] S. Mu, M. Cui, and X. Huang, “Multimodal data fusion in learning analytics: A systematic review,” Sensors, vol. 20, no. 23, p. 6856, 2020.

[18] K. Mangaroska, K. Sharma, D. Gaˇsevic, and M. Giannakos, “Multimodal learning analytics to inform learning design: Lessons learned from computing education.” Journal of Learning Analytics, vol. 7, no. 3, pp. 79–97, 2020.

[19] A. Andrade, “Understanding student learning trajectories using multimodal learning analytics within an embodied-interaction learning environment,” in Proceedings of the seventh international learning analytics & knowledge conference, 2017, pp. 70–79.

[20] K. E. Arnold and M. D. Pistilli, “Course signals at purdue: Using learning analytics to increase student success,” in Proceedings of the 2nd international conference on learning analytics and knowledge, 2012, pp. 267–270.

[21] P. J. Donnelly, N. Blanchard, B. Samei, A. M. Olney, X. Sun, B.Ward, S. Kelly, M. Nystran,

and S. K. D'Mello, "Automatic teacher modeling from live classroom audio," in Proceedings of the 2016 conference on user modeling adaptation and personalization, 2016, pp. 45–53.

[22] P. J. Donnelly, N. Blanchard, B. Samei, A. M. Olney, X. Sun, B. Ward, S. Kelly, M. Nystrand, and S. K. D'Mello, "Multi-sensor modeling of teacher instructional segments in live classrooms," in Proceedings of the 18th ACM international conference on multimodal interaction, 2016, pp. 177–184.

[23] J. Eickholt, "Supporting instructor reflection on employed teaching techniques via multimodal instructor analytics," in 2020 IEEE Frontiers in Education Conference (FIE). IEEE, 2020, pp. 1–5.

[24] Z. A. Pardos, R. S. Baker, M. O. San Pedro, S. M. Gowda, and S. M. Gowda, "Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes." Journal of Learning Analytics, vol. 1, no. 1, pp. 107–128, 2014.

[25] S. Teasley, "Student facing dashboards: One size fits all? technology, knowledge and learning, 22 (3), 377–384," 2017.

[26] V. Tinto, Leaving college: Rethinking the causes and cures of student attrition. University of Chicago press, 2012.

[27] N. Blanchard, S. D'Mello, A. M. Olney, and M. Nystrand, "Automatic classification of question & answer discourse segments from teacher's speech in classrooms." International Educational Data Mining Society, 2015.

[28] B. Samei, A. M. Olney, S. Kelly, M. Nystrand, S. D'Mello, N. Blanchard, X. Sun, M. Glaus, and A. Graesser, "Domain independent assessment of dialogic properties of classroom discourse." Grantee Submission, 2014.

[29] S. K. D'Mello, A. M. Olney, N. Blanchard, B. Samei, X. Sun, B. Ward, and S. Kelly, "Multimodal capture of teacher-student interactions for automated dialogic analysis in live classrooms," in Proceedings of the 2015 ACM on international conference on multimodal interaction, 2015, pp. 557–566.

[30] M. Nystrand, "Research on the role of classroom discourse as it affects reading comprehension,"

Research in the Teaching of English, pp. 392–412, 2006.
[31] A. N. Applebee, J. A. Langer, M. Nystrand, and A. Gamoran, “Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school english,” American Educational research journal, vol. 40, no. 3, pp. 685–730, 2003.
[32] M. Nystrand, A. Gamoran, R. Kachur, and C. Prendergast, Opening dialogue. New York: Teachers College Press, 1997.
[33] S. Bian, X. Liu, H. Zhao, X. Gong, and S. Jing, “An immersive learning system with multimodal cognitive processes inference,” in 2022 China Automation Congress (CAC). IEEE, 2022, pp. 6633–6637.
[34] K. Sharma and M. Giannakos, “Multimodal data capabilities for learning: What can multimodal data tell us about learning?” British Journal of Educational Technology, vol. 51, no. 5, pp. 1450–1484, 2020.
[35] M. J. Junokas, R. Lindgren, J. Kang, and J. W. Morphew, “Enhancing multimodal learning through personalized gesture recognition,” Journal of Computer Assisted Learning, vol. 34, no. 4, pp. 350–357, 2018.