JOURNAL OF LEARNING ANALYTICS

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

Volume 12(2), 224–236. https://doi.org/10.18608/jla.2025.8505

The Scope of Multimodal Learning Analytics in K–8: A Systematic Review

Secil Caskurlu[1], Ceren Ocak[2] and Chih-Pu Dai[3]

Abstract

This scoping review aims to provide an overview of how multimodal learning analytics has been applied in K–8 research and offers methodological insights and recommendations to bridge the gap between theory and practice. We identified 14 peer-reviewed empirical studies published between 2011 and 2023 through searches in relevant databases and reviews of journals that publish multimodal learning analytics–related research and publications from leading researchers in the field of multimodal learning analytics. The results revealed that 1) the application of multimodal learning analytics in K–8 research is varied; 2) even though studies explicitly stated how multimodal data were collected and analyzed, they rarely mentioned the data fusion techniques and its process, and ethics and transparency; and 3) multimodal learning analytics has the potential to improve learning outcomes, but more guidance and shared understanding are needed to conduct sound research using multimodal learning analytics. Implications and recommendations are provided.

Notes for Practice

- A small portion of MMLA research has focused on K–8 students. Our scoping review also underscores the continued need for further research in this area.
- Results support that intentional design is essential in MMLA research to make informed decisions about selecting and integrating different modes of interaction that best capture student learning processes.
- Our study suggests that conversations around data ethics, bias, and transparency are lacking, as these issues were frequently reported to be somewhat limited in MMLA studies.
- Transparency in reporting is critical in MMLA to enhance readers' understanding of the data collection, integration/fusion, and analysis processes. Our review also highlights the need for a shared understanding of the data fusion process and reporting guidelines.
- How MMLA supports complex learning processes and informs learning experience design is unclear.
- We propose that MMLA enhances our understanding of complex learning processes; however, a collaboration between researchers with strong computational experience and learning and cognition, as well as educators, is needed to make sense of multimodal data and data analytics, as well as how MMLA informs the learning experience design.

Keywords: Multimodal data, multimodal learning analytics, learning analytics, K–8

Submitted: 06/06/2024 — **Accepted:** 06/03/2025 — **Published:** 10/07/2025

Corresponding author [1]Email: scaskurlu@fsu.edu Address: Department of Educational Psychology and Learning Systems, 3205E Stone Building, Florida State University, Tallahassee, Florida 32304, USA. ORCID iD: https://orcid.org/0000-0001-8716-9586

[2]Email: cocak@georgiasouthern.edu Address: Department of Leadership, Technology, and Human Development, 3110 College of Education Building, Georgia Southern University, Statesboro, Georgia 30458, USA. ORCID iD: https://orcid.org/0000-0002-7231-918X

[3]Email: cdai@hawaii.edu Address: Department of Learning Design and Technology, College of Education, Wist Hall 227, University of Hawai'i at Mānoa, Honolulu, Hawaii 96822, USA. ORCID iD: https://orcid.org/0000-0003-0977-0039

1. Introduction

Multimodal Learning Analytics (MMLA) is a growing field within learning analytics (LA) that seeks to understand how learning can be studied through multiple modalities (Worsley et al., 2021). MMLA, being a specialized form of LA, involves "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens & Gašević, 2012, p. 1). Through a variety of techniques (e.g., sensor technologies and wearable recording devices), MMLA enables researchers to capture fine-grained data from multiple sources such as speech, text, and human action. By combining these analyses, researchers can have a nuanced understanding of how participants make meaning in a given situation (Abrahamson et al., 2022).

A decade of work on MMLA has resulted in diverse research practices. Several scoping and systematic reviews have expanded the breadth of research, each addressing specific themes within the field, such as ethics, bias, trustworthiness, and the application of MMLA in classroom settings (Alwahaby et al., 2022; Crescenzi-Lanna, 2020; Mu et al., 2020; Oviatt, 2018; Prinsloo et al., 2023). Building upon areas identified as under-researched in previous reviews, our study aims to critically analyze how these areas have evolved, focusing on emergent trends, gaps, and limitations. Unlike the reviews mentioned above, our focus is specifically on the application of MMLA in K–8 educational research, responding to a reported need for studies conducted in K–8 settings (Crescenzi-Lanna, 2020; Alwahaby et al., 2022). Therefore, this review identified trends and gaps across various aspects, including data collection tools, choice of data sources (e.g., behavioural, physiological), data analysis methods (e.g., supervised vs. unsupervised machine learning algorithms), outcomes, and ethical considerations within K–8 education. These findings contribute to the growing interest in MMLA and aim to bridge the gap between theory and practice, offering a recent perspective on leveraging MMLA in educational research.

2. Characteristics of Multimodal Learning Analytics

Recent technological advancements enable researchers to capture fine-grained data from multiple sources such as speech, text, and human action through diverse techniques (e.g., sensor technologies and wearable recording devices; Abrahamson et al., 2022). MMLA, therefore, offers researchers a variety of ways to combine those different modes of data to have a more nuanced understanding of a learning phenomenon (Worsley et al., 2021). Approaches such as text and speech analysis can be combined with gesture analysis to understand how participants make meaning in a given situation. Within each of those analyses lie a variety of computational techniques that could be employed to understand how learning is taking place (see Blikstein & Worsley, 2016; Spikol et al., 2017). For example, Spikol et al. (2017) collected high-frequency data across modalities such as face tracking, hand tracking, and interactions with the visual programming interface Arduino to study learners' group interactions. The authors found that the distance between learners' faces and hands is one strong indicator of group success when working on collaborative projects. Emerson et al. (2020) similarly explored how the combination of different multimodal data (MMD) can help predict learner performance and interests during game-based learning (i.e., learning environment: *Crystal Island*). They found that combining facial expressions and gameplay behaviour data increases the predictive power of the models that explain learner performance compared to behaviour-only data, while gaze data decreases the models' performance in models that utilize facial expression and gameplay behaviour data. Overall, these studies suggest how scholars have become increasingly interested in studying learning through various modes of interaction and taking advantage of technology and the latest developments in educational research, such as the use of machine learning to support MMLA.

Over the years, several attempts have been made to consolidate these practices and to characterize the domain. Most recently, Worsley et al. (2021) proposed 12 commitments for conducting high-quality MMLA research, which are grouped into three major categories: 1) data collection, 2) analysis and inference, and 3) feedback and data dissemination. Unlike previous studies that focus on a few prominent aspects of MMLA research (e.g., data fusion, transparency, and ethics), Worsley et al.'s commitments provide a comprehensive guideline for MMLA research to improve the rigour, impact, transparency, and ethics of MMLA studies. Thus, Worsley et al.'s commitments were used to inform this scoping review, serving as an overarching guideline when coding included studies, enabling us to analyze multiple and encompassing characteristics of those studies. Additionally, our review incorporates literature that adds to these commitments, expanding on specific aspects of MMLA research, including data analysis, multimodal data sources, data fusion, and transparency. The following section provides a summary of each commitment area, along with additional literature that supports the findings of Worsley et al. (2021).

2.1. Data Collection

The commitments related to data collection include MMD collection, authenticity of experiences, data privacy, and ethics. Unlike traditional LA and educational data mining (EDM) research, MMLA emphasizes using MMD (e.g., behavioural, physiological) to gain insights into interactions in collaborative learning environments (Alwahaby et al., 2022; Worsley et al., 2021). This, in turn, provides researchers with unique perspectives beyond log data analysis by monitoring student learning behaviours (e.g., collaborative learning behaviours, motivation, engagement) over time to provide support (Emerson et al., 2020) and inform learning experience design (Blikstein & Worsley, 2016; Vujovic et al., 2020; Worsley et al., 2021). To collect in situ, naturally occurring, and emergent data that highlight the authenticity of experiences, MMLA systems should be integrated into authentic, real-classroom settings. However, it is important to note that incorporating MMLA into real classrooms involves distinct research considerations, as classroom interactions play out differently than in laboratories. Thus, decisions such as choosing tools that are minimally invasive, and that benefit classroom participants gain more prominence in authentic settings (Worsley et al., 2021).

Despite the advantages of collecting MMD, significant concerns about ethics and data privacy arise (Worsley et al., 2021). While these concerns apply to any LA system, particularly sensor data carries an added significance due to the personal data it captures (Alwahaby et al., 2022). Moreover, MMD also raises concerns about data surveillance because of the systematic survey of learner behaviours and potential bias embedded in the computational methods used for analysis (Cukurova et al., 2020). Therefore, there is a strong need for more concrete guidelines related to data ethics, privacy, and surveillance during data collection, pre and post-processing, analysis, and interpretation in MMLA for researchers and participants (Alwahaby et al., 2022; Beardsley et al., 2020; Cukurova et al., 2020; Worsley et al., 2021).

2.2. Data Analysis and Inference

These commitments focus on data fusion techniques and data analysis in MMLA research. Data fusion involves combining data from different sources at different stages of data processing to create a unified data set. This remains a key challenge in MMLA due to integrating data of various forms and sizes as well as different approaches to data fusion (Chango et al., 2022; Mu et al., 2020; Worsley, 2014). For example, Mu et al. (2020) broadly categorizes fusion methods into three aspects: data type, learning indicators, and integration methods. Data type includes cross-type and non-cross-type data. Cross-type data combines different kinds of data such as digital and physical whereas non-cross-type data fuses similar types of data. Learning indicators are related to the constructs explored within the research questions — a single indicator (e.g., learning engagement) or multiple indicators simultaneously (e.g., collaboration and learning performance). Finally, integration methods are divided into three categories:

1. Many-to-One, where multiple data forms inform a single learning indicator using a data integration algorithm to enhance measurement accuracy by assigning appropriate weights to different data types, considering their relevance to the learning construct.
2. Many-to-Many, where multiple data forms are used independently to inform multiple learning indicators, providing a comprehensive understanding of the learning process without integrating different data types.
3. Mutual Verification Between MMD, similar to data triangulation, which uses cross-validation to increase confidence in research findings.

Another classification proposed by Worsley (2014) focuses on naïve fusion, low-level fusion, and high-level fusion. Naïve fusion, as the simplest fusion form, involves constructing multiple classifiers using aggregated features from different modalities. In naïve fusion, specific hypotheses are often not set for the study; instead, prior literature and researcher expertise guide the selection of modalities that might inform and complement each other. In low-level fusion, a common technique involves finding the most compatible time scale across different modalities for data integration — for data from each sensor, captured at potentially different scales, can be accurately combined. High-level fusion is a similar approach to low-level fusion, with one critical aspect marking the difference: conducting semantic analysis of the modalities before merging raw data from one or several data streams.

Finally, Chango et al.'s (2022) classification focuses on the timing of fusion, namely early fusion (feature-level), later fusion (decision-level), and hybrid fusion. Early fusion involves fusing the data before the machine learning classifier phase whereas later fusion involves developing a classifier for each data source independently, followed by the fusion of predictions from different classifiers. And hybrid fusion includes both early fusion and later fusion during the fusion process (Chango et al., 2022).

Overall, data fusion requires complex decisions and does not adhere to a one-size-fits-all approach. This process could be facilitated through collaboration between researchers with strong computational experience and a background in learning sciences. Equally important is transparency about data modelling, decision-making, and alignment with research questions and context (Worsley et al., 2021).

As for data analysis in MMLA, the use of supervised and unsupervised machine learning, or a combination of both (i.e., semi-supervised learning), has shaped the current landscape of AI-enhanced MMLA. These machine-learning methods are primarily employed for the classification and prediction of data outcomes. Supervised learning utilizes a training dataset with pre-labelled data to classify new data into predetermined categories or predict continuous outcomes. In contrast, unsupervised learning does not require a pre-labelled training dataset; instead, it finds patterns and structures in the data without predefined labels. Within these broader categories of supervised and unsupervised learning, a variety of algorithms can be applied. For instance, supervised learning includes algorithms such as linear regression, support vector machines, and decision trees. Linear regression represents one of the most basic methods used in supervised learning algorithms for predicting a continuous variable (Boehmke & Greenwell, 2020), while Support Vector Machines (SVM) are mainly used for classification but can also perform regression. Decision trees are used for both classification and regression tasks. Algorithms such as k-means clustering and principal component analysis (PCA) are categorized as unsupervised machine learning models. Algorithms such as random forests and neural networks are typically used in supervised learning; however, they can also be used to perform unsupervised tasks. Random forests, which is an ensemble of decision trees, improve prediction accuracy over using a single decision tree model. Neural networks, which are more complex, excel at tasks that require image classification and speech recognition.

Taken together, in addition to the complexities and different approaches to fuse MMD, analyzing different modalities and making sense of various data streams can also be challenging for researchers as well as other stakeholders (e.g., teachers). To reduce complexity in data inference, researchers can narrow their focus by analyzing specific parts of the task or including research participants in the sense-making process (see Echeverria et al., 2018). Another approach can involve presenting research data and findings to participants, using video playback of their actions and facial expressions to facilitate discussions about their thoughts and feelings during different parts of a learning experience (Worsley et al., 2021). However, asking teachers and learners to make sense of data and data analytics poses significant challenges because of the amount of data and the fact that raw data is usually noisy (Di Mitri et al., 2018).

2.3. Feedback and Data Dissemination

The commitments in this group aim to identify ways to represent data and MMLA findings to learners and other related stakeholders. These commitments highlight the need for providing users with access to data in useful multimodal representations and user-friendly interfaces to facilitate a deeper understanding of the materials and promote transparency (Alwahaby et al., 2022; Worsley et al., 2021) as well as to inform the learning experience design/redesign (Cukurova et al., 2019; Martinez-Maldonado et al., 2020). According to Martinez-Maldonado et al. (2020), most current practices are limited to representing basic descriptive information (e.g., amount of speech) rather than helping teachers and students make sense of complex computational analysis and outputs. One approach could be using data storytelling and data visualizations to present complex MMD to teachers and students to communicate about the key events occurring during the collaborative process (Martinez-Maldonado et al., 2020). Some other approaches use tools such as discussion capture and real-time transcriptions to provide learners with access to valuable resources for revisiting and reviewing elements from their prior learning experiences (Worsley et al., 2021). Taken together, these commitments emphasize the need to leverage multiple sensory channels for delivering feedback beyond traditional written or verbal methods. By incorporating various modalities into the feedback process, MMLA tools can enhance the learning experience and cater to the diverse needs of learners, fostering more engaging and effective learning environments.

3. Purpose Statement

As previously mentioned, prior research has explored various aspects of MMLA, including research trends, applications, data fusion techniques, feedback and dissemination, and ethical considerations. We acknowledge the existence of previous review studies on MMLA, each with a focus on different aspects of MMLA. As detailed in Appendix A, while most studies included all age groups, a few, such as Crescenzi-Lanna (2020), specifically target children under six. Different from these reviews, our study focuses on the application of MMLA in K–8 educational research for several reasons. First, as discussed earlier, MMLA offers many affordances for understanding complex learning processes, which can significantly enhance the design of learning experiences. However, the collection of sensitive and personal data in MMLA studies raises important concerns about data privacy and ethics, especially when involving younger learners (see Crescenzi-Lanna, 2020). Moreover, data privacy and ethics are abstract concepts for K–8 students that can be difficult for younger children to understand and engage with (see Ocak & Caskurlu, 2024). This makes it crucial to carefully consider how data is collected, analyzed, and shared in ways that maintain ethical practices and respect the vulnerabilities of younger learners, while also communicating effectively with teachers and parents or guardians. Second, the cognitive abilities of K–8 students differ significantly from those of high school and college students. As highlighted by Vujovic et al. (2020), younger learners process information and engage with the physical environment differently, which makes it essential to tailor MMLA approaches to student developmental stages. Finally, MMLA outputs can be complex and difficult for educators (Martinez-Maldonado et al., 2020), let alone young students, to interpret and utilize effectively. Therefore, by focusing specifically on K–8, this review contributes to understanding how to adapt MMLA practices to suit the developmental needs and ethical considerations of this age group, ultimately supporting more responsible and effective use of MMLA in K–8 education.

In addition to our K–8 focus, we also focused on studies that explicitly framed their design as MMLA. This focus stems from the importance of intentionality in MMLA research, which informs the selection and combination of different modes of interaction that best capture student learning (Worsley et al., 2021). Our scoping review, therefore, focused on studies with an explicit MMLA methodology, which was different from previous reviews that included both LA and EDM studies. Overall, based on the following research questions, this paper aims to provide an overview of characteristics of MMLA applications for the K–8 context, thereby offering recommendations for leveraging MMLA research and practice:

- How has MMLA been applied in K–8 educational research on data collection, data analysis and inference, and transparency and ethics?
  - What are the common study characteristics, data sources, and collection methods in K–8 MMLA research?
  - What data analysis techniques, including machine learning and data fusion, are employed in K–8 MMLA research?
  - How are ethical considerations, such as transparency and privacy, addressed in K–8 MLA research?

4. Methods

In this paper, we selected a scoping review method. Scoping reviews help identify the scope of the existing body of literature on a phenomenon of interest and offer a systematic approach to 1) identify and map emerging evidence, 2) reveal ways in which research is conducted, and 3) examine the characteristics that establish the boundaries of the field being studied (Munn et al., 2018). Considering the increasing number of MMLA studies conducted with young children and the emerging challenges in fields such as ethics and data fusion — a need highlighted by several scholars since 2020 (e.g., Crescenzi-Lanna, 2020) — we aimed to establish the boundaries and limitations of prior and current MMLA research that are sensitive to the developmental and ethical needs of K–8 students.

4.1. Sampling of Studies

We screened studies published between 2011 and 2023, a time frame that marked the emergence of MMLA as a distinct research field (Worsley et al., 2021). To identify potential studies, we conducted searches in databases such as *Web of Science, ERIC, ProQuest,* and *APA PsycInfo*. We also performed manual searches on Google Scholar and reviewed relevant journals and websites (e.g., Society for Learning Analytics Research, *International Journal of Child–Computer Interaction*, *Journal of the Learning Sciences*, and *British Journal of Educational Technology*). Additionally, we searched for studies by leading scholars in the field of MMLA to identify further studies.

We used the combination of following keywords, chosen based on the research team's expertise and insights from a prior literature review: multimodal learning analytics (i.e., a variety of keywords on MMLA, for example, "multimodal learning analytics," "MMD") and the targeted population (i.e., the education level, for example, "elementary education," "middle school"). Specifically, our database searches employed the following search query: ("multimodal learning analytics" OR "multi-modal learning analytics" OR "MMD" OR "multi-modal data" OR "artificial intelligence" OR "sensory data" OR "learning analytics") AND ("K–8" OR "elementary education" OR "middle school" OR "elementary school").

We retrieved 7,071 records from the initial search and then screened the title and overall relevance. To be included in the review, a study had to 1) be a peer-reviewed journal article and/or conference paper; 2) be empirical; 3) be published from 2011 to December 2023; and 4) include participants from K–8. Finally, we focused on studies that explicitly framed their research as MMLA since intentionality is a key factor that informs the selection and combination of different modes of interaction that best capture student learning (Worsley et al., 2021). The process is detailed in Figure 1.

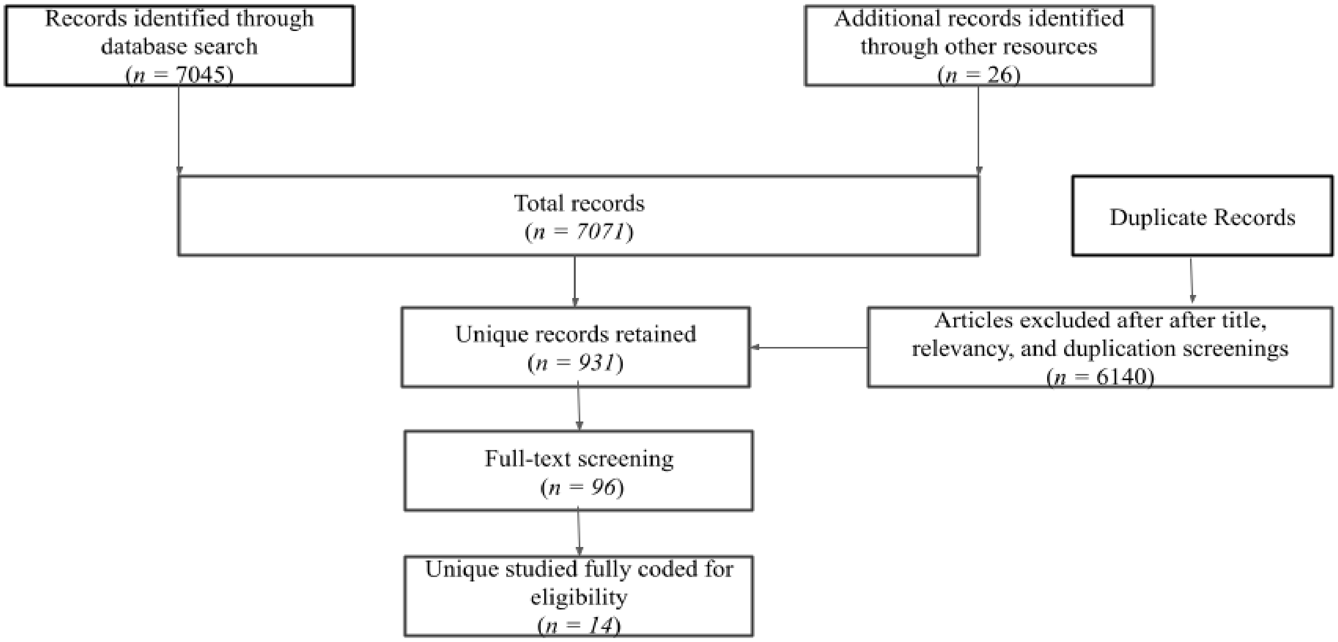

Figure 1. Study identification and selection process.

4.2. Data Extraction and Coding Process

We developed the initial coding form inductively and deductively using a checklist developed by our research team (see Ocak, Caskurlu et al., 2023). The checklist was designed and developed using the following resources: 1) Worsley et al.'s (2021) twelve commitments for grounding MMLA research, 2) Chango et al.'s (2022) multimodal data fusion classification, 3) a comprehensive literature review focused on multimodal data sources and its classification (e.g., Mu et al., 2020), 4) considerations related to transparency and ethics (e.g., Alwahaby et al., 2022; Ouhaichi et al., 2023), and 5) expert opinion.

In our initial checklist, we identified five major areas frequently practised by MMLA researchers: intentionality, theory-drivenness, nature of data, data collection and analysis, and ethics. Following the provided checklist, we developed and iteratively refined the coding form through recursive weekly meetings to discuss the items and ensure their clarity. The final coding categories included study characteristics (e.g., participants, learning environment), multimodal data sources, data collection channels, data analysis, machine learning algorithms, data fusion, and ethics. To note, the final category (i.e., feedback) was excluded due to a lack of information in the included studies. Our initial plan was to include strategies for providing actionable feedback and to explore how researchers offered MMLA feedback to help students and teachers interpret MMLA output effectively. However, most studies lacked this information, so we could not extract it from the included studies (this is discussed further in the discussion section).

To ensure coding reliability, all three authors independently coded the five sampled studies. Subsequently, the third author coded the remaining studies, with the first and second authors reviewing the coding. Discrepancies were discussed and addressed, resulting in a final intercoder reliability of 100%.

5. Results

5.1. Study Characteristics

This section provides an overview of the characteristics of the included studies related to participants, learning environment, and theoretical perspective(s) applied/mentioned (detailed further in Appendix B). Of the 14 studies, 10 were peer-reviewed journal articles, and four were conference papers. Participant demographics included elementary school students (*n*=8), middle school students (*n*=3), and both elementary and middle school students (*n*=3). Regarding learning environments, seven were conducted in a school setting, two were in a laboratory setting, one was a gameplay either in school or home, and one both in a school and museum. Finally, the theoretical frameworks highlighted in the included studies are as follows: embodied cognition (Andrade, 2017; Lee-Cultura et al., 2022; scaffolding, informed by van de Pol et al. (2019; Ouyang et al., 2022); engagement, measured through motivation (Camacho et al., 2020); and cognitive load and multimedia learning (Lee-Cultura et al., 2022.

5.2. Multimodal Data Sources and Collection Channels

As summarized in Table 1, the results showed that various modes were studied, either simultaneously or subsequently. We classified MMD based on where and how data were collected, as adapted from Mu et al. (2020), into the following categories: digital space (e.g., clickstream data), physical space (e.g., body posture, hand motion), physiological space (e.g., skin conductivity, heart rate), and psychometric space (e.g., measures of motivation).

First, log data (*n*=8) was identified as the most common type of data collected in digital spaces. Log data were primarily used to track learner activities and behaviours, such as time spent in the environment, attempts at solutions, and/or errors made (Camacho et al., 2020; Moon et al., 2022; Nasir et al., 2021; Ouyang et al., 2022; Papavlasopoulou et al., 2021; Shin et al., 2020). Log data collection included capturing interactions with computer-based systems and tracking hand gestures on devices (Shin et al., 2020) and recording student behaviours using wearable devices and smart objectives from physiological space (Camacho et al., 2020). Some studies utilized text data to analyze student interactions with tutoring systems (Min et al., 2019; Olsen et al., 2020).

After log data, gaze data (*n*=7) was the next most used data source (Andrade, 2017; Andrade et al., 2016; Lee-Cultura et al., 2022; Min et al., 2019; Moon et al., 2022; Nasir et al., 2021; Olsen et al., 2020). The rationale for collecting gaze data was its suggested link to students' cognitive and affective processes, such as attention and engagement (Min et al., 2019). Particularly with advancements in machine learning, which facilitate the collection and analysis of gaze data, there has been growing interest in this data source. For instance, Andrade (2017) utilized OpenFace, a deep-learning–based facial recognition software, to process eye gaze in recorded videos. Moreover, the tools for collecting gaze data have varied, including eye-tracking devices like SMI Red on desktop computers (Olsen et al., 2020), and more recent innovations such as Tobii eye-tracking glasses (Lee-Cultura et al., 2022).

Additionally, some research explored embodied interactions within physical spaces. For example, Andrade (2017) analyzed hand gestures to investigate student understanding of system dynamics. Furthermore, five studies have utilized reciprocal dialogue and discourse data to examine the interaction patterns of individuals and groups in physical spaces, employing audio or video recording devices for this purpose (Aslan et al., 2022; Nasir et al., 2021; Ocak, Kopcha et al., 2023; Olsen et al., 2020; Ouyang et al., 2022). For capturing physical movement during collaborative activities, Vujovic et al. (2020) employed motion capture systems.

Studies also gathered data in physiological spaces, often employing sensors and wearable devices (*n*=2). For example, Papavlasopoulou et al. (2021) used wristbands to monitor children's physiological arousal during game play. Similarly, Lee-

Cultura et al. (2022) combined wristband with Kinect and other MMD sources to capture play and problem-solving behaviours to explore their implications for learning. Finally, some studies (*n*=3) used pre- and post-test questionnaires (Andrade, 2017; Olsen et al., 2020) and questionnaires (Papavlasopoulou et al., 2021) within the psychometric space. These instruments measured such factors as perceived ease of use, enjoyment, game performance scores, stress levels, physiological arousal, and emotions.

Table 1. Multimodal Data Sources

Study	Log Data	Gaze Data	Embodied Interactions	Sensor Data	Surveys
Andrade (2017)		X	X		X
Andrade et al. (2016)		X			
Aslan et al. (2022)			X		
Camacho et al. (2020)	X				
Lee-Cultura et al. (2022)		X		X	
Min et al. (2019)	X	X			
Moon et al. (2022)	X	X			
Nasir et al. (2021)	X	X	X		
Ocak, Kopcha et al. (2023)			X		
Olsen et al. (2020)	X	X	X		X
Ouyang et al. (2022)	X		X		
Papavlasopoulou et al. (2021)	X			X	X
Shin et al. (2020)	X				
Vujovic et al. (2020)			X		

5.3. Multimodal Data Analysis, Methods and Algorithms

As with MMD sources and collection, the included studies employed a variety of methods for analyzing MMD (see Table 2). Among these, some studies (*n*=4) relied on conventional statistical analysis techniques including descriptive statistics (Aslan et al., 2022; Nasir et al., 2021), ANOVA (Ouyang et al., 2022; Vujovic et al., 2020), Kruskal-Wallis (Nasir et al., 2021), paired sample t-test, correlation (Papavlasopoulou et al., 2021), and computational techniques such as social network analysis and epistemic network analysis (Ouyang et al., 2022).

In addition, the included studies showcased a variety of approaches and algorithms, spanning both supervised and unsupervised machine learning algorithms. Among the supervised models employed in six studies, long short-term memory (LSTM) networks emerged as one of the most frequently utilized models (Min et al., 2019; Olsen et al., 2020). Random forest was another frequently used algorithm. For instance, Lee-Cultura et al. (2022) leveraged random forest algorithms to explore the synergies between various data modalities and student play and problem-solving behaviours. In Moon et al. (2022), after testing multiple classifiers using different algorithms (kNN, random forest, decision tree, and logistic regression), random forest was chosen as a classification model because it resulted in the best performance. Decision trees were another supervised model used to identify and classify student behaviour patterns and engagement (Camacho et al., 2020). Lastly, Ocak, Kopcha, et al. (2023) applied a deep neural network to classify student interactions, captured in video data, into pre-identified labels based on *a priori* theory.

Regarding unsupervised machine learning models (*n*=3), Andrade (2017) and Andrade et al. (2016) utilized hidden Markov models (HMM). Specifically, Andrade (2017) applied HMM to convert raw hands movement data into a vector of latent states. Andrade et al. (2016) also used the 3-cluster HMM algorithm to predict student behavioural frames gaze, gesture, speech, hedging, and frame during an interview. Finally, one study (Shin et al., 2020) applied different machine learning models to compare model performance. They employed logistic regression, support vector machine, decision tree, and random forests to compare them with baseline classifiers generated using synthetic minority over-sampling to distinguish between guessing and solution behaviour. Their results showed that random forest yielded the best performance in detecting guessing behaviours with motion features.

Table 2. Multimodal Data Analysis, Methods, and Algorithms

Study	Conventional Statistical Analysis	Computational Analysis (e.g., ENA, SNA)	Supervised Machine Learning Algorithms	Unsupervised Machine Learning Algorithms
Andrade (2017)				X
Andrade et al. (2016)				X
Aslan et al. (2022)	X			
Camacho et al. (2020)			X	
Lee-Cultura et al. (2022)			X	
Nasir et al. (2021)	X			
Min et al. (2019)			X	
Moon et al. (2022)			X	
Ocak, Kopcha et al. (2023)			X	
Olsen et al. (2020)			X	
Ouyang et al. (2022)	X	X		
Papavlasopoulou et al. (2021)	X			
Shin et al. (2020)				X
Vujovic et al. (2020)	X			

5.4. Data Integration and Fusion

As summarized in Table 3, although not explicitly reported — except in two studies (Moon et al., 2022; Olsen et al., 2020) — we observed the use of data fusion techniques in eight studies. We categorized these studies as employing early fusion, late fusion, or hybrid fusion as proposed by Chango et al. (2022).

Table 3. Data Fusion Techniques

Study		Fusion		
	No Fusion	Early Fusion	Late Fusion	Hybrid Fusion
Andrade (2017)		(O)		
Andrade et al. (2016)		(O)		
Aslan et al. (2022)	(O)			
Camacho et al. (2020)		(O)		
Lee-Cultura et al. (2022)	(O)			
Nasir et al. (2021)		(O)		
Min et al. (2019)		(O)		
Moon et al. (2022)			(R)	
Ocak, Kopcha et al. (2023)		(O)		
Olsen et al. (2020)		(R)		
Ouyang et al. (2022)	(O)			
Papavlasopoulou et al. (2021)	(O)			
Shin et al. (2020)	(O)			
Vujovic et al. (2020)	(O)			

**Observed (O), Reported (R)*

The results showed that seven studies applied early fusion techniques, with one explicitly reported (Olsen et al., 20202) and six observed based on the reported data analysis process. Olsen et al. (2020) combined different modalities (e.g., gaze and audio data) to assess the predictive power of their model in learning gains and post-test scores. We observed that six of the included studies (Andrade, 2017; Andrade et al., 2016; Camacho et al., 2020; Min et al., 2019; Nasir et al., 2021; Ocak, Kopcha et al., 2023) conducted some form of early data fusion with different forms of semantic analysis. For example, Andrade (2017) coded multimodal behaviour in video transcripts, such as body, gaze, gesture, hedging language, and speech, and then transduced those into categorical variables such as “making eye contact” and “looking at the material.” A similar type of fusion was observed in Ocak, Kopcha, et al. (2023), where they classified multimodal interactions in the video into pre-identified categories, such as the use of hands and body to represent numerical values, prior to training a deep neural network to label those interactions. Andrade (2017) employed multiple different techniques to interpret multiple modalities as a whole to address their research questions. Their analyses started with transforming simulation log data into vectors representing student hand movements. The computer tracked and predicted the direction of student hand movements (up, down, or static). They then converted these into vectors representing the student hand movements. This data was then used to identify behavioural clusters based on the motion patterns. Camacho et al. (2020) built a classification algorithm using multiple data streams to provide insights into student motivation and learning as they interact with smart objectives. In another study, Min et al. (2019), combined different data streams to evaluate model performance of three machine learning techniques. Finally, Ouyang et al. (2022) first transcribed classroom video, computer screen recordings, and audio data into Excel files and chronologically organized the data. Organizing the data into time-based segments provided a structured framework for analysis and facilitated temporal alignment. It indicates that the researchers deliberately engaged in the integration of MMD on very fine time scales, which could be considered as an early data fusion.

Only one study, Moon et al. (2022), explicitly reported employing a decision-level fusion approach, where the student’s stress state, their outcome variable, was coded as binary (0 and 1) based on learner emotions, gameplay performance logs, and in-game actions. The authors then utilized a supervised machine learning classifier to fuse the metadata from each data source.

Seven studies (Aslan et al., 2022; Lee-Cultura et al., 2022 Shin et al., 2020; Ouyang et al., 2022; Papavlasopoulou et al., 2021; Ouyang et al., 2022; Vujovic et al., 2020) neither reported nor were observed to use any data fusion technique. However, some of them mentioned data triangulation. For instance, Vujovic et al. (2020) employed aggregate statistical measures (minimum, maximum, mean) to interpret correlations. The MMD from video and motion were kept separate; they were used for triangulation purposes. Likewise, Aslan et al. (2022) collected and analyzed MMD (audio and video) along with conducting observational interviews. Each modality was analyzed separately. Lee-Cultura et al. (2022) collected MMD using eye trackers, wristbands, Kinect joint tracking, and a web camera to measure young children’s task performance. The authors employed many-to-many mappings between the video codes and MMD measurements, suggesting that although no algorithmic integration of data was used, employing multiple types of data concurrently improved the richness of the information gathered.

5.5. Transparency and Ethics

We identified explicit descriptions of transparency and ethics in seven studies that applied various approaches: obtaining approval from ethics committees (*n*=6; Aslan et al., 2022; Lee-Cultura et al., 2022 Ocak, Kopcha et al., 2023; Olsen et al., 2020; Ouyang et al., 2022; Vujovic et al., 2020), making their data available (*n*=2; Nasir et al., 2021; Olsen et al., 2020), sharing the network design including coding (*n*=1; Ocak, Kopcha et al., 2023), and informing participants and their parents/guardians and teachers about the research process (*n*=2), including potential benefits and risks associated with the use of technologies used to collect MMD (Lee-Cultura et al., 2022 Vujovic et al., 2020). Additionally, Lee-Cultura et al. (2022 used the Moral-IT Deck tool to ensure the transparency of sensor data usage.

6. Discussion

This study provided an overview of how MMLA has been applied in K–8 research and offered recommendations about how MMLA can be leveraged in educational research. Regarding the common study characteristics, first, we observed that various modes — gaze, facial expression, discourse, gesture, and writing — have been studied both simultaneously and/or in a time-series fashion to unfold student learning and cognition. These multiple modes of interaction have been collected through diverse channels, allowing MMLA researchers to gather insights from data that is not directly embodied; however, they still provide cues about learning, such as physiological data (e.g., heart rate, skin temperature) or digital trace data. This diversity not only enriches the data available for analysis but also opens new possibilities for understanding and enhancing complex collaborative learning processes. For example, the MMLA studies we reviewed captured data through multiple channels, such as sensor technologies, wearable devices, and video and audio recordings. One notable finding from the included studies is that certain combinations of modalities provide better model performance, depending on the context MMD collected from. For example, Olsen et al. (2020) found that combining various modalities, including gaze data (from eye tracking) and dialogue data (from audio recordings), enhanced the model’s ability to predict learning gains and post-test scores. Olsen et al. (2020) found that integrating audio and eye-tracking data was the most effective combination for prediction, among the combinations of audio-gaze-dialogue, audio-gaze-log, and audio-log. This might stem not only from the multimodal nature of the data but also from the unique insights emerging from the combination of each data stream (Olsen et al., 2020).

Another key finding from our study is the concern regarding the contexts in which MMLA research has been conducted. We observed that the majority of MMLA research has been conducted in laboratory settings, which may not have fully captured the complex interactions in real classrooms. Although lab research can be seen as a way of establishing ground truth or as a preliminary test space for what works in real classrooms, researchers need to be cautious about how they interpret the findings, keeping in mind that they might not fully reflect real practice. This is partially due to complex classroom dynamics that evolve over time, which might influence the outcomes of interventions. In this respect, Worsley et al. (2021) suggest avoiding seeing classrooms as "extensions of the laboratory" (p. 15); in real-world settings, MMLA research should gear towards minimally invasive interventions that align with the natural progression of learning in authentic environments. At the same time, several scholars have already suggested that in the current landscape little is known of the challenges that might arise from applying MMLA research in real-world contexts (Alwahaby et al., 2022). One concern is related to ethics, which plays out very differently in authentic settings compared to laboratories, potentially limiting MMLA's implications for real-world practice. This invites researchers to consider the relevance of the MMLA research to participants' lives. Overall, this suggests that designers need to consider the impact of context in MMLA research, along with the complexity of classroom interactions and ethical issues that might arise from its implementation. Moving forward, there is a clear need for more research conducted in authentic classroom settings to bridge the gap between research and practice (Cukurova et al., 2018) as well as to weigh contextual factors that have an impact on data collection and interpretation (Eradze et al., 2020). We also acknowledge that transitioning MMLA research from the laboratory to real classroom settings introduces additional challenges, including issues with data quality and technological complexities arising from noise, hardware malfunctions, and other difficulties, such as ethics and transparency. Therefore, addressing these challenges, as reported in several studies we reviewed, is crucial for improving the applicability of MMLA research in authentic educational environments (Chejara, 2020; Worsley et al., 2021).

Finally, our results revealed that the limited number of included studies apply theory to make informed decisions (e.g., identifying features, selecting analysis techniques) and interpreting the study results that align with theory and are related to learning practices. Aligning with Giannakos and Cukurova's (2023) findings, we observed that theory in the current MMLA research primarily serves two purposes: informing data collection and offering descriptive insights without explaining how theory is applied in the given research. Therefore, the affordances of MMD provide distinct insights into the learning process, which highlights the need for a more comprehensive integration of theory in future research, thereby leading to a better understanding of complex learning behaviours (Giannakos, 2023; Worsley, 2014). Finally, future research should go beyond MMLA to inform research design and contribute new perspectives to existing theories (Giannakos, 2023; Giannakos & Cukurova, 2023).

Regarding data analysis and data fusion practices, we identified two major approaches to analyzing MMD: conventional statistical approaches and the application of machine learning algorithms (i.e., LSTM networks, decisions trees, HMM, and random forest) for the classification and/or prediction of student outcomes and behaviours. Although there is methodological diversity and the choice of algorithms depends on the nature and objectives of a study, more research is needed to compare the strengths and weaknesses of unsupervised and supervised models and factors that reveal a model's performance in the context of learning and cognition. This would help guide the selection of the most appropriate approach based on the study's purpose. As for data fusion in the analysis phase, most of the included studies did not explicitly address data fusion, including the type of fusion and techniques used for fusing data, which leads to a lack of transparency in explaining their methodological decisions. It is important to note that this does not necessarily mean that these studies did not employ data fusion; rather, they were not explicit about their data fusion process, including the type of fusion and techniques used for fusion of MMD. This is particularly evident in studies that used machine learning algorithms, which often lacked detailed descriptions regarding fusion-related decisions made during data modelling and the communication of findings. Taken together, while data fusion is an essential feature of MMLA research, its implementation remains underexplored and often lacks transparency in methodology reporting, mainly because of the complex decisions required in the process, without a one-size-fits-all approach. Another reason may be that data fusion is still a grey area in MMLA and integrating modes with varying forms and sizes, ensuring clarity in data interpretation, is still challenging for researchers since there is no clarity on how to achieve it (Worsley et al., 2021). Moreover, some studies used data triangulation, instead of data fusion. While Mu et al. (2020) suggested that data triangulation can be considered a form of data fusion, others, such as Wang et al. (2024), have focused on integrating and fusing multiple data sources through various techniques. This goes beyond using different data sources separately to gain deeper insights into the studied phenomenon. Therefore, we need a shared understanding of data fusion and guidelines for reporting practices.

Additionally, ethical considerations, touched upon above, have also emerged as a significant concern in MMLA research. Consistent with previous findings, our results revealed limited to no effort to address the ethical concerns and to provide evidence of the impact on learning outcomes (Alwahaby et al., 2022; Ouhaichi et al., 2023). Moving forward, researchers must prioritize ethical guidelines and ensure transparency in their data handling practices, including how the data is managed, stored, and accessed (Worsley et al., 2021, p. 15). In addition, journals that publish MMLA research should continue to encourage researchers to engage in open science practices, as evident in Moon et al. (2022), which adapted publicly available data for their emotion recognition model.

Finally, we observed that the ways in which MMLA supports learning processes and informs learning experience design are still unclear, a finding consistent with others (Di Mitri et al., 2018). Although attempts have been made to understand MMD, there is a lack of clarity on how this data can be effectively used to support learning processes and inform learning experience design. For example, Vujovic et al. (2020) explored the effects of the physical environment on learner behaviour during collaborative problem tasks. The findings suggest that round tables positively impact the behaviour of elementary school students during group learning activities, which lead to higher levels of engagement and participation. There is a need for more studies like this, providing actionable feedback to both students and teachers. This remains a critical area for future research to enhance student learning experiences and learning outcomes (Martinez-Maldonado et al., 2020).

7. Conclusion

In conclusion, while MMLA research has made significant advancements in understanding the role of MMD in complex learning processes in K–8 education, there are still several challenges and areas for improvement, including data fusion, transparency in reporting, ethics, theory-drivenness, and practical implications. There are also limitations in this scoping review. First, we excluded numerous LA and EDM studies due to their lack of explicit focus on MMLA. These excluded papers might provide additional insights into MMD's role in explaining complex learning processes. Second, as our primary aim was to offer an overview of MMLA's application in K–8 educational research, we did not conduct a quality appraisal of studies that might include low-quality papers in the review. Future research could focus on developing criteria for assessing the quality of MMLA research. Moving forward, researchers and practitioners can benefit from our findings regarding the issues and challenges identified (e.g., data fusion, ethics, theory-drivenness, and practical implications) and contribute to the meaningful integration of MMLA into educational practice. Practitioners can use the review findings to make more informed decisions when integrating MMLA into learning experience design, thereby enhancing educational outcomes.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors declared no financial support for the research, authorship, and/or publication of this article.

References

Abrahamson, D., Worsley, M., Pardos, Z. A., & Ou, L. (2022). Learning analytics of embodied design: Enhancing synergy. *International Journal of Child–Computer Interaction*, Article 100409. https://doi.org/10.1016/j.ijcci.2021.100409

Alwahaby, H., Cukurova, M., Papamitsiou, Z., & Giannakos, M. (2022). The evidence of impact and ethical considerations of multimodal learning analytics: A systematic literature review. In M. Giannakos, D. Spikol, D. Di Mitri, K. Sharma, X. Ochoa, & R. Hammad (Eds.), *The multimodal learning analytics handbook* (pp. 289–325). Springer. https://doi.org/10.1007/978-3-031-08076-0_12

*Andrade, A. (2017). Understanding student learning trajectories using multimodal learning analytics within an embodied-interaction learning environment. In I. Molenaar, X. Ochoa, & S. Dawson (Eds.), *Understanding, informing and improving learning with data: LAK'17 conference proceedings: The seventh international learning analytics & knowledge conference* (pp. 70–79). ACM Press. http://dx.doi.org/10.1145/3027385.3027429

*Andrade, A., Delandshere, G., & Danish, J. A. (2016). Using multimodal learning analytics to model student behaviour: A systematic analysis of epistemological framing. *Journal of Learning Analytics*, *3*(2), 282–306. http://dx.doi.org/10.18608/jla.2016.32.14

*Aslan, S., Agrawal, A., Alyuz, N., Chierichetti, R., Durham, L. M., Manuvinakurike, R., Okur, E., Sahay, S., Sharma, S., Sherry, J., Raffa, G., & Nachman, L. (2022). Exploring kid space in the wild: A preliminary study of multimodal and immersive collaborative play-based learning experiences. *Educational Technology Research and Development*, *70*(1), 205–230. https://doi.org/10.1007/s11423-021-10072-x

Boehmke, B., & Greenwell, B. M. (2020). *Hands-on machine learning with R*. CRC Press. https://bradleyboehmke.github.io/HOML/

Beardsley, M., Martínez Moreno, J., Vujovic, M., Santos, P., & Hernández-Leo, D. (2020). Enhancing consent forms to support participant decision making in multimodal learning data research. *British Journal of Educational Technology*, *51*(5), 1631–1652. https://doi.org/10.1111/bjet.12983

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238. http://dx.doi.org/10.18608/jla.2016.32.11

*Camacho, V. L., de la Guía, E., Olivares, T., Flores, M. J., & Orozco-Barbosa, L. (2020). Data capture and multimodal learning analytics focused on engagement with a new wearable IoT approach. *IEEE Transactions on Learning Technologies*, *13*(4), 704–717. https://doi.org/10.1109/TLT.2020.2999787

Chango, W., Lara, J. A., Cerezo, R., & Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. *WIREs Data Mining and Knowledge Discovery*, *12*(4), Article e1458. https://doi.org/10.1002/widm.1458

Chejara, P. (2020). Multimodal learning analytics for understanding collocated collaboration in authentic classroom settings. In M. Chang, D. G. Sampson, R. Huang, D. Hooshyar, N.-S. Chen, Kinshuk, & M. Pedaste (Eds.), *Proceedings: IEEE 20th international conference on advanced learning technologies: ICALT 2020* (pp. 390–392). IEEE. https://doi.org/10.1109/ICALT49669.2020.00123

Crescenzi-Lanna, L. (2020). Multimodal learning analytics research with young children: A systematic review. *British Journal of Educational Technology*, *51*(5), 1485–1504. https://doi.org/10.1111/bjet.12959

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology*, *51*(5), 1441–1449. https://doi.org/10.1111/bjet.13015

Cukurova, M., Kent, C., & Luckin, R. (2019). Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring. *British Journal of Educational Technology*, *50*(6), 3032–3046 https://doi.org/10.1111/bjet.12829

Cukurova, M., Luckin, R., & Baines, E. (2018). The significance of context for the emergence and implementation of research evidence: The case of collaborative problem-solving. *Oxford Review of Education*, *44*(3), 322–337. https://doi.org/10.1080/03054985.2017.1389713

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 338–349. https://doi.org/10.1111/jcal.12288

Echeverria, V., Martinez-Maldonado, R., Buckingham Shum, S., Chiluiza, K., Granda, R., & Conati, C. (2018). Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling. *Journal of Learning Analytics*, *5*(3), 72–97. http://dx.doi.org/10.18608/jla.2018.53.6

Emerson, A., Cloude, E. B., Azevedo, R., & Lester, J. (2020). Multimodal learning analytics for game-based learning. *British Journal of Educational Technology*, *51*(5), 1505–1526. https://doi.org/10.1111/bjet.12992

Eradze, M., Rodríguez Triana, M. J., & Laanpere, M. (2020). Context-aware multimodal learning analytics taxonomy. In M. Scheffel, V. Kovanović, N. Pinkwart, & K. Verbert (Eds.), *Shaping the future of the field: Companion proceedings of 10th international conference learning analytics & knowledge LAK20* (pp. 533–538). SoLAR. https://www.solaresearch.org/wp-content/uploads/2020/06/LAK20_Companion_Proceedings.pdf

Giannakos, M. (2023). Theory and intermediate-level knowledge in multimodal learning analytics. In K. Bartimote, S. K. Howard, & D. Gašević (Eds.), *Theory informing and arising from learning analytics* (pp. 87–104). Springer. https://doi.org/10.1007/978-3-031-60571-0_6

Giannakos, M., & Cukurova, M. (2023). The role of learning theory in multimodal learning analytics. *British Journal of Educational Technology*, *54*(5), 1246–1267. https://doi.org/10.1111/bjet.13320

*Lee-Cultura, S., Sharma, K., & Giannakos, M. (2022). Children's play and problem-solving in motion-based learning technologies using a multi-modal mixed methods approach. *International Journal of Child–Computer Interaction*, *31*, Article 100355. https://doi.org/10.1016/j.ijcci.2021.100355

Martinez-Maldonado, R., Echeverria, V., Fernandez Nieto, G., & Buckingham Shum, S. (2020). From data to insights: A layered storytelling approach for multimodal learning analytics. In R. Bernhaupt, F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. Zhao, B. P. Samson, & R. Kocielnik (Eds.), *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–15). ACM Press. https://doi.org/10.1145/3313831.3376148

*Min, W., Park, K., Wiggins, J., Mott, B., Wiebe, E., Boyer, K. E., & Lester, J. (2019). Predicting dialogue breakdown in conversational pedagogical agents with multimodal LSTMs. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial intelligence in education: 20th international conference, AIED 2019, Chicago, IL, USA, June 25–29, 2019, proceedings, part II* (pp. 195–200). Springer. https://doi.org/10.1007/978-3-030-23207-8_37

*Moon, J., Ke, F., Sokolikj, Z., & Dahlstrom-Hakki, I. (2022). Multimodal data fusion to track students' distress during educational gameplay. *Journal of Learning Analytics*, *9*(3), 75–87. https://doi.org/10.18608/jla.2022.7631

Mu, S., Cui, M., & Huang, X. (2020). Multimodal data fusion in learning analytics: A systematic review. *Sensors*, *20*(23), Article 6856. https://doi.org/10.3390/s20236856

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(143). https://doi.org/10.1186/s12874-018-0611-x

*Nasir, J., Kothiyal, A., Bruno, B., & Dillenbourg, P. (2021). Many are the ways to learn identifying multi-modal behavioural profiles of collaborative learning in constructivist activities. *International Journal of Computer-Supported Collaborative Learning*, *16*(4), 485–523. https://doi.org/10.1007/s11412-021-09358-2

Ocak, C., & Caskurlu, S. (2024, April 14). *In-service teachers' perception of developing a critical lens toward technology integration* [Roundtable presentation]. AERA Annual Meeting, Philadelphia, PA, United States.

Ocak, C., Caskurlu, S., & Dai, C.-P. (2023, October 15–19). *Developing a checklist as a guide to conduct multimodal learning analytics research* [Paper presentation]. AECT International Convention, Orlando, FL, United States.

*Ocak, C., Kopcha, T. J., & Dey, R. (2023). An AI-enhanced pattern recognition approach to temporal and spatial analysis of children's embodied interactions. *Computers and Education: Artificial Intelligence*, Article 100146. https://doi.org/10.1016/j.caeai.2023.100146

*Olsen, J. K., Sharma, K., Rummel, N., & Aleven, V. (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology*, *51*(5), 1527–1547. https://doi.org/10.1111/bjet.12982

Ouhaichi, H., Spikol, D., & Vogel, B. (2023). Research trends in multimodal learning analytics: A systematic mapping study. *Computers and Education: Artificial Intelligence*, *4*, Article 100136. https://doi.org/10.1016/j.caeai.2023.100136

*Ouyang, F., Dai, X., & Chen, S. (2022). Applying multimodal learning analytics to examine the immediate and delayed effects of instructor scaffoldings on small groups' collaborative programming. *International Journal of STEM Education*, *9*, Article 45. https://doi.org/10.1186/s40594-022-00361-z

Oviatt, S. (2018). Ten opportunities and challenges for advancing student-centered multimodal learning analytics. In S. K. D'Mello, P. Georgiou, S. Scherer, E. M. Provost, M. Soleymani, & M. Worsley (Eds.), *ICMI'18: Proceedings of the 20th ACM International Conference on Multimodal Interaction* (pp. 87–94). ACM Press. https://doi.org/10.1145/3242969.3243010

*Papavlasopoulou, S., Sharma, K., Melhart, D., Schellekens, J., Lee-Cultura, S., Giannakos, M. N., & Yiannakakis, G. N. (2021). Investigating gaze interaction to support children's gameplay. *International Journal of Child–Computer Interaction*, *30*, Article 100349. https://doi.org/10.1016/j.ijcci.2021.100349

Prinsloo, P., Slade, S., & Khalil, M. (2023). Multimodal learning analytics—In-between student privacy and encroachment: A systematic review. *British Journal of Educational Technology*, *54*(6), 1566–1586. https://doi.org/10.1111/bjet.13373

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, *51*(5), 1450–1484. https://doi.org/10.1111/bjet.12993

*Shin, H., Kim, B., & Gweon, G. (2020). Guessing or solving? Exploring the use of motion features from educational game logs. In R. Bernhaupt, F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. Zhao, B. P. Samson, & R. Kocielnik (Eds.), *CHI'20: Extended abstracts of the 2020 CHI conference on human factors in computing systems* (LBW344). ACM Press. https://doi.org/10.1145/3334480.3383005

Siemens, G., & Gašević, D. (2012). Guest editorial: Learning and knowledge analytics. *Journal of Educational Technology & Society*, *15*(3), 1–2. https://www.proquest.com/docview/1287024909

Spikol, D., Ruffaldi, E., & Cukurova, M. (2017). Using multimodal learning analytics to identify aspects of collaboration in project-based learning. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a difference: Prioritizing equity and access in CSCL, 12th international conference on computer supported collaborative learning (CSCL) 2017* (Vol. 1, pp. 263–270). International Society of the Learning Sciences. https://repository.isls.org/handle/1/240

*Vujovic, M., Hernández-Leo, D., Tassani, S., & Spikol, D. (2020). Round or rectangular tables for collaborative problem solving? A multimodal learning analytics study. *British Journal of Educational Technology*, *51*(5), 1597–1614. https://doi.org/10.1111/bjet.12988

Wang, H., Barone, G., & Smith, A. (2024). Current and future role of data fusion and machine learning in infrastructure health monitoring. *Structure and Infrastructure Engineering*, *20*(12), 1853–1882. https://doi.org/10.1080/15732479.2023.2165118

Worsley, M. (2014). Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviours. In X. Ochoa, M. Worsley, K. Chiluiza, & S. Luz (Eds.), *MLA'14 proceedings of the 2014 ACM multimodal learning analytics workshop and grand challenge* (pp. 1–4). ACM Press. http://dx.doi.org/10.1145/2666633.2666634

Worsley, M., Martinez-Maldonado, R., & D'Angelo, C. (2021). A new era in multimodal learning analytics: Twelve core commitments to ground and grow MMLA. *Journal of Learning Analytics*, *8*(3), 10–27. https://doi.org/10.18608/jla.2021.7361

Appendix A

Summary of the Precedent Systematic Reviews

Authors	Foci	Characteristics
Alwahaby et al. (2022)	Exploring the ethical issues addressed in MMLA studies and the potential impact of MMLA on learning outcomes.	Age group: N/A Time frame: 2010–2020
Crescenzi-Lanna (2020)	Applications of MMLA research in the context of young children.	Age group: Children under six years old Time frame: 2015–2019
Mu et al. (2020)	Integration and analysis of multimodal data with a focus on types of data, learning indicators, and classification of data fusion methods.	Age group: N/A Time frame: 2017–2020
Ouhaichi et al. (2023)	Identification of research trends, methodologies, and applications in MMLA research.	Age group: N/A Time frame: 2014–2020
Prinsloo et al. (2023)	Investigating the rationale for MMLA applications, the nature and scope of data collected, the study contexts, evidence of commercial interests and/or downstream uses of students' data, and consideration of ethics, privacy, and the protection of student data.	Age group: N/A Time frame: 2011–2023
Sharma & Giannakos (2020)	Identification of research trends and applications in MMLA.	Age group: N/A Time frame: 2010–2020

Appendix B

Study Characteristics

Author(s) & Year	Participants	Multimodal data	Data collection channels	Learning environment	Data analysis	Algorithms	Data Fusion	Ethics
Andrade (2017)	15 3rd & 4th graders	—Hand gesture —Gaze direction —Pre- and post-tutorial questionnaire	—Video recordings —Questionnaire	N/A	—Dissimilarity measure —Hierarchical clustering —Categorical statistics —Heat plots	Hidden Markov model	Early fusion (Observed)	N/A
Andrade et al. (2016)	30 1st & 2nd graders	—Body language —Eye gaze —Gesture —Hedging language —Speech	Video recordings	School setting	—Descriptive statistics —Spearman correlation	Hidden Markov model	Early fusion (Observed)	N/A
Aslan et al. (2022)	10 1st graders	—Pose and gesture —Facial expression —Student interactions —Audio recordings	—Video recordings —Audio recordings —Touch detection	School setting	—Descriptive statistics —Qualitative data analysis method was not explicit but detailed description of process	N/A	N/A	Ethics committee approval
Camacho et al. (2020)	18 students (13–15 years old)	Log data (date, time, type/action, the student-initiated interactions were right, wrong, or duplicate)	—Wearable device —Smart objects —Log data	School setting	—Descriptive statistics	Decision Trees	Early fusion (Observed)	N/A
Lee-Cultura et al. (2022)	26 (10–12-year-old) children	—Videos —Eye gaze —Heart Rate —Skin Temperature —Blood Volume Pulse —EDA	—Video recording —Eye tracking —Wristbands —Kinect sensor	School setting	—Exploratory Factor Analysis —Predictive modelling —Qualitative data analysis method was not explicit but detailed description of process	Random forest	N/A	—Ethics committee approval —Informing participants, teachers, and parents about the research process —Moral-IT Deck tool for sensor data usage
Min et al. (2019)	92 middle school students	—Natural language utterance —Eye gaze traces —Task states —Gender	Log data	Game play in classroom	Developing a prediction model to detect breakdowns in student–agent conversations	Long Short-Term Memory Network	Early fusion (Observed)	N/A
Moon et al. (2022)	31 middle school students	—Log data —Facial expressions and gameplay screen recordings	—Log data —Video recordings	Gameplay either in school or home	Developing prediction model to track middle-school student distress states during educational gameplay	Random Forest	Late fusion (Reported)	N/A

Nasir et al. (2021)	96 students (9–12 years old)	—Log data —Video and audio recordings	Environment camera, RGB-D camera, microphones, logs (from touch screens)	N/A	—Kruskal-Wallis —Descriptive statistics —Qualitative interaction analysis of cases (method was not explicit but detailed description of process)	N/A	Early fusion (Observed)	Making data available
Ocak, Kopcha et al. (2023)	2 5th grade students	—Numerical representation —Imitating the robot —Use of the workbook —Use of the computer —Use of the robot	Video recordings	School setting	Machine learning model building for image classification	Deep Neural Network	Early fusion (Observed)	—Ethics committee approval —Sharing network design including coding
Olsen et al. (2020)	25 9–11-year-old dyads (50 students)	—Log data, —Audio (dialogue) data —Eye gaze —Pre- and post-test	—Eye tracking —Log data —Audio-only Skype connection —Pre- and post-test	Laboratory setting	—Descriptive statistics —Correlation —Temporal analysis	Deep LSTM	Early fusion (Reported)	—Ethics committee approval —Making data available
Ouyang et al. (2022)	28 3rd–5th graders	—Discourse audio —Class video —Log data —Computer screen recording (screen (clickstreams)	—Video recordings —Computer screen recording —Log data	School setting	—ANOVA, —ANCOVA —Social Network Analysis —Epistemic Network Analysis	N/A	N/A	Ethics committee approval
Papavlasopoulou et al. (2021)	29 children (8–14 years old)	—Log data —Video recording —Heart rate —EDA —Body temperature Blood volume pulse —Questionnaire	—Log data —Wristband —Video recording —Questionnaire	Science museum and school setting	—Paired samples t-tests —Pearson correlation —Spearman's correlations	N/A	N/A	N/A
Shin et al. (2020)	168 Tanzanian students	Log data	Log data (user actions and motions)	N/A	—Logistic Regression —Support Vector Machine —Decision Tree —Random Forest	—Logistic Regression —Support Vector Machine —Decision Tree —Random Forest	N/A	N/A
Vujovic et al. (2020)	24 elementary students	—Motion captures —Video recordings	—Motion capture systems (BTS Smart-DX motion) —Video recordings	Laboratory	—Quantitative analysis: Multifactorial ANOVA —Qualitative analysis: no detailed description	N/A	N/A	—Ethics committee approval —Informing participants about research process