British Journal of Educational Technology *Vol 51 No 5 2020* 1450–1484
doi:10.1111/bjet.12993

Multimodal data capabilities for learning: What can multimodal data tell us about learning?

Kshitij Sharma and Michail Giannakos

Kshitij Sharma is a senior researcher in the Department of Computer Science of NTNU. He received his PhD in Computer Science from the EPFL. His research interests include eye-tracking, MOOCs, collaborative learning, applied machine learning, multimodal learning and statistics. Michail Giannakos is a professor of interaction design and learning technology in the Department of Computer Science and leads the Learner–Computer Interaction lab at NTNU. His research interests center on developing new ways for humans to interact with interactive learning systems. Address for correspondence: Kshitij Sharma, Norwegian University of Science and Technology, Trondheim, Norway. Email: kshitij.sharma@ntnu.no

Abstract

Most research on learning technology uses clickstreams and questionnaires as their primary source of quantitative data. This study presents the outcomes of a systematic literature review of empirical evidence on the capabilities of multimodal data (MMD) for human learning. This paper provides an overview of what and how MMD have been used to inform learning and in what contexts. A search resulted in 42 papers that were included in the analysis. The results of the review depict the capabilities of MMD for learning and the ongoing advances and implications that emerge from the employment of MMD to capture and improve learning. In particular, we identified the six main objectives (ie, behavioral trajectories, learning outcome, learning-task performance, teacher support, engagement and student feedback) that the MMLA research has been focusing on. We also summarize the implications derived from the reviewed articles and frame them within six thematic areas. Finally, this review stresses that future research should consider developing a framework that would enable MMD capacities to be aligned with the research and learning design (LD). These MMD capacities could also be utilized on furthering theory and practice. Our findings set a baseline to support the adoption and democratization of MMD within future learning technology research and development.

Abbreviations: ACM, Association of Computer Machineries; AIED, Artificial Intelligence in Education; CS, Computer Science; CSCL, Computer-Supported Collaborative Learning; ECG, ElectroCardioGraphy; EEG, ElectroEncepheloGraphy; EDM, Educational Data Mining; GBL, Game-Based Learning; GSR, Galvanic Skin Response; HRV, Heart Rate Variability; IBL, Inquiry-Based Learning; ITS, Intelligent Tutoring System; LA, Learning Analytics; LAK, Learning Analytics and Knowledge; LD, Learning Design; LMS, Learning Management System; LSTM, Long Short Term Memory; MMD, Multimodal Data; MMLA, Multimodal Learning Analytics; MOOC, Massive Open Online Course; PBL, Problem-Based Learning; SoLAR, Society of Learning Analytics Research; SLR, Systematic Literature Review; SRL, Self-Regulated Learning; STEM, Science Technology Engineering Mathematics; TOCE, Transactions on Computing Education; UMUAI, User Modelling and User-Adapted Interaction; VR, Virtual Reality.

Practitioner Notes

What is already known about this topic

- Capturing and measuring learners' engagement and behavior using MMD has been explored in recent years and exhibits great potential.
- There are documented challenges and opportunities associated with capturing, processing, analyzing and interpreting MMD to support human learning.
- MMD can provide insights into predicting learning engagement and performance as well as into supporting the process.

What this paper adds

- Provides a systematic literature review (SLR) of empirical evidence on MMD for human learning.
- Summarizes the insights MMD can give us about the learning outcomes and process.
- Identifies challenges and opportunities of MMD to support human learning.

Implications for practice and/or policy

- Learning analytics researchers will be able to use the SLR as a guide for future research.
- Learning analytics practitioners will be able to use the SLR as a summary of the current state of the field.

Introduction

The confluence of multimodal data (MMD) with advanced computational analyses (multimodal learning analytics—MMLA, as the literature refers to them) enables us to understand and support complex learning phenomena (Blikstein & Worsley, 2016). For example, eye-tracking data and the different linguistic and prosodic features of speech can inform us about the students' expertise (Andrade, Delandshere, & Danish, 2016; Mangaroska, Vesin, & Giannakos, 2019); or video data can tell us about their engagement (Nguyen, Huptych, & Rienties, 2018; Pardo, Han, & Ellis, 2016). These insights can enable actionable feedback to be provided to the learners. For example, Hutt *et al.* (2019) used eye-tracking to automatically detect mind-wandering in online classes; while Grawemeyer *et al.* (2017) used students' speech and interaction to detect students' affective state. Such constructs (eg, mind-wandering, affective states) are used to provide feedback to the students (eg, by helping to determine the type of feedback that should be provided—reflective, instructive—and how it should be presented—evaluative, interpretive, supportive, probing).

Insights extracted from MMD enable us to investigate learners' behavior in ways that would not be possible with individual data sources. Giannakos, Sharma, Pappas, Kostakos, and Velloso (2019) found that the prediction of skill acquisition was far better using combined MMD (eg, eye-tracking, Electroencephalography [EEG] and facial video) than when using any individual stream. Research has shown that the fusion of MMD brings significantly better prediction of learning outcomes and helps us to interpret complex learning processes (Giannakos *et al.*, 2019; Liu *et al.*, 2019; Sharma, Papamitsiou, & Giannakos, 2019; Sharma, Pappas, Papavlasopoulou, & Giannakos, 2019). Despite the great potential of MMD for learning, and recent developments in the context of the CrossMMLA community, research in this direction has not reached its potential. For example, Worsley (2018) found that there are MMLA areas that remain largely underexplored (eg, supporting accessibility). In the same vein, other communities (eg, intelligent tutoring systems [ITS], Artificial Intelligence in Education [AIED], User Modeling and User-Adapted Interaction [UMUAI]) found that the implementation of MMLA in the areas where learning occurs (eg, classrooms, museums) is very challenging, and sometimes prohibitive (Baker & Ocumpaugh, 2015), with the result that MMLA's full potential is hindered (D'Mello & Kory, 2015; Giannakos *et al.*, 2019).

Proponents of MMD for learning have proposed several good arguments and benefits in seminal empirical and conceptual work. For example, Drachsler and Schneider (2018) suggested that MMD provide a more holistic picture of learning processes and success factors than the current form of knowledge extracted using individual data sources. Worsley and Blikstein (2018) argued for MMLA by suggesting that the existing strategies for analyzing MMD could provide more meaningful insights into complex learning processes than traditional approaches can. In the same vein, Spikol, Ruffaldi, Dabisias, and Cukurova (2018) and Blikstein and Worsley (2016) highlighted the importance and benefits of MMD in open-ended learning tasks, while Noel *et al.* (2018) and Liu *et al.* (2019) showcased the potential of MMD in rather restricted settings. Moreover, Prieto, Sharma, Kidzinski, Rodríguez-Triana, and Dillenbourg (2018) demonstrated that MMD provide insights both from the learners' side of the interaction and from the teachers' side of the interaction. Therefore, we see that there has been increased research in the field of MMD for learning and that it has focused on the following fronts. First, it had focused on the learning task: open-ended tasks such as, designing problems (Worsley & Blikstein, 2018) and close-ended tests like self-assessment tests (Sharma, Papamitsiou, *et al.*, 2019). Second, research has considered both sides of instruction: for students by predicting their learning task performance (Kaklauskas *et al.*, 2015) and for teachers in order to understand their classroom orchestration strategies and needs (Prieto, Sharma, Dillenbourg, & Jesús, 2016; Prieto *et al.*, 2018). Third, researchers have focused on the learning scenarios: face-to-face, to understand how teachers manage the classroom (Prieto *et al.*, 2016, 2018) and online learning scenarios, to predict students' learning-task performance (Florian-Gaviria, Glahn, & Gesa, 2013).

Learning analytics (LA) research has called for MMD actions with the formation of a CrossMMLA special interest group in the Society for Learning Analytics Research (SoLAR) (https://www.solaresearch.org/community/sigs/crossmmla-sig/) as well as for the organization of an annual workshop (http://crossmmla.org/) and dedicated special issues. A similar call for MMD research has also been made in recent literature review papers. For instance, Mangaroska and Giannakos (2018) identified that there is limited knowledge about how MMD can support the learning design (LD), and they suggested that the use of a broad set of complementary metrics (made possible through capturing MMD) will allow us to align the LD better with the student needs, thereby improving the learning process. This could be possible by finding the specific features that support understanding of complex learning experiences (Martinez-Maldonado *et al.*, 2016; Pantazos & Vatrapu, 2016) and by using these features to inform the LD (Bakharia *et al.*, 2016). Furthermore, Ruiz-Calleja, Prieto, Ley, Rodríguez-Triana, and Dennerlein (2017) suggested that LA (in the context of the workplace) could be enriched by exploring multiple data sources to overcome the problems of incomplete and scarce data caused by the low number of interactions between users and systems, thus, reducing the burden entailed by manual data gathering and increasing the chances of adoption. In the same vein, Worsley (2018) indicated that leveraging data from "non-traditional" modalities allows us to study and analyze learning that occurs in complex learning environments, thereby allowing us to collect data in ecological settings.

However, collecting and analyzing the MMD to answer a specific research question entails certain challenges as well. Challenges associated with MMD pose a significant impediment for many researchers. For example, each data stream has a different sampling rate (eg, eye-tracking 60–250 Hz, EEG 120–500 Hz, Video 10–60 FPS, Audio 44.1 KHz, heart rate 4 Hz) that incurs additional processing of the collected data to ensure that all the modalities are at the same temporal resolution. Moreover, all the data streams should be synchronized so that they can be analyzed together (Ochoa *et al.*, 2018; Worsley, 2018). Further, each data stream carries a different set of noise sources, and hence, improving the signal-to-noise-ratio for each data stream and having it at similar levels might be a tedious task (Sharma, Papamitsiou, *et al.*, 2019). In addition, each data stream entails a separate type of features and measures (emotions from faces, D'Mello & Graesser, 2012; attention from eye-tracking, Mangaroska, Sharma, Giannakos, Træteberg, & Dillenbourg, 2018; mental workload from EEG, Doppelmayr, Klimesch, Schwaiger, Auinger, & Winkler, 1998), and once the different measures and features are extracted from the collected, cleaned (noise-removal) and synchronized data streams, extracting LA-specific guidelines is another challenge (Bakharia *et al.*, 2016; Giannakos *et al.*, 2019). Given that most researchers on MMLA rely on custom-developed scripts and manual data alignment (Worsley, 2018), MMD can be inaccessible to those who are not already invested in this type of research and/or do not have the necessary technical competence.

Keeping the aforementioned benefits and challenges in mind, this paper presents a systematic literature review (SLR), strictly following the guidelines of Kitchenham and Charters (2007), with the aim of examining the empirical evidence on the capabilities of the insights extracted from MMD for learning. This SLR will allow us to provide information about the implementation and impact of MMD across a wide range of learning settings, contexts and empirical methods, and to provide robust and transferable evidence to other fields (eg, learner modeling, Educational Data Mining [EDM], Learning Analytics and Knowledge [LAK] at large, AIED). This paper presents an overview of what and how MMD have been used to inform learning and in what contexts. Although the MMLA field is still relatively young, enough work has already been done in the context of the CrossMMLA community to conduct a review.

Specifically, in this contribution, we systematically review the existing literature from the following two points of view.

RQ1: What is the current status of MMD for learning research, seen through the lens of areas of implementation (eg, learning scenario, learning environments), technologies used and methodologies (eg, types of data and data analysis techniques)?

RQ2: What insights the MMD has given us about learning (ie, MMD capabilities for learning)?

Our motivation for this work is based on developments in the area of MMLA creating momentum for increased use of multiple data sources to understand learners and learning processes. This study can provide a springboard for other scholars and practitioners, especially in the area of learning technologies, to examine MMD potential and MMLA approaches by taking into consideration the prior and ongoing research efforts.

Previous works that have reviewed the state of the field are concerned with individual modes (Blikstein, 2013; Blikstein & Worsley, 2016), or provide a conceptual framework using handpicked studies (Di Mitri, Schneider, Specht, & Drachsler, 2018), or analyze architectures (and are thus nonempirical works; Shankar, Prieto, Rodríguez-Triana, & Ruiz-Calleja, 2018) and do not engage in a systematic collection of empirical evidence (Blikstein & Worsley, 2016; Worsley, 2018). Apart from these review studies, there have been several MMLA and CrossMMLA workshops that have discussed the challenges and opportunities of MMLA. However, they are limited by the submission length, and thus, are neither comprehensive nor systematic. This study complements a vast amount of research in related fields, such as EDM, LAK, AIED, UMUAI, that have a long tradition of utilizing more than one data source to analyze and support students' behavior/performance. However, in light of the special issue of MMLA, this SLR aims to offer a solid review of works that have been conducted in the context of the CrossMMLA community and to discuss the results of the review through the lens of the neighboring fields (eg, UMUAI, AIED, EDM).

Methodology

In this SLR, we follow a transparent and widely accepted procedure (especially in the area of software engineering and information systems, as well as in educational technology) to minimize potential biases (researchers) and support reproducibility (Kitchenham & Charters, 2007). Besides the minimization of bias and the support of reproducibility, SLRs enable information to be provided about the impact of a phenomenon across a wide range of settings, contexts and empirical methods. Therefore, if the selected studies give consistent results, SLRs have the capacity to provide evidence that the phenomenon is robust and transferable (Kitchenham & Charters, 2007).

Articles collection

Several procedures were followed to ensure a high-quality review of the literature on MMD for learning. A comprehensive search of peer-reviewed articles was conducted in September 2019 (short papers, posters, dissertations, editorials and reports were excluded). The term "multimodal learning analytics" (also written as "multi-modal") was selected, since it is an umbrella term that captures the major works published in the confluence of MMD and LA in the developing MMLA community. Publications were selected from 2010 onward, because there have been tremendous advances since 2010 (eg, the LA field emerged) in the area of data-driven LA. A wide variety of databases were searched, including SpringerLink, Wiley, Assembly of Computer Machineries [ACM] Digital Library, IEEE Xplore, Science Direct, SAGE and ERIC. The search process uncovered 375 peer-reviewed articles.

Inclusion and exclusion criteria

The selection phase determines the overall validity of the literature review, and thus, it is important to define specific inclusion and exclusion criteria. We applied eight quality criteria (see Appendix A) informed by related works (eg, Dybå & Dingsøyr, 2008). Therefore, studies were eligible for inclusion if they were focused on MMLA. The aforementioned criteria were applied in Stage 2 and Stage 3 of the selection process (Figure 1), when the researcher had to assess the papers based on their titles and abstracts (Stage 2), and then, on the full papers (Stage 3).

Data analysis

In total, 42 studies met the quality criteria. We coded these studies according to specific areas of focus. These areas allowed us to consolidate the essence and the main focus of the studies. We selected categories that represent the MMD utilized as well as the objectives and content of the paper. This categorization enabled us to record all the necessary details from the papers in our literature review and to use them to address our research questions. In particular, each collected study was analyzed using the following elements:

1. Category: Experiment, Case study, Secondary Data Analysis, Ethnography.
2. Research topic: Social Sciences, Science Technology Engineering Mathematics (STEM), Computer Science (CS), Economics, Other (eg, Game Design, Robotics).

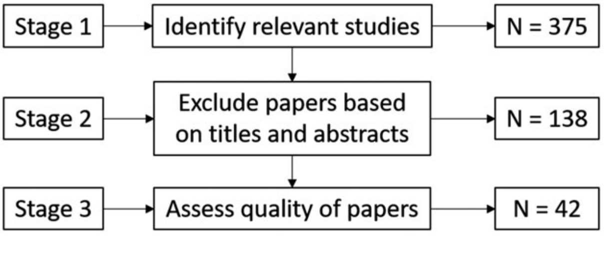

Figure 1: Stages of the selection process

3. Learning environment: LMS, ITS, Massive Open Online Course (MOOC), else.
4. Learning scenario: Formal, informal, nonformal.
5. Population: primary, secondary, high-school, undergraduate, graduate, master, teachers.
6. Sample size: Size of sample population.
7. Unit of analysis: individual or team.
8. Pedagogical approach: PBL, SRL, IBL, GBL, CSCL, Constructivism, Other.
9. Data collection: Type of data source/collection methods.
10. Methodology: Qualitative, quantitative, mixed.
11. Research objective: What was the main research objective of the contribution?
12. Behavior performance: What was the impact (if mentioned) of the intervention/innovation on learners' behavior?

The first 10 categories answer RQ1 and the latter two address RQ2. It is important to highlight that articles were coded based on reported information, that different authors reported information at different levels of granularity, and that in some cases the information was missing from the paper. Overall, the authors did their best to code the article as accurately and completely as possible. Details on the paper coding are shown in Appendix B.

Research findings

Domain, population and research topic

Most of the contributions presented a case study (30) or an experiment (10), while only one presented an ethnography. This could further showcase the exploratory phase of MMLA research since most of the contributions present an analytical/conceptual framework, and then, used the data collected from a study as an example "case study." For example, some studies proposed a method to obtain relevant predictions from the MMD and used a data collection based on a certain context as a case study (Sharma, Papamitsiou, *et al.*, 2019), or they proposed a method to use MMD to understand user stories in a software engineering course and used the data collected from a specific case study to test the proposed method (Noel *et al.*, 2018). The majority of the case studies (27) were conducted in formal learning settings, with a quarter of them (11) being in informal learning environments (four of them did not explicitly mention the learning settings). This shows that the MMLA community focuses on methods/artifacts-driven interventions through case studies conducted in convenient settings (formal learning/classroom settings), as we can notice by looking at the particular papers (Spikol *et al.*, 2018; Worsley & Blikstein, 2015, 2018). Moreover, of the 27 studies in formal learning scenarios, 20 were in university settings (where most of the teaching practices are formal). On the contrary, 6 out of 11 studies in informal learning were in primary, secondary or high school settings (where there is scope for higher integration of informal education, such as museum visits).

The unit of analysis employed was individual data for 23 studies and teams' data (collaborative settings) for 17 studies; there were also six studies that did not explicitly mention this information. In terms of the educational domain employed, most of the MMLA studies focused on STEM topics in general (13), a good number of studies (10) specified the domain area of CS education, while there were studies focusing on teaching overall and some on gaming and memorization (Figure 2, left), with most of the studies recruiting university students (Figure 2, right). Moreover, there were seven studies focusing on diverse domains such as doctoral students' activities (2), economics (1), environmental studies (1), communication skills (1), construction (1) and nursing (1). There were also papers in which the study was highly contextualized and the context influenced the scenario, the research design and the MMD collected; for example, Martinez-Maldonado, Echeverria, Santos, Santos, and Yacef (2018) analyzed MMD in the collaborative nursing context, where the participants were told to analyze a mannequin in a simulation. The

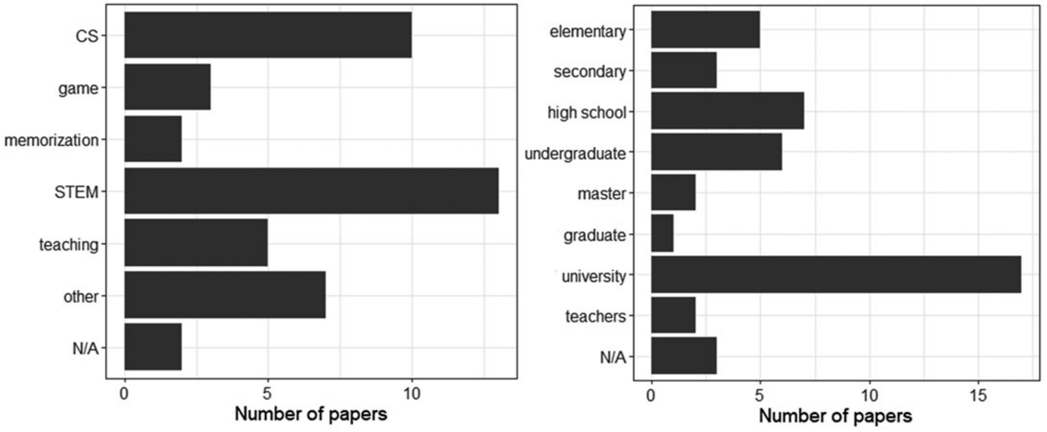

Figure 2: The different educational domains (left) and sample populations used in the contributions

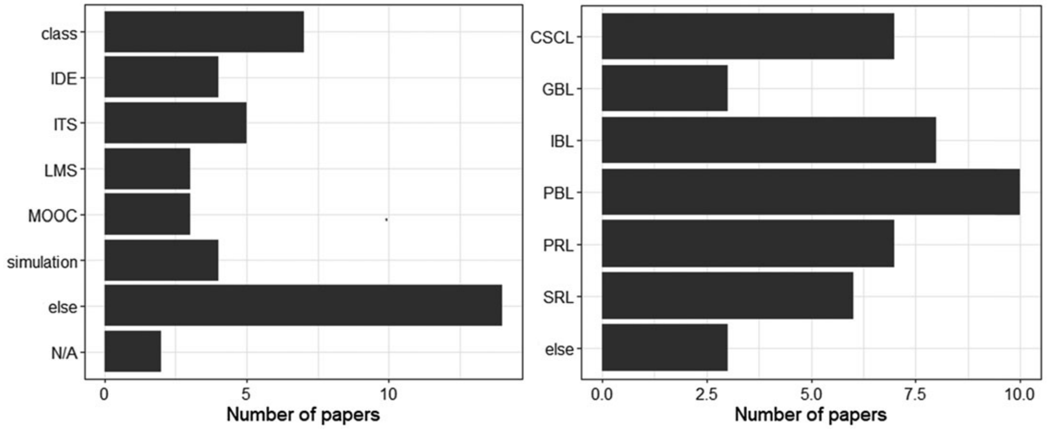

Figure 3: The different learning environments (left) and pedagogical underpinnings (right) employed in the MMLA studies

majority of the papers seemed to use educational domains and contexts that were convenient for them. For example, Ochoa *et al*. (2018) utilized MMD to provide feedback on students' communication skills, and Worsley and Blikstein (2018) analyzed MMD in the context of a collaborative construction task, with participants recruited through university mailing lists.

Learning environment and pedagogical approach

The MMLA contributions used a varied set of learning environments (Figure 3, left). The predominant learning environment was a face-to-face classroom setting (7), while others used ITS (5), interactive development environments (4), simulations (4), learning management systems (LMSs) (3) and MOOCs (3). There were 14 contributions that used specific tools outside these categories and four contributions that did not mention their learning environment. When it comes to the pedagogical underpinnings, the MMLA studies addressed a variety of approaches (Figure 3, right). In particular, problem-based learning (PBL) was the most frequent approach (10), while there was also a good distribution of other active learning approaches, such as inquiry-based learning (IBL, 8) and problem-regulated learning (PRL, 7). Other pedagogical underpinnings employed were computer-supported collaborative learning (CSCL, 7), self-regulated learning (SRL, 6), game-based learning (GBL, 3), example-based learning, principle-based learning and constructivism. The predominance of face-to-face and active learning approaches indicates that the focus of the community is to extract MMD from noninstructional learning activities and physical settings (non-digitally mediated learning). In particular, Worsley and Blikstein (2018) used an IBL framework where the students gathered three examples to understand engineering design; also, Junokas, Lindgren, Kang, and Morphew (2018) used a similar framework to ask the participants to understand the gesture recognition itself. In the same vein, Mangaroska *et al*. (2019) and Ezen-Can, Grafsgaard, Lester, and Boyer (2015) asked the participants to solve coding problems to learn debugging and programing, respectively. Furthermore, Martinez-Maldonado *et al*. (2018) and Spikol *et al*. (2016) analyzed MMD in collaborative nursing and collaborative programing contexts, respectively. Although most of the studies were not in the space of digitally mediated learning, we found a few studies, such as those by Di Mitri *et al*. (2017) and Florian-Gaviria *et al*. (2013), that focused on blended and online learning, respectively.

Data collection, sample size and methodology

Similar to the kind of population, the size of the population (Figure 4) participating in the contributions varied from under 10 (3) to more than 200 (1). There were six contributions with 10–20 participants, 16 with 20–40 participants, seven with 40–60 participants, one with 80–100 participants and six with more than 100 participants. There was one contribution that did not mention the sample size used in the study. A post hoc mapping between the modalities used in the studies and the sample size reveals that the majority of the studies with a sample size larger than 40 used data streams such as audio, video, logs and surveys (for the details of the mapping for the studies with more than 40 participants, see Appendix D). On the contrary, most of the other studies had more sophisticated data collection equipment, such as eye-tracking (Mangaroska *et al*., 2018), EEG (Giannakos *et al*., 2019; Sharma, Papamitsiou, *et al*., 2019), EDA sensors (Worsley & Blikstein, 2015) and Kinect (Kosmas, Ioannou, & Retalis, 2018). This ease of collection of data possibly explains the higher number of participants in the first set of studies.

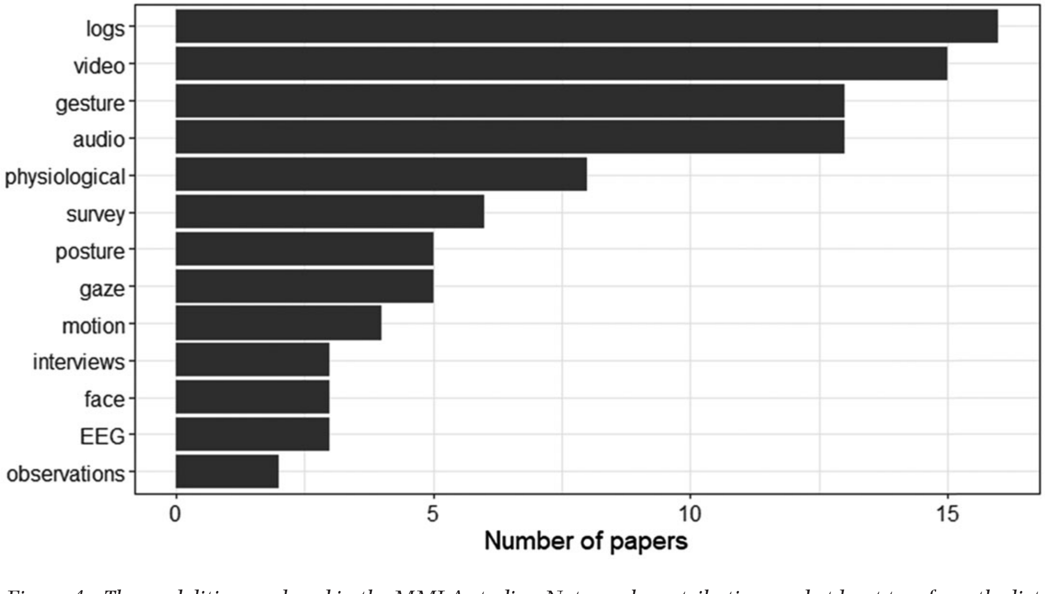

Figure 4: The modalities employed in the MMLA studies. Note: each contribution used at least two from the list

The most common data analysis methodology adopted a quantitative approach (24), while 10 contributions used a qualitative approach, and 10 studies reported results from mixed methods. Considering the modalities applied in the MMLA studies, a wide range of tools and techniques were used for data collection (Figure 4). The most common modalities captured were logs (16), videos (15), gesture (15), audio (13) and physiological data (8; physiological data includes heart rate variability [HRV], blood volume pulse, galvanic skin response [GSR]). Other modalities captured in the selected contributions were eye-tracking (5), face (3), EEG (3), motion (4), posture (5), interviews (3) and human observations (2). We report the modes as identified in the respective papers, but different authors described their modes with different levels of granularity (eg, face, video) and had different interpretations of what constitutes a modality (eg, is observation a modality in the MMLA context?).

In most of these contributions, there was a consistent use of these modalities. For example, videos were mostly used to annotate gestures and interactions (Prieto *et al.*, 2016, 2018; Worsley & Blikstein, 2015, 2018), eye-tracking was mostly used for finding attentional patterns and pupillary response (Mangaroska *et al.*, 2018; Prieto *et al.*, 2018), EEG was mostly used to extract features that correspond to deep mental processes such as long/short-term memory (LSTM) load and cognitive workload (Giannakos *et al.*, 2019; Prieto *et al.*, 2016; Sharma, Papamitsiou, *et al.*, 2019), and faces were mostly used to extract features that correspond to emotional responses like boredom, happiness, sadness and engagement (Florian-Gaviria *et al.*, 2013; Ochoa *et al.*, 2018).

Further investigation into how these different modalities were used revealed that logs and videos were mostly used to evaluate the performance of the participants, either in a quantitative way using logs (Liu, Stamper, & Davenport, 2018; Liu *et al.*, 2019; Mock *et al.*, 2016; Sharma, Papamitsiou, *et al.*, 2019) or environment variables (Mangaroska *et al.*, 2018), or in a qualitative manner using videos (Worsley & Blikstein, 2015, 2018). The rest of the multimodal sources were used to quantify behavioral trajectories, such as interaction behavior using touch gestures (Mock *et al.*, 2016), engagement with problem space using EDA and audio (flow, stress; Worsley & Blikstein, 2015, 2018), understanding and misconceptions using physiological data (Liu *et al.*, 2018, 2019), and problem-solving behavior using faces and EEG (Sharma, Papamitsiou, *et al.*, 2019) and eye-tracking (Mangaroska *et al.*, 2018).

With regard to the number of modes used in each of the studies, 21 studies used two modalities, nine used three modalities, four used four modalities and four used five modalities. The distribution of the number of modalities might also reflect the availability of the data capturing devices to the researchers. For example, three out of four studies using five modalities used both eye-tracking and EEG data (Giannakos *et al.*, 2019; Prieto *et al.*, 2016; Sharma, Papamitsiou, *et al.*, 2019), which involves using equipment that is costly and requires specialized technical competence. Therefore, it is worth further investigating whether the high cost of such off-the-shelf devices and the need for specialized technical competence were limiting factors when planning the data collection part of studies.

With regard to the sample size and number of modalities employed, Figure 5 shows that most of the studies employed two modalities and covered sample sizes of 10–200.

Research objective

When coding the papers, we also collected the research objective of each study. We then grouped the objectives and formed the following categories:

1. Explain the learning trajectories of the students (n = 16). By learning trajectories, we mean the MMD's capacity to portray learners' dialogues, strategies, behavior, interaction with the system or certain states of the students.

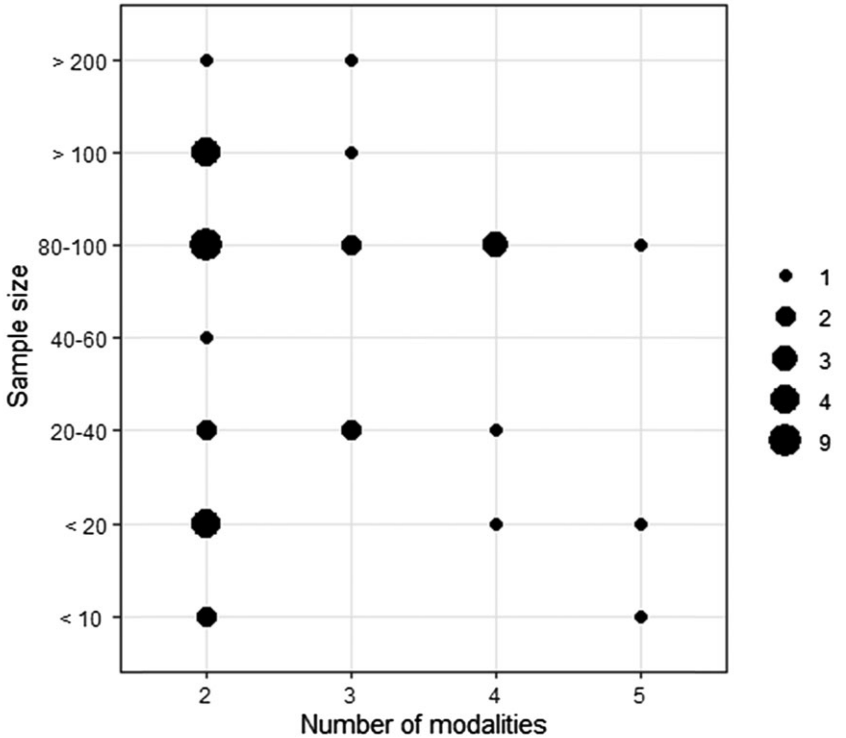

Figure 5: Number different modalities and the sample size in the MMLA studies

2. Predict the learning outcome (n = 5), such as grade, memorization capacity and skill acquisition.
3. Predict the learning performance on a task (n = 14).
4. Contribute toward teacher support (n = 6) in a manner that provides an awareness or reflection tool to the teacher or support in decision making.
5. Explain/predict student engagement (n = 2).
6. Present/experiment with a student feedback system (n = 2).

To investigate the connection between MMD and their capacity to assist us in understanding human learning, we mapped the various MMD with the research objectives (Figure 6, left) and the learning settings (Figure 6, right).

MMD for learning behavior and performance

We observed six major trends concerning the research objectives. In this section, we report on MMD's capacity to inform us about each of these six categories, focusing on the papers identified in each of the categories.

Behavioral trajectories (process)

Several researchers have pointed out how MMLA is capable of explaining learning processes or students' trajectories. Specifically, Kosmas *et al.* (2018) showed in a Kinect-based game that the play-time and range of motion improve students' short-term memory; Junokas *et al.* (2018) used skeleton positions and kinematics features to predict participants' recall; Andrade, Danish, and

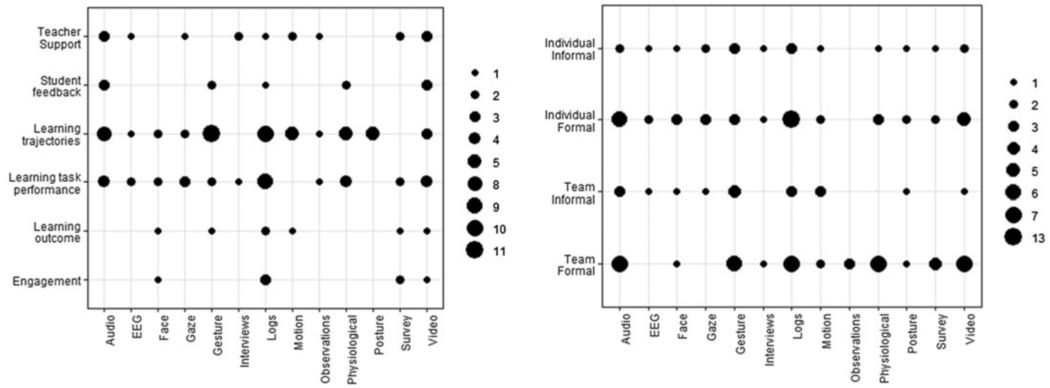

Figure 6: Association of research objectives of the MMLA studies with modalities employed (left); association of learning settings of MMLA studies with modalities employed (right). Note: Appendix C presents the detailed mapping, showing which papers belong in each of the circles

Maltese (2017) used hand positions and head poses to predict students' understanding in a predator-prey simulation; Spikol *et al.* (2018) used distance between students' faces, hand motion speed and the distance between hands to predict quality of students' projects; and Mock *et al.* (2016) used interaction logs from a touchscreen and the hand movements to predict the cognitive workload of the students. Therefore, the studies report that MMD can be successfully used to explain students' trajectories while the students engage in the learning task. From the aforementioned studies, we can see that interaction logs, gestures and posture were related to memory (Junokas *et al.*, 2018; Kosmas *et al.*, 2018), conceptual understanding (Andrade *et al.*, 2017), the artifact quality (Spikol *et al.*, 2018) and cognitive workload and motivation (Mock *et al.*, 2016).

In collaborative conditions, MMD have been used to identify key moments of collaboration (Noel *et al.*, 2018; Noroozi *et al.*, 2019; Pijeira-Díaz, Drachsler, Järvelä, & Kirschner, 2019). MMD have also been used to distinguish between help-seeking and help-giving behavior (Cukurova, Kent, & Luckin, 2019), tentative and casual problem-solving behavior (Andrade *et al.*, 2017), nonverbal behaviors (Cukurova, Luckin, Millán, & Mavrikis, 2018), solving versus guessing (Sharma, Papamitsiou, *et al.*, 2019) and the reasoning behavior of students (Worsley & Blikstein, 2015, 2018). Therefore, the studies identified in this category demonstrate the rich insights researchers can extract from MMD. Depending on the context, technologies and interactions involved, researchers have focused on understanding different micro-level (eg, nonverbal behaviors and guessing behaviors) and macro-level (eg, key collaborative moments, help giving and seeking and reasoning) aspects of students' trajectories.

Learning outcome

The MMLA studies report that MMD could be a key enabler for understanding and distinguishing the different levels of learning outcomes. With one exception, all studies utilized the logs of the system as one of the modalities, and all the studies used more than one modality to predict the learning outcome. In particular, Blikstein, Gomes, Akiba, and Schneider (2017) conducted a study with a focus on the laws of Newtonian physics, where students were required to build a tower and a bridge using a physics microworld. In the study, the authors did not find any effect of the experimental manipulation (ie, detailed vs generic instruction) on students' learning outcome and time to completion. However, the activation measured by the GSR was different for the two conditions; that is, for generic instruction the activation was monotonously decreasing, while the authors reported a "U-shaped" curve for the different parts of detailed instruction.

In another study, Andrade (2017) investigated MMLA's capacity to enrich our understanding of how elementary students explore the concept of feedback loops while controlling an embodied simulation of a predator-prey ecosystem using hand movements. The results of the study show five distinct motion sequences in students' embodied interactions, and these motion patterns are statistically associated with initial and post-tutorial levels of students' understanding of loops. Andrade (2017) demonstrated that embodiment "interacts" with the learning outcome and increases conceptual understanding, and that for this multidimensional association to be revealed the qualities of MMLA were critical. Overall, the MMLA studies identified in the learning outcome space used different modalities as a means to understand and explain the association of different complex instructional techniques and technologies (eg, embodiment, haptic, virtual reality [VR]) with the learning outcome.

Learning-task performance

The studies that focused on predicting and/or explaining learning-task performance mainly used logs, audio, video, eye-tracking and physiological modalities, with fewer studies using modalities like EEG, posture and survey. Di Mitri *et al.*, (2017) utilized MMD such as heart rate, step count, weather condition and learning activity to predict learning performance in SRL settings, with the results presenting "decent" prediction accuracy. In particular, the participants of the study were doctoral students who were asked to wear a Fitbit wristband and to rate their learning activities from 7 a.m. to 7 p.m. every hour while working in their normal routine (Di Mitri *et al.*, 2017).

In another study, Worsley and Blikstein (2018) investigated three MMLA approaches from a making-based learning activity. The authors concluded that there are many different approaches to applying MMLA and that each approach can provide a meaningful "glimpse" into a complex dataset, a glimpse that may be difficult to identify using non-MMLA approaches. In a controlled lab setting, participants were asked to play a version of a Pacman game focusing on movement-motor learning; Giannakos *et al.* (2019) compared MMD (EEG, facial video, eye-tracking, heart rate, electrodermal activation and blood volume pulse) to unimodal (logs) data and identified that traditional click streams (18% error rate) were significantly outperformed by MMD models in predicting movement-motor learning performance when the authors fused MMD (the error drops to 6%).

The MMLA studies identified in the learning-task performance category utilized the richness of MMLA to investigate different approaches (eg, making: Worsley & Blikstein, 2018; debugging: Mangaroska *et al.*, 2018; playing: Giannakos *et al.*, 2019) with the ultimate goals of understanding the experience (Di Mitri *et al.*, 2017; Giannakos *et al.*, 2019), improving either the technology (eg, Mangaroska *et al.*, 2018) or the design of the process (eg, Liu *et al.*, 2019; Worsley & Blikstein, 2018) and achieving a higher level of success on the given task. An interesting observation is that all the studies identified in this category were highly contextualized.

Teacher support

Prieto *et al.* (2018) reported that the temporal relationships of the events occurring at different points in a lesson can be captured using MMD and found that they are important in predicting a teacher's activity. Florian-Gaviria *et al.* (2013) demonstrated that integrated applications for adopting a well-grounded framework in teaching practice are required since they can assist teachers in creating contextual awareness. Prieto *et al.* (2016) reported reasonable accuracy using MMD features to predict teachers' activities in the classroom, which was later used as a reflection tool for teachers. Rodríguez-Triana, Prieto, Martínez-Monés, Asensio-Pérez, and Dimitriadis (2018) showed the positive impact of such a tool on prediction performance, as well as on the teacher's ability to interpret and react according to the results. Further, Cukurova *et al.*, (2019) argue that knowledge embedded in the analytical nature of classification models could be used to track/monitor each factor's contribution to decisions made in the classroom. Overall, MMLA to support teaching is a category with several studies, and most of them focus on less-sophisticated modalities (eg, audio, video and motion) that can be easily implemented in the classroom. Most studies converged on the fact that MMD's ability to distil information that is either nonobvious or demands high cognitive load for the teacher (eg, posture, gesture, level of collaboration) has important implications for the teaching design and teaching temporal actionable insights.

Engagement

The two studies that investigated engagement offer several insights into students' engagement patterns using MMLA. In particular, Pardo *et al.* (2016) proposed that a positive student experience of self-efficacy, tests, motivation, self-regulation and positive interaction with many online events, particularly those that offer feedback, reflection and reasoning, would correlate with academic achievement. In another study to investigate students' engagement in an online course, Nguyen *et al.* (2018) found, that high-performing students were studying in advance, while low-performing students showed more catching-up activities. For example, Nguyen *et al.* (2018) showed that for certain lessons in the later parts of the course all the students spent significantly more time on learning. However, the passing students spent more time in learning in advance while failed students spent more time catching-up with the material. Therefore, the results from both studies implicate that linking LD with MMLA-based insights, such as pointing out students to certain materials (eg, relevant online events: Pardo *et al.*, 2016) or frontloading and scaffolding learning materials that cause engagement and were avoided; and the ones causing disengagements (Nguyen *et al.*, 2018). Both studies also call for further research actions using MMD-based behavioral patterns to uncover student disengagement and to highlight the potential of MMD-based insights to inform the LD and to evaluate the implications of this for engagement (Nguyen *et al.*, 2018; Pardo *et al.*, 2016).

Student feedback

Interestingly, we identified only two studies that used MMD to provide feedback. Those two studies report high agreement between the human expert feedback and the MMD-based feedback. In particular, Ochoa *et al.* (2018) found that the system's feedback highly agrees with human feedback and that students considered this feedback to be useful and meaningful in developing their oral presentation skills. The authors concluded that an affordable system (camera and microphone) is able to provide feedback to avoid common errors in oral presentations. The feedback was given to the students off-line and post-presentation about their posture, gaze toward the audience and speech (pauses and volume). The initial evaluation showed that the majority of the students rated the system from acceptable to excellent (from 6 to 10 on a 10-point scale) in terms of overall experience, usefulness and learnability (Ochoa *et al.*, 2018).

Moreover, Ochoa *et al.* (2018) reported that the results from the hand-annotated data, which all correlated with the system's feedback, provided a consistent picture of how principle-based reasoning may be related to success and learning. It is, however, agreed that the more the complexity of the system increases (Worsley & Blikstein, 2015), the more difficult it becomes to provide high-quality feedback. In the fields of AIED and UMUAI, there have been a lot of contributions dealing with student feedback. For example, Grawemeyer *et al.* (2017) adapted the feedback based on students' affective states, reduced boredom and off-task behavior. Similarly, Davaris *et al.* (2019) used hand movement data and game logs in a VR setting not only to help the trainee surgeons understand basic anatomy, but also to provide feedback on the students' mistakes.

Discussion and future research directions

From the review process, we observe certain clear trends. STEM and CS are domains that have extensively researched and defined problems and problem-solving approaches. These domains are also heavily studied in the related fields, such as CSCL, LA and EDM, with dedicated special issues and venues on the intersection of these content areas (eg, In ACM Transactions on Computing

Education [TOCE], LA and CS education). Although there is no clear evidence about the reasons behind these trends, the convenience of conducting studies in the university context and the focus on action learning (which is very common in STEM) are two potential explanations (as shown in the left panel in Figure 2, CS and STEM are the domains that receive the most focus). Further, face-to-face classroom settings and ITS provide a better opportunity for MMD collection than do other learning environments because physiological sensors are easily integrated in such environments (as shown in the left panel in Figure 3, face-to-face classes and ITS have been the most used learning environments). In ITS, an eye-tracker and/or a webcam can easily be connected to the system, and the devices can use the system's (eg, ITS) timestamp for synchronization (the appropriateness of using MMD with ITS and standalone systems was also highlighted in a review and meta-study by D'Mello & Kory, 2015).

The challenging contexts for implementing MMD are MOOCs and LMS, since these platforms were not designed for multimodal input interaction. The following barriers are encountered: (1) MMD collection and integration are difficult (huge amount of data); (2) off-the-shelf devices are not yet "at-scale" (and when webcams, eg, are available, privacy and ethical requirements make their utilization impossible); and (3) students use such environments asynchronously (making collaborative MMD nonfunctional). However, the challenging contexts could also benefit from the use of MMD with the advancements in technology and proper ethical permissions (eg, in an LMS, if the students have the capacity to upload their Fitbit data while working on an assignment, it would help the researchers understand the arousal and engagement patterns).

Concerning data collection, MMLA researchers have used a plethora of data streams to obtain a deeper insight about the learners' rich and "multimodal" behavior. The most frequently used modalities are logs, audio, video and gestures (as shown in Figure 4). These modalities have a smaller cost and overhead in setting up than high-quality off-the-shelf sensor devices. Moreover, these modalities are the least obtrusive. This way, researchers could set up an experiment/case study seamlessly. However, implications from such studies are limited in terms of the depth of the understanding these modalities provide about the learners' behaviors and states.

By contrast, gaze, EEG, face and physiological data provide in-depth access to learners' affect, attention and cognition. In spite of gaze and EEG being highly informative, they are one of the least used modalities (as shown in Figure 4. However, high-quality off-the-shelf devices are expensive, and data collection and analysis need high expertise and entail several ethical issues. Furthermore, processing these data modalities is not trivial: because these are high-frequency modalities, there are also high noise levels, their synchronization is often challenging and there are several important methodological decisions that require high competence and contextual knowledge (eg, going from low-level signals to high-level abstractions, such as features and measurements).

Contemporary research on valid ways of collecting, preprocessing, synchronizing and analyzing these data streams can be found in related fields, such as human-computer interaction and ubiquitous computing (Grosse-Puppendahl *et al.*, 2017; Mansoorizadeh & Charkari, 2010; Newell, 1994; Rasmussen, 1983), ITS (D'Mello, Dieterle, & Duckworth, 2017), EDM (Blikstein & Worsley, 2016) and LAK (Di Mitri *et al.*, 2018). Therefore, the community can benefit from these approaches, techniques and recommendations. For example, D'Mello and Kory (2015) highlighted the trade-off between "accuracy" and "authenticity," since accurate results (high-data quality) are usually obtained in nonauthentic contexts (low ecological validity). In addition, "standard procedures," such as the one-off collecting labeled data to train supervised classifiers, are inherently limited due to the manual annotation process (D'Mello & Kory, 2015). Therefore, general techniques and recommendations are important and useful, but the final methodological decisions and applied techniques should be guided by specific application contexts.

Considering the target population (Figure 2, right panel), there are a few studies that focused on primary and secondary school children; while most studies focused on high school or university (undergraduate, master's, graduate) students. This could be because recruitment of participants depends on teachers and/or parents (legal guardian). MMD has not yet achieved a high level of social acceptance (Koelle *et al.*, 2018; Lee, Lee, Shin, & Oakley, 2018), which could be the reason why parents do not consent to their children participating in a study. Moreover, teachers might also think that data collection would hinder the normal learning process and be too intrusive.

Finally, we observe the following downward trend: when the researchers have used a high number of modalities to collect data (as shown in Figure 5), the sample size decreases (with the exception of Liu *et al.*, 2019). However, when we remove the studies with two modalities (the minimum to be considered "multimodal") we see that there is a significantly negative Spearman correlation between the number of modalities in MMD and the sample size ($r = .49$, $p = .05$, Figure 5). This indicates difficulties in collecting, preprocessing and analyzing MMD in large-scale studies. Moreover, such problems are aggravated when there are many psycho-physiological sensors (eg, EEG, GSR, heart rate, eye-tracking) involved (Giannakos *et al.*, 2019; Prieto *et al.*, 2018; Sharma, Pappas, *et al.*, 2019). Thus, from a practical perspective, when designing an MMLA study it is important for the researcher to consider the targeted population (eg, size, age), the objectives of the study (improving learning outcomes, engagement) and to carefully decide the modalities, taking into consideration their inherent qualities: noninvasive (cameras, audio, logs), invasive (EEG, eye-tracking glasses), sensitivity to noise, processing complexity, individual capacities (selection of appropriate modalities that produce useful for the under investigation phenomenon data) and collective capacities (ie, their ability to combine certain modalities to investigate different phenomena, such as combining video to find the moments of collaboration with eye-tracking to investigate the quality of those moments).

Opportunities

Several combinations of audio data, eye-tracking, system logs, video data and physiological data (eg, HRV, GSR) have been used to predict/explain learners' task performance (Florian-Gaviria *et al.*, 2013; Giannakos *et al.*, 2019; Sharma, Papamitsiou, *et al.*, 2019). On the contrary, several combinations of audio data, system logs, video data and physiological data have been used to define different behavioral trajectories (eg, Noel *et al.*, 2018; Spikol *et al.*, 2016; Spikol *et al.*, 2018). These two MMD sets provide deep insights into the performance and learning process from different standpoints. For example, gaze data can provide attentional information (Poole & Ball, 2006), physiological data can provide information about stress and arousal states of learners (Sharma, Pappas, *et al.*, 2019), and gesture, posture and motion data can tell us about the way learners interact/communicate with each other or with the system in a certain learning environment (eg, ITS, embodied play). Combining two or more of these datasets could improve our understanding and the predictive power of the ML algorithms in terms of the performance and learning process. However, contemporary MMLA applications are context-dependent, requiring customized and sometimes cumbersome methods that cannot be easily reused and standardized. Therefore, working toward the modularization and standardization of MMD for learning (eg, identifying data features that are not context-dependent) is a promising avenue that can allow us to further MMLA research and application.

Further, we observe that combinations of audio and video data have been used to provide teacher support, both in terms of post-class reflection (Prieto *et al.*, 2016, 2018) and decision making (Cukurova *et al.*, 2019). A plausible explanation for this could be that these two modalities are the most unobtrusive when it comes to monitoring a classroom, and they present almost no hindrance to the normal execution of routines in any given face-to-face classroom. Also, most of the recording devices have these two modalities synchronized at the hardware level, so there is no requirement for lengthy preprocessing in such cases.

With regard to the second mapping (Figure 6, right), we observe four clear trends. First, while analyzing a collaborative and formal setting, certain combinations of audio, video, physiological data and system logs were used (Martinez-Maldonado, Power, *et al.*, 2017; Noel *et al.*, 2018). Second, for informal situations in collaborative learning, audio, system logs, gesture and motion data were preferred (Spikol *et al.*, 2016, 2018). Third, in the case of formal instruction to individuals, audio, system logs, gesture and video data were used (Blikstein, 2013; Mock *et al.*, 2016). Finally, for informal and individual cases, audio, gestures and log data were utilized by most contributions (Andrade *et al.*, 2017; Ochoa *et al.*, 2018). These choices reflect peculiarities of instructional settings. For example, most informal collaborative settings are carried out with learners talking to each other (audio and gesture) in an environment (eg, museum) where they are required to move (motion and posture); while in collaborative formal settings, learners interact with each other and with the system (audio, video and logs), and during these interactions argumentation and negotiations take place (causing stress and arousal, hence, the usage of physiological data). In addition, the selection and utilization of different data streams might also be connected to the equipment and technical competence of the researchers.

Combining the aforementioned observations, we can reckon that there is a body of knowledge suggesting that there are certain combinations that are useful for predicting/explaining learners' behavior/performance/outcome, such as audio, eye-tracking, system logs, video and physiological data (see Figure 6 and Appendix C for the detailed mapping). However, further exploitation (new features, advanced ML methods) of these data sources is necessary to advance the current knowledge. Another interesting insight has to do with a limitation in the range of the learning scenarios that have been studied. There have been only a few learning scenarios explored (face-to-face: Prieto et al., 2016, 2018; scripted group-work: Martinez-Maldonado et al., 2018). This limitation becomes even more important when we start considering whether the students or the teacher in these learning scenarios are being monitored. This also suggests that MMLA is still in an exploratory phase, with most studies and prototypes addressing specific learning scenarios. The contribution of this paper will provide insights into MMD's capabilities to inform learning based on the empirical works published in the MMLA community; in addition, it will allow researchers to make informed decisions about which MMD to employ in their studies (eg, different goals and learning scenarios).

Challenges and future directions

During the mapping of the different modalities, we also came across a few gaps in the research. For example, we observed that EEG data are underutilized. Only three contributions (Giannakos *et al.*, 2019; Prieto *et al.*, 2018; Sharma, Papamitsiou, *et al.*, 2019) used a combination (or a subset) where EEG data were part of MMD. This might be for one or more of the following reasons: the low signal-to-noise ratio in EEG data; the requirement (which might sometimes be an overhead) of preprocessing (noise cancelation from eye movements, jaw movements and other possible sources due to involuntary motions) is higher for EEG data than for any other data source (as shown in Figure 6); and EEG costs the most in terms of ecological validity. However, there are emerging EEG devices that reduce the invasiveness of EEG (Vourvopoulos, Niforatos & Giannakos, 2019) and provide a good proxy (compared to other data sources in MMLA) to deeper cognitive states, such as cognitive load (Doppelmayr *et al.*, 1998), memory load (long term: Jensen & Tesche, 2002; short term: Ryu & Myung, 2005) and higher mental functions (Georgiadis, van

Oostendorp, & van der Pal, 2015). Therefore, it is necessary that EEG data collection and preprocessing is done carefully to obtain the most information from this valuable data source.

Another future direction for MMLA would be the use of facial videos, skeleton data, system logs and physiological data to provide actionable feedback to students. These MMD are the most ubiquitous, and many of them are noninvasive (eg, depth cameras such as Microsoft Kinect) or have minimized their invasiveness in recent years (eg, skin conductance mouse, wristbands), making them more appropriate for utilization in pedagogical settings. However, there is much research required to understand the timing and the type of feedback (Schwartz & Bransford, 1998). As we found in this SLR (which might be limited due to the special focus on MMLA), there is limited work on whether the feedback should be proactive (preventive feedback; eg, should the system forecast the mistake to occur and act before the students make the mistake?) or reactive (corrective feedback; eg, should the system wait for the students to make the mistake, and then, act accordingly?), on whether it should be initiated by the teacher, by the student or by an artificial agent and on how much information it should provide to the student about their behavior (Nicol, 2013; Ochoa *et al.*, 2018). Therefore, future work needs to focus on how to leverage MMD's capacities (eg, temporal information, affective information) to provide more effective and efficient feedback to the learner.

Future work should also focus on not only on designing the feedback tools based on MMD, but also on testing the tool's effectiveness and efficiency. There are plenty of feedback tools/practices in related fields, such as LAK, EDM, AIED and UMUAI. For example, Zhu, Xing, Costa, Scardamalia, and Pei (2019) triangulated the speech, sentiment and discourse analysis to provide personalized support to students and improve their collaborative knowledge building. Similarly, Grawemeyer *et al.* (2017) used students' speech and interaction data to adaptively scaffold their affective states so that the off-task behavior could be reduced and the learning gains could be increased. Finally, Davaris *et al.* (2019) used logs and movement data to facilitate support that reduced students' mistakes. However, further research is required to be able to provide feedback using multiple channels and to validate the effectiveness and efficiency of such feedback tools.

Finally, only two contributions combined attention (eye-tracking), cognition (EEG) and affect (face) measurements (Giannakos *et al.*, 2019; Sharma, Papamitsiou, *et al.*, 2019). Baker, D'Mello, Rodrigo, and Graesser (2010), D'Mello and Graesser (2010) and Mangaroska and Giannakos (2018) have called for these three measurements to be combined so that a holistic picture of the learners' performance, outcomes and behavior can be formed. Especially in formal settings, the synchronization of eye-tracking, EEG and facial videos can be achieved with a relatively lower level of difficulty than in informal settings.

Theoretical and practical implications

A number of implications emerge from this SLR on MMLA empirical published work in recent years. In the following sections, we summarize the implications derived from the reviewed articles by considering six thematic areas.

Enhanced prediction/explanation using MMLA

Using automatic analyses of the data, one key performance (methodological) trend in the MMLA research has been the showcasing of the improved and generalized performance of the studies. For example, Worsley and Blikstein (2015) suggested that MMLA can significantly enhance our understanding in detecting and modeling student learning. Giannakos *et al.* (2019) highlighted limitations of standalone clickstream models in a specific context, quantified the expected benefits of using a variety of MMD that employ physiological sensing, and proposed guidelines for design learning technologies. Liu *et al.* (2018) and Andrade *et al.* (2016) claimed their MMD analytic approach to be useful and generalizable to any branch of LA research in which data from multiple sources or modalities must be integrated for analysis. Cukurova *et al.* (2019), Liu *et al.* (2019) and Sharma, Papamitsiou, *et al.* (2019) demonstrated the improved accuracy of prediction over unimodal variables and proposed a generalized methodology for creating MMLA pipelines.

Automatic methods and machine Learning
MMD provide a unique opportunity for researchers to apply state-of-the-art ML techniques, since the data has different granularities and is sufficient for properly training the algorithms. Spikol *et al.* (2018) showed that new and promising approaches, such as neural networks, and more traditional regression approaches can both be used to classify MMLA data. Ezen-Can, Boyer, Kellogg, and Booth (2015) proposed the application of unsupervised models to asynchronous communication, which can enable massive-scale automated discourse analysis and mining to better support students' learning. Finally, Casey (2017) proposed methods to discover what constitutes relevant data within a particular learning context in programing using MMLA patterns.

Deep temporal Learning
Most of the methods used in the selected contributions either used the aggregated version of the MMD across the dependent variable that does not use time as a factor, or, when using temporal analysis, employed discrete classes/clusters of behavior (Worsley & Blikstein, 2018). Both these methods have their limitations. For example, clearly aggregating the data at the dependent variable level does not tell us anything about learners' behavioral trajectories. On the contrary, using discrete classes/clusters of behavior does not produce a holistic portrayal either. In related fields, such as EDM and teaching analytics, there are a few examples of using deep learning methods with temporal data to provide better predictability than when using clustering or aggregation (Grafsgaard, Duran, Randall, Tao, & D'Mello, 2018; Prieto *et al.*, 2018; Stewart, Keirn, & D'Mello, 2018). These methods could be further exploited in MMLA to further our understanding of learning performance/outcomes/trajectories.

Adaptation, personalization and moments to intervene
MMLA focuses on providing deep insights into the learning processes and learners' behavior. In addition, MMLA provides guidelines (and, in some cases, systems) to move toward adaptive and personalized learning scenarios. Junokas *et al.* (2018) demonstrated a promising way to create productive experiences with gesture-based educational simulations, promoting personalized interfaces and MMLA scenarios. Ezen-Can, Grafsgaard, *et al.* (2015) presented a step toward achieving a better understanding of student utterances by incorporating natural language processing based MMD features within adaptive learning environments. Kaklauskas *et al.* (2015) proposed a system that integrates the strong points of a student's personal learning style preference with a choice of available adaptive learning materials for the most appropriate learning adaptation (applying text, video, audio, computer learning systems and virtual and augmented realities). Finally, Cukurova *et al.* (2018) outlined implications for design, research and development of educational technology based on MMD measures to empower teachers with information that they can use to obtain a better view of the whole picture so that they can plan and adapt instruction accordingly.

Another facet of adaptation and personalization has been to discover the moments when, during the learners' interaction with the learning technology, personalized feedback should be provided. For example, Holmes, Latham, Crockett, and O'Shea (2017) presented a method for identifying timely intervention points at which an intelligent e-learning platform could scaffold the experience of learners who had failed to comprehend tutorial information. Di Mitri *et al.* (2017) presented a system for designing and developing predictive LA systems that exploit MMD about the learners, their contexts and their activities to predict the learners' current learning state, and thus, to generate timely scaffolding. For example, Amarasinghe, Hernández-Leo, and Jonsson (2019) used the engagement levels, measured from the system logs and interactive discourse, to adapt the group creation and scaffolding in a collaborative setting. Santos, Saneiro, Salmeron-Majadas, and Boticario (2014) used affective states, which were detected using physiological and log data, to provide motivational feedback adapted to students' personality and self-efficacy. Conati and MacLaren (2009) used emotions, detected by bodily expressions and game logs, to scaffold students' emotional states based on their personal goals. D'Mello *et al.* (2005, 2010) designed a conversational agent using students' discourse, facial expressions and body movements to provide the students with support according to their affective states. Finally, Grawemeyer *et al.* (2015, 2017) developed an ITS to support children learning fractions by using their speech and interaction logs.

Generalizability

MMLA often results in big datasets, which involve the risk of mis-implicating the results (Kidzinski, Sharma, Shirvani Boroujeni, & Dillenbourg, 2016). The mis-implication could surface for two main reasons. First, the effect sizes are a few orders of magnitude smaller than we are used to expect in classical educational psychology studies; and the results are still significant due to the large sample. Second, "black-box" approaches, like support vector machines or neural networks, give us great predictive power of the models, but they do not have a transparent predictor. Therefore, another challenging facet of the MMLA research is to seek generalizability. Only two (Liu *et al.*, 2019; Martinez-Maldonado, Shum, *et al.*, 2017) out of 42 contributions in this SLR have more than one case study (or experiment) reported. This means that the implications generally emerge from only one case study that was implemented in a specific context. This raises a big question about whether the results (or part of the results) are generalizable. One should use the basic principle of transfer learning (Bleidorn & Hopwood, 2019; Phillips, Wheeler, & Kochenderfer, 2017; Saari, Eerola, & Lartillot, 2010) to use a pretrained (or partially trained) model from one MMLA dataset to verify the implications of other datasets. In this direction, early results indicate that there are certain capacities of MMLA data that are generalizable across contexts, while there are also capacities that are context-dependent (Sharma, Niforatos, Giannakos, & Kostakos, 2020).

Importance of contextualization and ecological validity

All of the aforementioned advantages of MMLA come with a cost of ecological validity, which requires attention. The transfer of MMLA findings and implications from a constrained laboratory setting to an open-ended democratized setting is important (Blikstein, 2013; Martinez-Maldonado *et al.*, 2018; Pijeira-Díaz *et al.*, 2019). Martinez-Maldonado *et al.* (2018) set out guidelines for other researchers and developers seeking to provide enhanced support in simulation laboratories as well as in more generic collocated settings. Blikstein (2013) indicated how MMLA could be used to devise naturalistic assessments, which would, at the same time, be social, ecologically valid, more inclusive as to the types of knowledge they measure, and enabling of real-time evaluation in realistic tasks, either offline or online. As a step toward ecologically valid settings, Pijeira-Díaz *et al.* (2019) provided an ecologically valid picture of group dynamics in terms of arousal direction, level and contagion among triad members during collaborative learning.

In summary, the challenges in MMLA are concerned with: (1) the ecological validity of the studies; (2) the transfer of the knowledge from controlled studies to real-life scenarios; (3) using the findings to provide actionable feedback; (4) the manual annotation work, such as in labeled data, needed for various purposes (eg, to train supervised classifiers); (4) the lack of modularization and standardization of methods, since all the MMLA studies are custom and context-dependent; and (5) the fact that the insights extracted are also context-dependent, making it difficult to extract generalizable insights. MMLA also presents certain technical challenges to researchers who are not familiar with MMD noise reduction, synchronization, preprocessing and feature extraction. A few MMLA system architectures attempt to bridge the technological gaps across different contexts (for a review, see Shankar *et al.*, 2018). Furthermore, certain MMLA architectures are not completely context-agnostic, but are more generalizable than individual contexts/purposes (eg, ubiquitous learning: Muñoz-Cristóbal *et al.*, 2017; blended learning: Rodríguez-Triana *et al.*, 2018; incorporating annotations: Di Mitri, Schneider, Klemke, Specht, & Drachsler, 2019).

Ongoing research in MMLA

In recent years, in the context of LA research, there has been growing interest and momentum in MMLA. This is reflected in the formation of the CrossMMLA SIG, the continuous growth of the respective workshop and the increasing interest in this special issue. Given the timeframe of this SLR and this special issue, it would not be possible to incorporate the content of the special issue within our SLR, despite this collection of articles naturally fitting here. However, we would like to discuss the results of our SLR with respect to this special collection of MMLA studies.

The contributions in this collection (Cukurova, Giannakos, & Martinez-Maldonado, 2020) fall into one or more of the six major research trends identified in this literature review. For example, Larmuseau, Cornelis, Lancieri, Desmet, and Depaepe (2020) measure cognitive load (*behavioral trajectories*) using GSR, Electrocardiography (ECG) and HRV in an informal online problem-solving task. Dindar, Oulun, Järvelä, Haataja, and Oulun (2020) monitor meta-cognitive experiences (*behavioral trajectories*) and explain group performance (*learning performance*) using data coming from learners' EDA and survey responses in an informal collaborative problem-solving task. Olsen, Sharma, Rummel, and Aleven (2020) use LSTM on data features coming from learners' gaze, log, dialogue and audio to predict their posttest score (*learning performance*) and leaning gain (*learning outcomes*) in a formal collaborative setting of an ITS. Ahn and Harley (2020) use gaze and facial data to understand emotional processes (*behavioral trajectories*) and their relationship with learning gains (*learning outcomes*) in the context of GBL. Vujovic, Hernàndez-Leo, Tassani, and Spikol (2020) use motion and video data to explain level of participation and movement range (*behavioral trajectory*) in a collaborative formal learning setting. Emerson and Lester (2020) use gaze, facial and log data to predict students' motivation (*behavioral trajectories*) and *learning outcome*. Finally, Ochoa and Domínguez (2020) present a controlled evaluation against experts' opinion about the oral presentation skills (*feedback*) in the context of informal learning, using data coming from learners' video, audio and presentation files.

The collection includes two more papers that place strong emphasis on the ethical perspectives of MMLA. Although they do not focus on one of the six major research trends, they cover one or more in a transversal manner. Crescenzi-Lanna (2020) presents an SLR with a focus on MMLA in children under six. The author analyzed the contributions using seven scales: performance analytics (students' understanding and engagement); use of machine learning; use of eye-tracking; Kinect; biometrics; human-coded quantitative data; and qualitative data (interview, observation). Crescenzi-Lanna (2020) also provides a commentary on the ethical issues (eg, children being lied to in a "Wizard of Oz" experiment, longer and obtrusive exposure to experiments, effect on data quality of the interaction with strangers who are researchers, anonymity). Beardsley, Martínez-Moreno, Vujovic, Santos, and Hernàndez-Leo (2020) also tackle the ethical aspects of MMLA and introduce an informed consent comprehension test for MMLA research. The authors assessed its effects based on learners' comprehension and rates of enrolment in an MMLA study.

This special collection provides a good coverage of the different dimensions of MMLA research, as we saw them in this review; at the same time, it opens up new avenues for research with the six research trends to be major areas of focus in coming years, as well as the cross-cutting category of ethics.

Limitations and interpretation of the results

This work should be seen through the lens of some limitations. First, the authors had to make some methodological decisions (eg, selection of databases, the search query) that might introduce certain biases into the results. However, we did the best we could to avoid such biases by considering all the major databases and following the steps set out by Kitchenham and Charters (2007). Second, the selection of empirical studies and coding of the papers might pose another possible bias. However, the focus was clearly on the empirical evidence, and the coding of papers was performed by two independent researchers. Third, some elements of the papers were not described accurately, leading to some missing information in the coding of the papers. However, the amount of missing information was very small and could not affect the results significantly.

Another important limitation of this study is the focus on the MMLA-based theme of the special issue. This special issue responds to the growing interest in this theme as observed in a series of workshops organized by the MMLA community and in the respective terminology. Therefore, the selection criteria we employed captured relevant papers that used the keyword "multimodal learning analytics" (which is common in the MMLA community), but we did not capture potentially relevant papers that used "multimodal" and "analytics/modelling" as standalone keywords, such as affective learner modeling and multimodal affect detection, despite the potential of some of these papers being conceptually relevant (SLR refers to them as Type I errors). This positions the results of our work at the center of the MMLA community, but we also need to acknowledge that relevant works from neighboring communities, such as AIED, ITS and UMUAI, might have been excluded. This has several implications for the results, since areas in the intersection of multimodality and ITS, ubiquitous systems for learning, mobile learning, sensors and feedback systems, to mention but a few, might have been missed or underrepresented.

This is even more intense in the AIED, ITS and UMUAI communities where the development of models that automatically detect learner affect to inform feedback and regulate engagement is at the epicenter (eg, see extensive reviews in Calvo & D'Mello, 2010; D'Mello & Kory, 2015). This line of research is closely related to MMLA, with MMLA being an emerging domain of LA that analyses evidence from multimodal and multisystem data and extracts meaning from the increasingly fluid and complex data that come from different kinds of transformative learning situations (Giannakos *et al.*, 2020).

Both MMLA and multimodal affective detection communities focus on physical sensors, capture physiological and behavioral manifestations of emotion (Paquette *et al.*, 2016). Especially with recent advances in sensing technologies and computational methods, the interest in (and promise of) multimodality informing and supporting learning is ongoing, but it still requires plenty of effort to reach a matured stage. However, both the communities agree that further works are needed, because their implementation in the areas where learning occurs (eg, homes, classroom, museums) is challenging and sometimes prohibitive (Baker & Ocumpaugh, 2015), with researchers from both communities arguing that this line of work has not yet reached its full potential (Blikstein & Worsley, 2016; D'Mello & Kory, 2015; Giannakos *et al.*, 2019).

Conclusions

We have presented an SLR of 42 contributions in the field of MMLA from the last 9 years. We analyzed the papers from the perspective of the study design (learning context, environment, population and so on) and the insights they provide about the learners' task-based performance/outcome or behavior. We categorized the main findings of the selected papers in six thematic areas and discussed the challenges and opportunities emerging from the current review in terms of both the MMD used and the impact it could have on our understanding of learners' outcome and behavior. Finally, based on the current state of the field, we have proposed further possible advancements.

Acknowledgements
This work is supported by the Norwegian Research Council under the projects FUTURE LEARNING (number: 255129/H20) and Xdesign (290994/F20). We would like to thank Manolis Mavrikis, the editor-in-chief of BJET, for his comments and suggestions.

Statements on open data, ethics and conflict of interest

Since this is a systematic literature review, there is no data to be shared. The coding scheme is presented in Appendix B.

Appropriate permissions and ethical approval were requested and approved.

There is no potential conflict of interest in this study.

References

Ahn, T. B., & Harley, J. (2020). Exploring emotions and multimodal learning analytics: Eye-tracking and facial recognition. *British Journal of Educational Technology*.

Amarasinghe, I., Hernández-Leo, D., & Jonsson, A. (2019). Data-informed design parameters for adaptive collaborative scripting in across-spaces learning situations. *User Modeling and User-Adapted Interaction*, *29*, 869–892. https://doi.org/10.1007/s11257-019-09233-8

Andrade, A. (2017). Understanding student learning trajectories using multimodal learning analytics within an embodied-interaction learning environment. In *Proceedings of the Seventh International Learning Analytics and Knowledge Conference* (pp. 70–79). New York, NY: ACM.

Andrade, A., Danish, J. A., & Maltese, A. V. (2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, *4*(3), 18–46.

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.

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241.

Baker, R. S., & Ocumpaugh, J. (2015). Interaction-based affect detection in educational software. In R. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford handbook of affective computing* (pp. 233–245). New York, NY: Oxford University Press.

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gasevic, D., Mulder, R., … Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. In T. Reiners, B. R. von Konsky, D. Gibson, V. Chang, L. Irving, & K. Clarke (Eds.), *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 409–413). New York, NY: ACM.

Barmaki, R., & Hughes, C. E. (2018). Embodiment analytics of practicing teachers in a virtual immersive environment. *Journal of Computer Assisted Learning*, *34*(4), 387–396.

Beardsley, M., Martínez-Moreno, J., Vujovic, M., Santos, P., & Hernàndez-Leo, D. (2020). Enhancing consent forms to support participant decision making in multimodal learning research. *British Journal of Educational Technology*.

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, *23*(2), 190–203.

Blikstein, P. (2013). Multimodal learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 102–106). New York, NY: ACM
Blikstein, P., Gomes, J. S., Akiba, H. T., & Schneider, B. (2017). The effect of highly scaffolded versus general instruction on students' exploratory behavior and arousal. *Technology, Knowledge and Learning*, *22*(1), 105–128.
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.
Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, *1*(1), 18–37.
Casey, K. (2017). Using keystroke analytics to improve pass-fail classifiers. *Journal of Learning Analytics*, *4*(2), 189–211.
Conati, C., & MacLaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, *19*, 267–303.
Crescenzi-Lanna, L. (2020). Multimodal learning analytics research with young children: A systematic review. *British Journal of Educational Technology*. 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*.
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.
Cukurova, M., Luckin, R., Millán, E., & Mavrikis, M. (2018). The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers and Education*, *116*, 93–109.
D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S., Picard, R. W., & Graesser, A. C. (2005). Affective interactions: The computer in the affective loop. In *Workshop at the International Conference on Intelligent User Interfaces* (pp. 7–13).
D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, analytic, automated (AAA) measurement of engagement during learning. *Educational Psychologist*, *52*(2), 104–123.
D'mello S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, *20*(2), 147–187.
D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157.
D'Mello, S., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys*, *47*(3), 1–36. https://doi.org/10.1145/2682899
D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., … Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In *Tenth International Conference on Intelligent Tutoring Systems*. ITS.
Davaris, M., Wijewickrema, S., Zhou, Y., Piromchai, P., Bailey, J., Kennedy, G., & O'Leary, S. (2019, June). The importance of automated real-time performance feedback in virtual reality temporal bone surgery training. In *International Conference on Artificial Intelligence in Education* (pp. 96–109). Cham, Switzerland: Springer.
Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., & Specht, M. (2017). Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data. In *Proceedings of the Seventh International Learning Analytics and Knowledge Conference* (pp. 188–197). New York, NY: ACM.
Di Mitri, D., Schneider, J., Klemke, R., Specht, M., & Drachsler, H. (2019). Read between the lines: An annotation tool for multimodal data for learning. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (pp. 51–60).
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.
Dindar, M., Oulun, Y., Järvelä, S., Haataja, E., & Oulun, Y. (2020). What does physiological synchrony reveal about metacognitive experiences and group performance? *British Journal of Educational Technology*.

Doppelmayr, M., Klimesch, W., Schwaiger, J., Auinger, P., & Winkler, T. (1998). Theta synchronization in the human EEG and episodic retrieval. *Neuroscience Letters*, *257*(1), 41–44.

Drachsler, H., & Schneider, J. (2018). JCAL special issue on multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 335–337.

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology*, *50*(9–10), 833–859.

Emerson, A., & Lester, J. (2020). Multimodal learning analytics for game-based learning. *British Journal of Educational Technology*.

Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 146–150). New York, NY: ACM.

Ezen-Can, A., Grafsgaard, J. F., Lester, J. C., & Boyer, K. E. (2015). Classifying student dialogue acts with multimodal learning analytics. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 280–289). New York, NY: ACM.

Florian-Gaviria, B., Glahn, C., & Gesa, R. F. (2013). A software suite for efficient use of the European qualifications framework in online and blended courses. *IEEE Transactions on Learning Technologies*, *6*(3), 283–296.

Georgiadis, K., van Oostendorp, H., & van der Pal, J. (2015). EEG assessment of surprise effects in serious games. *International Conference on Games and Learning Alliance* (pp. 517–529). Cham, Switzerland: Springer.

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, *48*, 108–119.

Giannakos, M., Spikol, D., Molenaar, I., Di Mitri, D., Sharma, K., Ochoa, X., & Hammad, R. (2020). CrossMMLA in practice: Collecting, annotating and analyzing multimodal data across spaces. In *Companion Proceedings of LAK20*.

Grafsgaard, J., Duran, N., Randall, A., Tao, C., & D'Mello, S. (2018). Generative multimodal models of nonverbal synchrony in close relationships. In *2018 Thirteenth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018)* (pp. 195–202). IEEE.

Grawemeyer, B., Mavrikis, M., Holmes, W., Gutiérrez-Santos, S., Wiedmann, M., & Rummel, N. (2017). Affective learning: Improving engagement and enhancing learning with affect-aware feedback. *User Modeling and User-Adapted Interaction*, *27*(1), 119–158.

Grawemeyer, B., Mavrikis, M., Holmes, W., Hansen, A., Loibl, K., & Gutiérrez-Santos, S. (2015). Affect matters: Exploring the impact of feedback during mathematical tasks in an exploratory environment. In *Proceedings of the 17th International Conference on Artificial Intelligence in Education (AIED 2015): Lecture Notes in Computer Science* (pp. 595–599). Berlin, Germany: Springer.

Grosse-Puppendahl, T., Holz, C., Cohn, G., Wimmer, R., Bechtold, O., Hodges, S., Smith, J. R. (2017, May). Finding common ground: A survey of capacitive sensing in human-computer interaction. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3293–3315).

Holmes, M., Latham, A., Crockett, K., & O'Shea, J. D. (2017). Near real-time comprehension classification with artificial neural networks: Decoding e-learner non-verbal behavior. *IEEE Transactions on Learning Technologies*, *11*(1), 5–12.

Hutt, S., Krasich, K., Mills, C., Bosch, N., White, S., Brockmole, J. R., & D'Mello, S. K. (2019). Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction*, *29*(4), 821–867.

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscience*, *15*(8), 1395–1399.

Junokas, M. J., Lindgren, R., Kang, J., & Morphew, J. W. (2018). Enhancing multimodal learning through personalized gesture recognition. *Journal of Computer Assisted Learning*, *34*(4), 350–357.

Kaklauskas, A., Kuzminske, A., Zavadskas, E. K., Daniunas, A., Kaklauskas, G., Seniut, M., … Radzeviciene, A. (2015). Affective tutoring system for built environment management. *Computers and Education*, *82*, 202–216.

Kidzinski, Ł., Sharma, K., Shirvani Boroujeni, M., & Dillenbourg, P. (2016). On generalizability of MOOC models. In *Proceedings of the Ninth International Conference on Educational Data Mining* (pp. 406–411).

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (Technical Report). EBSE.

Koelle, M., Boll, S., Olsson, T., Williamson, J., Profita, H., Kane, S., & Mitchell, R. (2018, April). (Un) Acceptable!?! Re-thinking the social acceptability of emerging technologies. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (p. W03). New York, NY: ACM.

Kosmas, P., Ioannou, A., & Retalis, S. (2018). Moving bodies to moving minds: A study of the use of motion-based games in special education. *TechTrends*, *62*(6), 594–601.

Larmuseau, C., Cornelis, J., Lancieri, L., Desmet, P., & Depaepe, F. (2020). Multimodal learning analytics to investigate cognitive load during online problem solving. *British Journal of Educational Technology*.

Lee, D., Lee, Y., Shin, Y., & Oakley, I. (2018, October). Designing socially acceptable hand-to-face input. In *The 31st Annual ACM Symposium on User Interface Software and Technology* (pp. 711–723). New York, NY: ACM.

Liu, R., Stamper, J. C., & Davenport, J. (2018). A novel method for the in-depth multimodal analysis of student learning trajectories in intelligent tutoring systems. *Journal of Learning Analytics*, *5*(1), 41–54.

Liu, R., Stamper, J., Davenport, J., Crossley, S., McNamara, D., Nzinga, K., & Sherin, B. (2019). Learning linkages: Integrating data streams of multiple modalities and timescales. *Journal of Computer Assisted Learning*, *35*(1), 99–109.

Mangaroska, K., & Giannakos, M. N. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*.

Mangaroska, K., Sharma, K., Giannakos, M., Træteberg, H., & Dillenbourg, P. (2018). Gaze-driven design insights to amplify debugging skills: A learner-centred analysis approach. *Journal of Learning Analytics*, *5*(3), 98–119.

Mangaroska, K., Vesin, B., & Giannakos, M. (2019). Cross-platform analytics: A step towards personalization and adaptation in education. In *Proceedings of the Ninth International Conference on Learning Analytics & Knowledge* (pp. 71–75). New York, NY: ACM.

Mansoorizadeh, M., & Charkari, N. M. (2010). Multimodal information fusion application to human emotion recognition from face and speech. *Multimedia Tools and Applications*, *49*(2), 277–297.

Martinez-Maldonado, R., Echeverria, V., Santos, O. C., Santos, A. D. P. D., & Yacef, K. (2018). Physical learning analytics: A multimodal perspective. In *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (pp. 375–379). New York, NY: ACM.

Martinez-Maldonado, R., Power, T., Hayes, C., Abdiprano, A., Vo, T., Axisa, C., & Buckingham Shum, S. (2017). Analytics meet patient manikins: Challenges in an authentic small-group healthcare simulation classroom. In *Proceedings of the Seventh International Learning Analytics and Knowledge Conference* (pp. 90–94). New York, NY: ACM.

Martinez-Maldonado, R., Schneider, B., Charleer, S., Shum, S. B., Klerkx, J., & Duval, E. (2016). Interactive surfaces and learning analytics: Data, orchestration aspects, pedagogical uses and challenges. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 124–133).

Martinez-Maldonado, R., Shum, S. B., Schneider, B., Charleer, S., Klerkx, J., & Duval, E. (2017). Learning analytics for natural user interfaces. *Journal of Learning Analytics*, *4*(1), 24–57.

Mock, P., Gerjets, P., Tibus, M., Trautwein, U., Möller, K., & Rosenstiel, W. (2016). Using touchscreen interaction data to predict cognitive workload. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (pp. 349–356). New York, NY: ACM.

Muñoz-Cristóbal, J. A., Rodriguez Triana, M., Bote-Lorenzo, M. L., Villagrá-Sobrino, S. L., Asensio-Pérez, J. I., & Martínez-Monés, A. (2017). Toward multimodal analytics in ubiquitous learning environments. In *Joint Proceedings of the Sixth Multimodal Learning Analytics (MMLA) Workshop and the Second Cross-LAK Workshop Co-located with Seventh International Learning Analytics and Knowledge Conference* (Vol. *1828*, pp. 60–67). CEUR.

Newell, A. (1994). *Unified Theories Of Cognition*. Cambridge, MA: Harvard University Press.

Nguyen, Q., Huptych, M., & Rienties, B. (2018). Using temporal analytics to detect inconsistencies between learning design and student behaviours. *Journal of Learning Analytics*, *5*(3), 120–135.

Nicol, D. (2013). Resituating feedback from the reactive to the proactive. In D. Boud, & E. Molloy (Eds.), *Feedback in higher and professional education: understanding it and doing it well* (pp. 44–59). London, UK: Routledge.

Noel, R., Riquelme, F., MacLean, R., Merino, E., Cechinel, C., Barcelos, T. S., … Munoz, R. (2018). Exploring collaborative writing of user stories with multimodal learning analytics: A case study on a software engineering course. *IEEE Access*, *6*, 67783–67798.

Noroozi, O., Alikhani, I., Järvelä, S., Kirschner, P. A., Juuso, I., & Seppänen, T. (2019). Multimodal data to design visual learning analytics for understanding regulation of learning. *Computers in Human Behavior*, *100*, 298–304.

Ochoa, X., & Domínguez, F. (2020). Controlled evaluation of a multimodal system to improve oral presentation skills in a real learning setting. *British Journal of Educational Technology*.

Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G., & Castells, J. (2018). The rap system: Automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors. In *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (pp. 360–364). New York, NY: ACM.

Olsen, J., Sharma, K., Rummel, N., & Aleven, V. (2020). Using multimodal data to temporally analyze collaborative learning outcomes: Benefits and challenges. *British Journal of Educational Technology*.

Pantazos, K., & Vatrapu, R. (2016). Enhancing the professional vision of teachers: A physiological study of teaching analytics dashboards of students' repertory grid exercises in business education. In *Proceedings of the 49th Hawaii International Conference on System Sciences* (pp. 41–50).

Paquette, L., Rowe, J., Baker, R., Mott, B., Lester, J., DeFalco, J., … Georgoulas, V. (2016). *Sensor-free or sensor-full: A comparison of data modalities in multi-channel affect detection*. International Educational Data Mining Society.

Pardo, A., Han, F., & Ellis, R. A. (2016). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies*, *10*(1), 82–92.

Phillips, D. J., Wheeler, T. A., & Kochenderfer, M. J. (2017). Generalizable intention prediction of human drivers at intersections. In *2017 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1665–1670). IEEE.

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2019). Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members? *Computers in Human Behavior*, *92*, 188–197.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia of human computer interaction* (pp. 211–219). Hershey, PA: IGI Global.

Prieto, L. P., Sharma, K., Dillenbourg, P., & Jesús, M. (2016). Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 148–157). New York, NY: ACM.

Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M. J., & Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of Computer Assisted Learning*, *34*(2), 193–203.

Rasmussen, J. (1983). Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*, 257–266.

Reis, R. C. D., Isotani, S., Rodriguez, C. L., Lyra, K. T., Jaques, P. A., & Bittencourt, I. I. (2018). Affective states in computer-supported collaborative learning: Studying the past to drive the future. *Computers & Education*, *120*, 29–50.

Rodríguez-Triana, M. J., Prieto, L. P., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2018). The teacher in the loop: Customizing multimodal learning analytics for blended learning. In *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (pp. 417–426). New York, NY: ACM.

Ruiz-Calleja, A., Prieto, L. P., Ley, T., Rodríguez-Triana, M. J., & Dennerlein, S. (2017). Learning analytics for professional and workplace learning: A literature review. In *European Conference on Technology Enhanced Learning* (pp. 164–178). Springer.

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, *35*(11), 991–1009.

Saari, P., Eerola, T., & Lartillot, O. (2010). Generalizability and simplicity as criteria in feature selection: Application to mood classification in music. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(6), 1802–1812.

Santos, O. C., Saneiro, M., Salmeron-Majadas, S., & Boticario, J. G. (2014). A methodological approach to elicit affective educational recommendations. In *Proceedings of the 14th International Conference on Advanced Learning Technologies (ICALT 2014)* (pp. 529–533).

Schönborn, K. J., Bivall, P., & Tibell, L. A. (2011). Exploring relationships between students' interaction and learning with a haptic virtual biomolecular model. *Computers & Education*, *57*(3), 2095–2105.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475–5223.

Shankar, S. K., Prieto, L. P., Rodríguez-Triana, M. J., & Ruiz-Calleja, A. (2018). A review of multimodal learning analytics architectures. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)* (pp. 212–214). IEEE.

Sharma, K., Niforatos, E., Giannakos, M., & Kostakos, E. (2020). Assessing cognitive performance using physiological and facial features: Generalizing across contexts. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(3).

Sharma, K., Papamitsiou, Z., & Giannakos, M. (2019). Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach. *British Journal of Educational Technology*, *50*(6), 3004–3031.

Sharma, K., Pappas, I., Papavlasopoulou, S., & Giannakos, M. (2019). Towards automatic and pervasive physiological sensing of collaborative learning. In K. Lund, G. P. Niccolai, E. Lavoué, C. H. Gweon, & M. Baker (Eds.), *Thirteenth International Conference on Computer Supported Collaborative Learning (CSCL)* (pp. 684–687). International Society of the Learning Sciences (ISLS).

Spikol, D., Avramides, K., Cukurova, M., Vogel, B., Luckin, R., Mavrikis, M., & Ruffaldi, E. (2016). Exploring the interplay between human and machine annotated multimodal learning analytics in hands-on STEM activities. In *Proceedings of the Sixth International Learning Analytics and Knowledge Conference* (Vol. *6*, pp. 522–523). New York, NY: ACM.

Spikol, D., Ruffaldi, E., Dabisias, G., & Cukurova, M. (2018). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning*, *34*(4), 366–377.

Stewart, A. E., Keirn, Z. A., & D'Mello, S. K. (2018). Multimodal modeling of coordination and coregulation patterns in speech rate during triadic collaborative problem solving. In *Proceedings of the 2018 International Conference on Multimodal Interaction* (pp. 21–30). New York, NY: ACM.

Vourvopoulos, A., Niforatos, E., & Giannakos, M. (2019). EEGlass: An EEG-eyeware prototype for ubiquitous brain-computer interaction. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (pp. 647–652).

Vujovic, M., Hernàndez-Leo, D., Tassani, S., & Spikol, D. (2020). Studying collaborative learning and space design with multimodal learning analytics. *British Journal of Educational Technology*.

Worsley, M. (2018). *Multimodal learning analytics' past, present, and potential futures*. CrossMMLA@ LAK.

Worsley, M., & Blikstein, P. (2015). Leveraging multimodal learning analytics to differentiate student learning strategies. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 360–367). New York, NY: ACM.

Worsley, M., & Blikstein, P. (2018). A multimodal analysis of making. *International Journal of Artificial Intelligence in Education*, *28*(3), 385–419.

Zhu, G., Xing, W., Costa, S., Scardamalia, M., & Pei, B. (2019). Exploring emotional and cognitive dynamics of knowledge building in grades 1 and 2. *User Modeling and User-Adapted Interaction*, *29*(4), 789–820.

APPENDIX A

Critical appraisal skills programme criteria

The selection phase determines the overall validity of the literature review, and thus it is important to define specific inclusion and exclusion criteria. As Dybå and Dingsøyr (2008) specified, the quality criteria needs to cover three main issues (ie, rigour, credibility, and relevance) that needs to be considered when evaluating the quality of the selected studies. We applied eight quality criteria informed by related works (eg, Dybå & Dingsøyr, 2008). Following are those criteria:

1. Does the study clearly address the research problem?
2. Is there a clear statement of the aims of the research?
3. Is there an adequate description of the context in which the research was carried out?
4. Was the research design appropriate to address the aims of the research?
5. Does the study clearly determine the research methods (subjects, instruments, data collection, data analysis)?
6. Was the data analysis sufficiently rigorous?
7. Is there a clear statement of findings?
8. Is the study of value for research or practice?

APPENDIX B

Table B1: Coding details Part 1

ID	Category	Research topic	Learning Environment	Learning scenario	Sample size	Population
1	Case study	STEM	Else	Informal	20	High-school and undergraduate
2	Ethnography	Game	Else	Informal	31	Elementary
3	Case study	STEM	simulation	Formal	19	High-school
4	Case study	STEM	ITS	Informal	18	University
5	Case study	STEM	CLASS	Formal	1	Teacher
6	Case study	memorization	Else	Formal	21	Undergraduate
7	Case study	Teaching	Simulation	Formal	30	University
8	Case study	CS	Else	Formal	60	University
9	Experiment	CS	LMS	Formal	145	University
10	Case study	CS	MOOC	Formal	20	Teachers
11	Case study	CS	Simulation	Formal	44	Undergraduate
12	Case study	Other	Class	Formal	9	Graduate
13	Case study	Other	Class	Formal	56	Undergraduate
14	Case study	Other	Class			
15	Case study	CS	IDE	Formal	20	University
16	Experiment	STEM	IDE	Informal	6	University
17	Case study	STEM	Else	Formal	30	Elementary
18	Case study	STEM	Else	Formal	13	Undergraduate
19	Case study	Teaching	Class			Elementary
20	Experiment	Teaching	MOOC	Formal	155	University
21	Case study	Teaching	LMS	Formal	165	Students
22	Case study	Other	Else	Formal	83	University
23	Experiment	CS	IDE	Formal	34	University
24	Case study	Game	Else	Informal	15	Elementary
25	Case study	Other	Else	Formal	20	High-school and undergraduate
26	Experiment	Memorization	Else	Informal	17	University
27	Case study	Other	ITS	Formal	206	Master
28	Experiment	STEM	Else	Formal	20	Master

Table B1: (Continued)

ID	*Category*	*Research topic*	*Learning Environment*	*Learning scenario*	*Sample size*	*Population*
29						
30	Experiment	CS	else	Informal	45, 36	Secondary, university
31	Experiment	STEM	ITS	Formal	24	High-school
32	Experiment	STEM	Class	Formal	24	High-school
33	Case study	other	LMS	Formal	182	University
34	Case study	STEM	ITS	Formal	56	High-school
35	Experiment	STEM	SIMULATION	Informal	31	Elementary
36	2 case studies					
37	Case study	Game	Else	Informal	15	Secondary
38	Case study	CS	IDE	Informal	40	University
39	Case study	CS	MOOC	Formal	111	University
40	Case study	Teaching	Class	Informal	41	University
41	2 case studies	STEM	ITS	Formal	59, 104	High-school and secondary
42	Case study	CS	Else	Formal	32	University

ITS = intelligent tutoring system; LMS = learning management system; MOOC = massive open online courses; IDE = integrated development environment; CS = computer science education.

Table B2: Coding details Part 2

ID	Unit of analysis	Pedagogical approach	Data collection	Methodology
1	Team	IBL	Video, audio, gesture and bio-physiology	Quantitative
2	Team	GBL	Psychometric prepost testing, games-usage analytics, a student attitudinal scale, teachers' reflection notes and teacher interviews, Kinect sensor	Mixed
3	Team	GBL	Survey, coding	Quantitative
4	Team	else	Capture of objects; the positions of people, hand movements, face tracking; audio and video	Quantitative
5	Individual	PBL	Eye track, audio-visual	Quantitative
6	Team	IBL	Gesture data, skeleton positions, kinematics feature	Mixed
7	Team	IBL	Visual feedback, movements	qualitative
8	Team	PBL	Microphones	Mixture
9	Team	PBL	Academic performance, video, digital footprint, questionnaire	Quantitative
10	Individual	SRL	Survey	qualitative
11	Individual	PRL	Web camera image streams	Quantitative
12	Individual	SRL	Software tracking tool installed on laptop	Quantitative
13	Team	PRL		
14		PRL		
15	Individual	PRL	Examples, challenges, and coding exercises	Quantitative
16	Team	CSCL	Arm Tracking, Audio Levels, Motion, Objects, facial and object tracking	Qualitative
17	Individual	PRL	Touch point detection	Quantitative
18	Individual	PRL	Video and gesture tracking	Quantitative
19	Team	CSCL	Eye-tracking, EEG, accelerometer, audio and video	
20		CSCL		Quantitative
21	Team	CSCL	Analysis of the learning designs, teacher interviews, researcher observations, student questionnaires, system logs, and student-generated artefacts	Both
22	Individual	SRL	Video, audio, slides	qualitative
23	Individual	PBL	Posture, gesture features	Quantitative
24	Individual	Else	Hand movements, video frames	Qualitative
25	Team	Else	Audio, gesture and video, measuring stress and/arousal	Quantitative

Table B2: (Continued)

ID	*Unit of analysis*	*Pedagogical approach*	*Data collection*	*Methodology*
26	Individual	GBL	Click-stream data, as well as eye-tracking, electroencephalography (EEG), video, and wristband	Quantitative
27	Individual	SRL	Voice, grades	Quantitative
28	Individual	IBL	Movements, students' interaction with the model	Both
29		CSCL		
30	Individual + Team	IBL	Video data, sensor data	Quantitative
31	Team	CSCL	EDA sensor recordings	Quantitative
32	Team	IBL	Sensor data	Quantitative
33	Individual	SRL	Clicks, views	Quantitative
34	Individual	IBL	Screen video and audio captures	Both
35	Individual	IBL	Gesture, speech, hedging, gaze. body, eye contact	Both
36		PBL, PBL		
37	Individual	PBL	Movement data, hand position	Both
38	Individual	PBL	Eye tracking, gaze, debug solving	Quantitative
39	Individual	PBL	Time spent on each slide of the learning materials, IP address, keystroke timings, successful compiles, failed compiles (again recording the source code) and GUI interactions such as menu clicks and window opening/closing.	Quantitative
40	Individual	Else	Audio, experience, survey	Quantitative
41	Individual and Team	PBL, CSCL	Logs, audio, video	Both
42	Individual	SRL	Logs gaze eeg wristband face	Quantitative

PBL = problem based learning; IBL = inquiry based learning; PRL = problem regulated learning; CSCL = computer supported collaborative learning; SRL = self regulated learning; GBL = game based learning.
Key (number to reference) for Tables B1 and B2.
1 Worsley and Blikstein (2018). 2 Kosmas *et al.* (2018). 3 Blikstein *et al.* (2017). 4 Spikol *et al.* (2018). 5 Prieto *et al.* (2018). 6 Junokas *et al.* (2018). 7 Barmaki and Hughes (2018). 8 Noel *et al.* (2018). 9 Pardo *et al.* (2016). 10 Florian-Gaviria *et al.* (2013). 11 Holmes *et al.* (2017). 12 Di Mitri *et al.* (2017). 13 Martinez-Maldonado, Power, *et al.* (2017). 14 Martinez-Maldonado *et al.* (2018). 15 Mangaroska *et al.* (2019). 16 Spikol *et al.* (2016). 17 Mock *et al.* (2016). 18 Blikstein (2013). 19 Prieto *et al.* (2016). 20 Ezen-Can, Grafsgaard *et al.* (2015). 21 Rodríguez-Triana *et al.* (2018). 22 Ochoa *et al.* (2018). 23 Ezen-Can, Grafsgaard *et al.* (2015). 24 Andrade (2017). 25 Worsley and Blikstein (2015). 26 Giannakos *et al.* (2019). 27 Kaklauskas *et al.* (2015). 28 Schönborn, Bivall & Tibell (2011). 29 Reis *et al.* (2018). 30 Cukurova *et al.* (2018). 31 Noroozi *et al.* (2019). 32 Pijeira-Díaz *et al.* (2019). 33 Nguyen *et al.* (2018). 34 Liu *et al.* (2019). 35 Andrade *et al.* (2016). 36 Martinez-Maldonado, Power, *et al.* (2017). 37 Andrade *et al.* (2017). 38 Mangaroska *et al.* (2018). 39 Casey (2017). 40 Cukurova *et al.* (2019). 41 Liu *et al.* (2018). 42 Sharma, Pappas, *et al.* (2019).

APPENDIX C

Table C1: Mapping between the research objectives and the MMD

	Engagement	*Learning outcome*	*Learning task performance*	*Learning trajectories*	*Student feedback*	*Teacher support*
Audio		13	1, 27, 34, 41	1, 4, 8, 16, 25, 34, 35, 41	22, 25	5, 19, 40
EEG			26, 42	42		19
Face		11	10, 26, 42	16, 42		
Gaze			10, 26, 38, 42	35, 42		5, 19
Gesture		24	1, 18, 24	1, 4, 6, 16, 17, 23, 25, 30, 35, 37	25	
Interviews			17			21, 40
Logs	9, 33	3, 11, 13, 28	9, 12, 17, 26, 27, 38, 39, 41, 42	2, 4, 8, 16, 17, 32, 41, 42	22	14, 21
Motion		13, 28	24	2, 4, 6, 16, 37		7, 14, 19
Observations			1	1		21
Physiological			1, 12, 26, 42	1, 25, 31, 32, 42	25	
Posture			24	6. 23. 30. 35		
Survey	9, 33	3	9, 17			21, 40
Video	9	24	9, 18, 34, 41	4, 25, 34, 41	22, 25	5, 7, 19

Table C2: Mapping between the Learning settings and the MMD

	Team formal	*Team informal*	*Individual formal*	*Individual informal*
Audio	8, 13, 25, 41	1, 4, 16, 19	5, 27, 34, 35, 41	22, 40
EEG		19	42	26
Face		10, 16	11, 42	26
Gaze		10, 19	35, 42	26, 38
Gesture	6, 25	1, 4, 16, 30	17, 18, 23, 35	24, 30, 37
Interviews	21		15	40
Logs	3, 8, 9, 13, 21, 32, 41	2, 4, 16	11, 12, 14, 15, 17, 27, 28, 33, 39, 41, 42	22, 26, 38
Motion	6, 7	4, 16, 19	14, 28	37
Observations	21	1		
Physiological	25, 31, 32	1	12, 42	26
Posture	6	30	23, 35	30
Survey	3, 21		15, 33	40
Video	6, 7, 9, 25, 41	4, 19	5, 18, 34, 41	22, 24

Key (number to reference) for Tables C1 and C2.
1. Worsley and Blikstein (2018). 2 Kosmas *et al*. (2018). 3 Blikstein *et al*. (2017). 4 Spikol *et al*. (2018). 5 Prieto *et al*. (2018). 6 Junokas *et al*. (2018). 7 Barmaki and Hughes (2018). 8 Noel *et al*. (2018). 9 Pardo *et al*. (2016). 10 Florian-Gaviria *et al*. (2013). 11 Holmes *et al*. (2017). 12 Di Mitri *et al*. (2017). 13 Martinez-Maldonado, Power, *et al*. (2017). 14 Martinez-Maldonado *et al*. (2018). 15 Mangaroska *et al*. (2019). 16 Spikol *et al*. (2016). 17 Mock *et al*. (2016). 18 Blikstein (2013). 19 Prieto *et al*. (2016). 20 Ezen-Can, Grafsgaard *et al*. (2015). 21 Rodríguez-Triana *et al*. (2018). 22 Ochoa *et al*. (2018). 23 Ezen-Can, Grafsgaard *et al*. (2015). 24 Andrade (2017). 25 Worsley and Blikstein (2015). 26 Giannakos *et al*. (2019). 27 Kaklauskas *et al*. (2015). 28 Schönborn *et al*. (2011). 29 Reis *et al*. (2018). 30 Cukurova *et al*. (2018). 31 Noroozi *et al*. (2019). 32 Pijeira-Díaz *et al*. (2019). 33 Nguyen *et al*. (2018). 34 Liu *et al*. (2019). 35 Andrade *et al*. (2016). 36 Martinez-Maldonado, Power, *et al*. (2017). 37 Andrade *et al*. (2017). 38 Mangaroska *et al*. (2018). 39 Casey (2017). 40 Cukurova *et al*. (2019). 41 Liu *et al*. (2018). 42 Sharma, Pappas, *et al*. (2019)

APPENDIX D

Table D1: Mapping the studies with sample size more than 40 to the type of data collected

ID	*Sample*	*Data collection*
Noel *et al*. (2018)	60	Microphones
Pardo *et al*. (2016)	145	Academic performance, video, digital footprint, questionnaire
Holmes *et al*. (2017)	44	Web camera image streams
Martinez-Maldonado, Power, *et al*. (2017)	56	Video, logs
Ezen-Can, Grafsgaard, *et al*. (2015)	155	Logs, text
Rodríguez-Triana *et al*. (2018)	165	Analysis of the learning designs, teacher interviews, researcher observations, student questionnaires, system logs, and student-generated artefacts
Ochoa *et al*. (2018)	83	Video, audio, slides,
Kaklauskas *et al*. (2015)	206	Voice, grades
Cukurova *et al*. (2018)	45, 36	Video data, log data
Nguyen *et al*. (2018)	182	Clicks, views
Liu *et al*. (2019)	56	Screen video and audio captures
Casey (2017)	111	Time spent on each slide of the learning materials, IP address, keystroke timings, successful compiles, failed compiles (again recording the source code) and GUI interactions such as menu clicks and window opening/closing
Cukurova *et al*. (2019)	41	Audio, experience, survey
Liu *et al*. (2018)	59, 104	Logs, audio, video