THIS IS A DRAFT VERSION.

Technologies for automated analysis of co-located, real-life, physical learning spaces: Where are we now?

Yi Han Victoria Chua
Nanyang Technological University
50 Nanyang Avenue
Singapore
vicchuayh@gmail.com

Justin Dauwels
Nanyang Technological University
50 Nanyang Avenue
Singapore
jdauwels@ntu.edu.sg

Seng Chee Tan
National Institute of Education
1 Nanyang Walk
Singapore
sengchee.tan@nie.edu.sg

ABSTRACT

The motivation for this paper is derived from the fact that there has been increasing interest among researchers and practitioners in developing technologies that capture, model and analyze learning and teaching experiences that take place beyond computer-based learning environments. In this paper, we review case studies of tools and technologies developed to collect and analyze data in educational settings, quantify learning and teaching processes and support assessment of learning and teaching in an automated fashion. We focus on pipelines that leverage information and data harnessed from physical spaces and/or integrates collected data across physical and digital spaces. Our review reveals a promising field of physical classroom analysis. We describe some trends and suggest potential future directions. Specifically, more research should be geared towards a) deployable and sustainable data collection set-ups in physical learning environments, b) teacher assessment, c) developing feedback and visualization systems and d) promoting inclusivity and generalizability of models across populations.

CCS CONCEPTS

• **Human-centred computing**~Visualization design and evaluation methods •**Information systems**~Data analytics

KEYWORDS

Face-to-face classroom analysis, co-located learning, physical learning analytics, educational data mining, educational technologies

1 INTRODUCTION

The field of educational data mining (EDM) is concerned with developing models, methods and algorithms originally used in data mining to extract and make sense of the large volume of data derived from educational settings to provide automated detection and analysis (for a review, see [77]). On the other hand, learning analytics (LA) refers to the field of measuring, collecting and analyzing data from educational settings with the aim of quantifying, improving and optimizing learning.

To date, the majority of EDM and LA studies focus on characterizing students when they are interacting with computer-based educational systems (CBESs), in which a computer system is an integral part of the learning process [67]. These studies examine human-computer interactions of teachers and students within variants of CBESs such as learning management systems (LMSs), massive open online courses (MOOCs), intelligent tutoring systems (ITSs) etc. In these environments, the learning tasks are more structured, the entirety of interactions happen in the digital space, and users are often seated or situated right in front of a computer screen, making it easier to collect digital and physical data with less noise [69], ensuring sufficient fidelity and ecological validity. As observed by Berland and colleagues [68], the degree of structure in these environments makes it easier to infer associations between student behavioral responses and learning constructs of interest. For this reason, observations and frameworks developed in the research literature of EDM and LA are largely based in computer-based educational environments.

The motivation for this paper is derived from the fact that while majority of learning still occurs in face-to-face educational settings, there is an under-representation of learning processes that do not occur through or are mediated by a computer within the EDM and LA research literature [75]. We observed a shift in educational paradigms to adopt blended learning models, where face-to-face learning is often combined and facilitated with technologies, but most LA and EDM studies are fixated on the online portions of such learning curriculums, leveraging only the digital footprints and online logs of teacher and student interactions.

In face-to-face, co-located learning environments, students communicate and interact with their peers via speech, facial expressions and body gestures while teachers or facilitators monitor these cues and reciprocate accordingly in real-time. Examples of such learning environments include project-based learning, embodied interaction, constructionist, or simply, traditional teacher-students classrooms, each of these with ranging degrees of structure in its curriculums. There has been increasing interest among researchers and practitioners in developing technologies that model and analyze learning and teaching experiences beyond computer-based learning environments and capturing learner and teacher data beyond digital spaces [69], [42]. As such, the term "*physical learning analytics*" was recently coined to refer to research and paradigms which brings learning analytics methods and innovations into physical learning spaces and attempts to leverage and make sense of physical data to aid teaching practices and learning processes [74].

There exist various comprehensive literature reviews on the current state of EDM and LA, its key methodologies and data analysis tools [73], [77], but these tools and technologies are primarily developed for computer-based or computer-assisted learning or rarely explore using face-to-face physical data in learning environments. There are also papers, which focus on multi-modal analytics (MMLA) or natural user interfaces, offering brief reviews of, but not limited to, face-to-face classroom analysis (e.g. [69], [75, 76])

In this paper, we focus on tools and technologies designed and developed for face-to-face classroom analysis, with a specific focus of leveraging from the physical sphere. As far as we know, no such review exists at the point of conception of our paper. We review case studies that deploy tools and technologies to harness behavioral data beyond digital and online logs in face-to-face, co-located learning environments, with the purpose of quantifying learning processes and supporting improvement in learning and teaching in an automated fashion. We identify current trends, research gaps, challenges and future directions by categorizing case studies according to their data sources, data modalities, outcomes, targets of assessments, deployment settings, units of analysis and the degree of maturity of these technologies and tools developed. The following research questions (RQ) motivate the direction of the paper:

- **RQ1:** What are the types of data harnessed in automated face-to-face classroom analysis?
- **RQ2:** What are the research objectives and outcomes in automated face-to-face classroom analysis?
- **RQ3:** What is the maturity of such automated tools and technologies in terms of application and evaluation?
- **RQ4:** What are the open issues and main challenges of applying automated techniques to face-to-face classroom analysis?

The rest of the paper is structured as follows. In Section 2, we outline our definitions of face-to-face classroom, physical data, and automated analysis, our methodology and inclusion/exclusion criterion for the selection of case studies, and the frameworks we adopted to categorize, evaluate and synthesize findings from the reviewed studies. In Section 3, we describe the trends we observed from reviewing the case studies, with emphasis towards RQs 1 to 3. We then focus on RQ4, discussing trends, challenges, and future directions in the field of face-to-face classroom analysis in Section 4, followed by our conclusions in Section 5.

2 EXPERIMENTAL AND COMPUTATIONAL DETAILS

2.1 Definitions

2.1.1 Definition of face-to-face classroom. A face-to-face classroom can be defined as a "real-world, non-computer mediated environment" differing from computer-based learning environments "where the computer features as an active component in the learning process" [75]. Such learning environments include traditional lectures, open-ended hands-on activities and collaborative learning activities. Most in-class discourse, interaction and activity amongst students and between teachers and students should occur in the same physical space and in real-time. As such, we also include learning contexts where learning occurs largely in the real-world but can be facilitated by computer interfaces such as table top interfaces (e.g. [52]).

2.1.2 Definition of automated analysis. In this review, we define automated analysis as the use of custom scripts to cluster, model or predict teaching and student behaviors. The selected case studies included some form of self-developed scripts, use of existing toolkits (such as Linguistic Inquiry Word Count [LIWC] and OpenSmile, to name a few) or custom tools adapted from existing APIs and SDKs (e.g., Kinect for Windows) to conduct analyses. Many studies also used machine learning algorithms (mainly, supervised) to predict learning outcomes or cluster classroom activities. Case studies ranged from developing algorithms to analysing and classifying face-to-face large-scale classroom data (e.g., [15]) to predicting and modelling students' behaviours within small groups or dyads (e.g., [63]).

2.1.3 Definition of physical learning analytics. The term "*physical learning analytics"* was recently coined by Martinez-Maldonado and colleagues [74] to understand the theoretical foundations of adapting similar technologies and philosophies in MMLA to co-located learning scenarios, bridging between the physical and digital spheres of a learning space. As we have outlined, the physical aspects of learning activities (e.g. team/pair dynamics, teachers' actions) must be emphasized. For this review, we focus on studies that collected and analysed data from the physical sphere (e.g. video and audio recordings of the face-to-face classroom lesson) but can be combined with data from the digital sphere (such as natural user interfaces log and digital pen strokes) in cases where learning is accompanied by such devices.

2.2 Search Process and Selection of Case Studies

We utilized a three-pronged approach to select studies for review. First, relevant journals and conference proceedings, such as journals and proceedings of known conferences targeted at showcasing advancements and innovation in the field of learning sciences and learning analytics, were shortlisted and searched using a targeted search strategy. We include the following journals: Journal of Educational Data Mining, Journal of Learning Analytics, International Journal of Learning Sciences, International Journal of Artificial Intelligence in Education, Journal of Computer Assisted Learning, British Journal of Educational Technology, and the Computers in Human Behavior, and shortlisted the following conferences: International Conference of the Learning Sciences, International Conference on Multimodal Interaction, International Conference on Learning Analytics and Knowledge, International Conference on Educational Data Mining, International Conference on Artificial Intelligence in Education, International Conference of Computer-Supported Cooperative Work, and the International Conference of Computer-Supported Collaborative Learning. We carried out the secondary search by combing through the bibliography sections of articles retrieved from the targeted search and identified additional articles. For the third and final approach, we commenced the informal search by querying Google Scholar with the following search keywords: (face-to-face OR co-located OR physical) classroom analysis AND learning analytics. We restricted our searches to articles published within 2010 to July 2018.

To map out and understand the design space of learning analytics in physical learning settings, we adopted a rather liberal inclusion/exclusion criterion to maximize the number of reviewed studies. Any peer-reviewed publication that provided accessible information on their methods, data collection set-up and metrics was included in our review. We include papers that utilize or harness at least one form of data from the physical sphere and perform or outline some sort of automated analysis of the data. Selected studies may include a combination of physical data sources and digital data sources. However, we exclude studies which harnessed only information from digital realms (e.g. solely relying on dashboard logs) from our review as they do not meet our criteria of physical data analysis as outlined in Section 2.1.2.

In total, we selected 66 studies based on our search strategy and inclusion/exclusion criteria.

2.3 Theoretical Framework

We used a variety of theoretical frameworks to categorize, analyze, and review the selected case studies. To answer our first three research questions (RQ1-3), we coded the studies along 7 dimensions: (1) *types and numbers of modalities* and (2) *data sources* harnessed, (3) the *types of outcomes* and (4) *targets of assessment* examined, (5) the *settings* and (6) *units of analysis* the in which the tool/technology was deployed, and (7) the *level of maturity* of innovation. We explain each dimension and their theoretical underpinnings in this section.

2.3.1 Modality types and data sources. The field of MMLA is aimed at harnessing multiple data modalities and combining data processing techniques to build richer, more comprehensive understanding of learning processes. MMLA researchers have explored: (1) behavioral data streams of motoric captures of head pose, eye gaze, hand gestures, facial expressions and physiological captures of heart, brain, skin and respiratory systems and (2) contextual environmental (e.g., location, weather), social (e.g. proximity) and situational (e.g., activity) data streams [72]. As it was important for our review to differentiate between data collected from the physical and digital realms, we adapted categories from the multimodal data taxonomy outlined by Di Mitri and colleagues [72] and divide the case studies based on *from whom* the data was collected (*teacher*, *student*, and *classroom*) and the *modalities of data* being collected and analyzed (*audio*, *video, biomarkers,* various combinations of *audio/video/biomarker with* or *without digital logs*).

2.3.2 Outcomes and Targets of Assessments. Based on the research outputs, we divided each case studies in terms of what learning or teaching *outcomes* were investigated. We acknowledge that most studies are early developments and deployments, hence it is challenging and impractical to define strict boundaries of stakeholders for these tools, thus we chose to define *target of assessment* (whether *teacher* or *student* or both i.e. *classroom* outcomes are being assessed).

2.3.3 Settings and Units of Analysis. Given that the purpose of our review is to understand the maturity level and design space of physical learning analytics, we were interested in the types of *settings* (*ecological* versus *laboratory*) these pipelines and tools were tested or deployed in. Most studies in the field of MMLA leveraged laboratory studies due to complex data collection setups, while learning analytics tools applied in CBESs are usually deployed in-the-wild due to the structured nature of tasks and learning paradigms. Data collection during *ecological* studies took place in-the-wild within the learning sessions, while *laboratory* studies were defined as controlled environments, often with set-ups for high-fidelity data collection and participants assigned into conditions depending on research objectives and participated in tasks related to but not part of their curriculum.

Based on different sample sizes of our selected studies, we also divided the studies according to *units of analysis* (*individuals*, *dyads*, *groups* and *classroom-level)* to understand the potential scale of these tools and technologies. *Individual* indicates data from only one student or teacher is analyzed or assessed at a single point in time while *dyads* and *groups* usually refer to pairs or groups of students being analyzed at a single point in time, and finally at *classroom-level*, data from all students and teacher(s) are collected simultaneously.

2.3.4 Level of Maturity. This dimension refers to the degree of development and deployment of the tools, technologies and pipelines in the selected case studies. Clow [71] describes the LA Cycle of four iterative steps, where *learners*' gaps and needs for technological/analytics support are first identified, then *data* is collected. After which, the data generates *metrics* and visualizations to create *interventions* to impact and influence learners. This cycle is similar to the Multimodal Learning Analytics Model (MLeAM) [72], a conceptual model to evaluate progress of MMLA tools and technologies. The MLeAM consists of four processes: *sensor capturing* which refers to the: 1) collection of multi-modal data streams, 2) manual labelling and *annotation* of data collected, 3) application of *machine learning* to develop models, generate prediction labels and validate generalizability of models on new data through iterations, and lastly, 4) *feedback interpretation* where generation of feedback drives positive behavioral change in students and teachers. Within the *metrics* stage of LA cycle and the *Annotation* and *Machine Learning* processes in MLeAM, we observe that automated analysis and metrics generation comprises of 2 different stages of maturity namely, (a) development of automated metrics and (b) generation of visualizations and insights (Table 1).

Thus, we divided the studies into 4 levels of maturity: *sensing and capturing data* where papers describe a certain data collection set-up within a small controlled sample of students with little to no analysis of data collected, *generation of automated metrics* where studies demonstrate the possibility of using automated analytic techniques to cluster certain learning and teaching processes and outcomes, *visualizations and elucidations* where studies provide insight to the learning/teaching process by elucidating how and why certain behavioral differences can account for the difference in learning gains, expertise and teaching abilities or fit developed models to new data and lastly, *provision of feedback for behavioral change* in which tools and technologies are deployed over multiple sessions and provide feedback with the aim of improving learning or teaching processes.

Table 1: Mappings between Clow's Learning Analytics Cycle, MLeAM and our proposed levels of maturity.

Learning Analytics Cycle	Our Proposed Levels of Maturity	MLeAM Processes
Learners	Sensing and capturing data	Sensor capturing
Data		
Metrics	Generation of automated metrics	Annotation + Machine Learning
	Visualizations and elucidations	
Interventions	Provision of feedback for behavioral change	Feedback interpretation

3 RESULTS AND DISCUSSION

3.1 Types of Modalities and Data Sources (RQ1)

To understand the nature of data utilized in the reviewed studies, we coded each study according to the types of modalities and types of data sources (see Table 2). Most studies extracted and analyzed data from students. Out of 66 studies, 41 (62.1%) considered data streams collected from students, 11 (16.7%) studies collected data from teachers, while the remaining 14 (21.2%) studies collected and analyzed data at a classroom level from both teacher and students.

The studies differed in the modalities of data collected from their target population, be it student, teacher or at classroom level. A total of 4 major modalities groups were identified and categorized as audio signals, video recordings, biomarkers and digital data. More specifically, the audio modality consists of classroom discourse, teacher speech, student speech; while the video modality consisted of face expressions, body movements, head pose, hand gestures and posture. The modality of biomarkers included measures such as eye tracking, EEG, electro-dermal activation and thermal signatures from heat sensors (e.g. in [9]). The last modality of digital behaviors included innovative use of location tracking from wearable badges and action logging from manikin in [18] or integration of user interfaces in curriculum such as tabletop interfaces [52] [53], digital pen strokes [34-36], presentations slides (e.g. [10], [11], [19]).

Table 2: Overview of studies concerning their use of modality and data source.

Data sources	No. and type of modalities	References
Teacher	1 (Audio only)	[5], [6], [7], [8], [15], [17], [33]
	1 (Video Only)	N/A
	1 (Biometric-only)	[12], [41]
	2 (audio and video)	N/A
	2 (audio/video + another modality)	N/A
	More or equal to 3	[42], [43]
Student	1 (Audio only)	[21], [26], [27], [64]
	1 (Video Only)	[25], [29], [44], [47], [48], [52], [57], [62]
	1 (Biometric-only)	[3], [9], [14], [24], [40], [53], [9]
	2 (audio and video)	[2], [11], [20], [30], [39], [51], [55], [56]
	2 (audio/video + another modality)	[1], [18], [28]
	More or equal to 3	[10], [19], [31], [32], [34-37], [58], [63], [65], [66]
Classroom	1 (audio only)	[13], [16], [22], [23], [38], [49], [50], [59], [60]
	1 (Video Only)	[46], [61], [62]
	1 (Biometric Only)	[4]
	2 (audio or video)	N/A
	2 (audio/video + another modality)	[45]
	More or equal to 3	N/A

Audio is more frequently explored in single modality analysis when data is collected from teachers or at the classroom level, while a wider distribution and combinations of modalities can be observed for data collected from students. For biometric measures, newer studies published from 2017 to 2018 frequently investigate EEG for single modality analysis while eye-tracking appears to be a recurrent choice for single modality analysis for earlier studies.

Notably, out of 66 papers, 26 (39.3%) of them leveraged a combination of modalities instead of focusing on only one modality. Due to the ease of collecting video and audio data in physical classrooms, we observed that amongst the 25 studies, most studies (n = 21) explored amalgamations of audio-visual features, while 2 studies attempted to harness the fusion of either audio or visual features with eye-tracking [1], [45], and the remaining two studies harnessed location sensors [19] and visual features of presentation slides [28] respectively. Amongst studies that attempt to leverage 3 or more modalities, the most common combinations were audio-visual features with digital behavior, followed by audio-visual features with biomarkers, with eye tracking being the mose frequent choice. In most studies, researchers tend to leverage a fusion of audio-visual features with digital log behaviour as equipment was easy to deploy, supplemented the learning actitivies or were existing part as the curriculum, such as a digital pen [34-36] or presentation slides, [10], [11].

3.2 Types of outcomes and assessment targets (RQ2)

To understand the research objectives in the selected studies, (RQ2), we labelled each tool, technology or pipeline in terms of the *target* that was assessed and the *types ot outcomes* (see Table 3). Some studies present overlapping scenarios assessing multiple targets and outcomes within one study. For example, we classify understanding classroom activities and instructional segments under teacher and classroom outcomes, depending on the data source and research objective of the study (e.g. [42, 43], [49]). Such studies showcase instances of pipelines and tools that automatically classify a set of classroom activities, such as lecture, group work, pair work, question-and-answer, across different classroom settings ranging from elementary school classes to university classes by modelling teachers' actions.

Amongst the case studies, the distribution of outcomes investigated is skewed towards student outcomes with the majority related to assessing students for their individual performance across different domains such as presentation quality, mathematics expertise, increase in understanding of study material (measured by pre- and post-questionnaires). We observed that in some studies, in addition to collaborative group/dyadic outcomes, researchers also analyzed and predicted individual outcomes such as expertise and performance (e.g., [34-37]). Moreover, we acknowledge that student outcomes can be a proxy for teacher assessment and an indicator of teachers' effectiveness (e.g. student attention). However, we chose to classify only studies that specifically assessed teachers on their actions, such as authenticity of their questions [17], [23] and instructional quality [12], under teacher outcomes. As such, coupled with our previous observations in Section 3.1, we notice that tools and technologies developed for face-to-face learning spaces are currently more focused on mining physical data from students and making learning processes visible so that teachers can scale and regulate their strategies according to learners' responses.

We observe a consensus that audio indicators are most useful when automatically classifying classroom activity type (e.g. in universities [38], elementary schools [60] and middle schools [16]). In contrast, indicators of performance outcomes (be it individual or collaborative) are understandably more varied. Across different contexts, differentiation of expertise has been found to correlate with the following audio (verbal and non-verbal) indicators: lesser linguistic displays of uncertainty [37], [64], and more instances of dominance [37], [51]. On the other hand, indicators of collaboration quality are largely dependent on learning contexts, but often explored via indicators of synchrony and proximity. Some studies observed physical proximity [55, 56], while others investigated synchrony in speech [26], hand or body gestures [20], [52], [62], or eye gaze [53].

Finally, despite the focus on affective computing methodologies in LA for CBESs, we note a lack of affective computing studies on students' and teachers' emotions in face-to-face physical learning environments. Of note, only one study [63] investigated affective states beyond engagement/attention and disengagement. Other studies focus on the cognitive state of attention to develop systems to help teachers more effectively monitor large classes and adapt accordingly.

Table 3: Overview of studies in terms of their targets of assessment and types of outcomes.

Target of assessment	Types of outcomes	References
Student	Student emotions, engagement and attention	[3], [14], [24], [40], [45], [47], [48], [57], [63], [66]
	Student attendance	[3], [9], [25], [29]
	Collaborative team	[14], [18], [30], [34], [35], [55], [56], [58]
	Collaborative dyadic	[14], [20], [26], [52], [53], [62]
	Individual performance	[1], [3], [10], [11]. [19], [21], [25], [27], [28], [31], [32], [35-37], [39], [41], [51], [63-66]
	Modelling student actions	[2], [18], [25], [30], [35], [44], [46], [54], [61], [63], [66]
Teacher	Instructional quality	[5-7], [12], [17], [23], [49], [50], [59], [61]
	Modelling of teacher actions	[33], [42], [43]
Classroom	Teacher-student rapport	[4], [12], [22]. [45], [46], [49], [50], [61]
	Modelling classroom activities	[9], [13], [15], [16], [38], [42], [43], [59], [60]

3.3 Settings, units of analysis and level of maturity (RQ3)

Researchers have looked at different units of analysis (i.e. dyads, groups or classroom level) across controlled laboratories studies or ecological studies (see Table 4). We noticed a balance between the use of laboratory and ecological studies. To our surprise, we reviewed a sizeable number (n = 21; 31.8%) of ecological studies at collecting and analyzing data at the classroom level. The distribution of modalities leveraged is as follows: 8 studies harnessed video analytics (i.e. classroom videos in [29], [45-48], [57], [61]), 8 studies utilised audio signals (i.e. clasroom discourse in [13], [15], [22], [23], [49], [50], [59], [60]), 2 studies leveraged biomarkers only (i.e. EEG in [4], [14]) and the remaining 4 studies utilized multimodal methods and innovations in proximitiy, motion and heat sensors [9], [42], [43]. These ecological studies and the distribution of modalities leveraged highlight the increasing potential and feasibility of applying state-of-the-art technological solutions to collect different data streams and develop automated tools for real-life, ecological classroom analysis. Even though we only observed 1 ecological dyadic study, we identified that most of the dyadic lab studies recruited actual students and administered realistic educational tasks that matched their curriculum and were evaluations of interfaces and set-ups designed for capturing signals during learning in real-time [52], [63], [65-66].

Table 4: Overview of Papers Utilizing Lab or Ecological Studies and Their Respective Units of Analysis.

Type of study	Unit of analysis	References
Laboratory	Individual	[1], [2], [32], [41], [64]
	Dyads	[20], [26], [52], [62], [63], [65], [66]
	Groups (3 to 6 people)	[18], [27], [30], [31], [34-37], [55], [56]. [58]
	Classroom/Lecture level	[3], [24], [25], [40]
Ecological	Individual	[5-8], [12], [16], [17], [21], [33], [39]
	Dyads	[53]
	Groups (3 to 6 people)	[10], [11], [19], [28]
	Classroom/Lecture level	[4], [9], [13], [14], [15], [22], [23], [29] [42-50], [57], [59-61],

To assess the maturity of the reviewed tools and technologies, we coded the outcomes of each study reviewed according to the end-goal achieved at the end of each study reviewed study (see Table 5). Studies under *Sensing and Capturing Data* often exist as descriptive papers of methodologies and position papers with informal testing and experiments within laboratory settings. These papers (3%) often serve as background papers to provide readers with comprehensive information of the innovative set-ups and pipelines employed (e.g. the set-up described in [30] was evaluated and deployed in [18] while experimental pipeline described in [45] was used in [46-48]).

Under *Generation of automated metrics,* we observed that studies often showcase the potential of leveraging, customizing and adapting existing toolkits, APIs and SDKs to mine, cluster, monitor and model student, teacher or classroom activities and behavior. Research objectives in these studies often revolve around exploring if certain low-level motoric or physiological features are useful predicting or clustering different types of performance or learning outcomes. These studies often utilized machine learning algorithms (mainly, supervised) to classify outcomes and cluster classroom activities.

We noticed several studies that attempt to develop tools and pipelines to automate human coding (and provide automated metrics of) traditionally human-annotated performance feedback (e.g. predicting instructional quality through eye tracking in [12], classroom emotional climate through low-level audio cues in [22] and authenticity of teacher's questions through audio transcriptions [23]). As the provision of feedback or visualizations are beyond the scope of these studies, they are categorized under the current level of maturity (*Automated Metrics)* instead of the subsequent levels. A total of 45 studies (68.2%) were mapped to this maturity level.

Next, the 17 studies (25.8%) that fall under the maturity level of *visualizations and elucidations* typically showcase results that differentiate and make sense of how students learn, e.g., behavioral differences in reasoning strategies in project-based learning [65] and embodied-interaction learning [1]. Importantly, these studies not only leverage data to cluster behavioral groups, they attempt to elucidate and explain motoric (in speech, face and body) and physiological differences between different levels of student or teacher competencies to make learning and teaching processes more visible. Studies in this category often attempt to answer the research question of how and why certain behavioral differences captured by the 4 different modalities (*audio*, *video*, *biomarkers* and *digital*) can account for the difference in learning gains, expertise and teaching abilities. Studies that provide meaningful visualizations of student (e.g. location and space usage visualization in [18]) and teacher activities (e.g. orchestration graph in [43]) but exclude investigations of effectiveness of such visualizations are also categorized under this level of maturity.

Finally, we categorized tools and technologies that act on generated insights and provide feedback with the intention to improve learning and/or teaching behaviors or validate and generalize developed across different populations under *Provision of Feedback and Iteration*. We only managed to shortlist 2 studies (3%). In these studies, varying types of feedback were provided and evaluated for their effectiveness for learning and/or teaching outcomes. For example, researchers in [59] investigated if providing teachers with a post-hoc automated metric of active discussion time would result in a change in teaching strategies to increase productive dialogue in the next class. Ideally, the tools and technologies under this maturity level should undergo iterative processes within and between cohorts for validation and generalizability. However, we recognize the limitation in resources to carry out iterative processes and validation studies, hence, tools tested on small populations which provided actionable feedback with the overarching aim to improve students and teachers' outcomes (e.g. [32], [59]), were categorized under this maturity level.

Table 5: Overview of the Distribution of Case Studies Across Maturity Levels.

Level of maturity	References
Sensing and Capturing Data	[30], [45]
Generation of Automated Metrics	[3], [5-13], [15-17], [19-27], [29], [31], [33], [38-40], [41], [44], [46], [48], [49], [50-53], [55-58], [60-62], [64]
Visualizations and Elucidations	[1], [2], [4], [14], [18], [28], [34-37], [42], [43], [47], [54], [63], [65], [66]

Provision of Feedback and Iteration	[32], [59]

4 DISCUSSIONS

4.1 Current Progress and Challenges (RQ4)

In our review, there was an overwhelming number of studies which focused on garnering data from students and analyzing student-related outcomes. The spotlight on learner-centric technologies and dearth of studies on teacher practice was noted by many researchers [42]. We observed that many EDM and LA studies were directed at helping teachers monitor or change students' behaviors, instead of helping teachers monitor and reflect on their teaching behaviors. This is precedent as trends of education paradigms point towards the importance of student-centric learning [72]. This is similarly reflected in how most studies within the EDM and LA community concentrated on collecting and analyzing data from students [68].

Given the unpredictability and unstructured nature of face-to-face classroom environments, it is savvy for researchers to harness the advantage of multimodality in order to more accurately model and investigate learning processes in these environments. The review reveals promising results of utilising a myriad of approaches ranging from traditional methods of audio and visual features to biomarkers and movement markers such as EEG and accelerometry. However, the small sample size present in most studies reflect how collecting data at a large scale is still significantly impeded for researchers within the realm of face-to-face classroom analysis. Perhaps this is why during our review, most ecological studies held at classroom-level, utilised only one modality, exclusively video or audio, and even EEG. The reliance on multiple modalities neccessitates complex set-ups which may interfere with learning processes, making them impractical and cumbersome to deploy in realistic learning environments.

Interestingly, within the handful of studies that leverage data from teachers, the researchers more often utilised the audio modality, than when collecting data from students. Firstly, this is in line with the research objective to assess and provide feedback to teachers' with regard to their instructional quality, as audio information contains a wealth of verbal, non-verbal, conversational and linguistic data. In addition, audio information can be easy to collect; from a lapel microphone attached to a teacher, with minimal obstruction to the teaching activities. From our personal experiences, utilising audio modality alone was sufficient at achieving an acceptable accuracy when predicting teachers' ability to facilitate classroom interactions [22]. Apart from developing better analysis and tracking techniques and sensors, future work should also concentrate on methods that support practical and sustainable data collection and explore optimal combinations of modalities most relevant for different learning environments.

Our review reveals that analytics tools geared towards visualizing and giving feedback on learning and teaching processes in a classroom are not common. Other researchers have also commented on how current and existing developments are focused at the first stage of gathering and analyzing data from the learning environment [42], [65]. Most of the reviewed studies applied state-of-the-art techniques to mine, analyze, model and cluster data of real-world classrooms. These studies have shown that subtle differences (in speech, face and body) could be automatically detected, analysed and translated into insights that make learning processes in physical classrooms more visible. For example, a study on nursing simulation [18] leveraged the pioneering use of localization sensors, in combination with a digital manikin and video recordings to detect and visualize students' actions and behaviours. From these visualizations, the researchers were able to differentiate significant behaviours that differentiated a low and high performing group.

Innovations that provide interventions and feedback derived from these distinctions to improve low-performing groups is largely underdeveloped in this field. Notably, a recent study by Ochoa and colleagues [32] harnessed audio, video and visual quality of slides to create an automated feedback system to rate students on their verbal presentations. The effectiveness of such a system was not reported. Similarly, teacher interventions guided by data insights appear to be infrequent. In fact, only one study [59] designed a subsequent teacher intervention, following insights derived from the speech discourse data of students and teacher. Another study [43] generated an orchestration graph; although its impact on teaching styles was beyond the scope of the study, such an approach underlines a way of making outputs of face-to-face classroom analysis more intepretable. Given the current state-of-the-art, these studies showcase the promising potential of harnessing learning analytics to create useful feedback and visualization systems for both students and teachers to optimize learning and teaching processes in the future.

The current progress and focus on collecting, validating, and optimzing data collection and analysis is hardly surprising as automated assessment and feedback systems have to overcome realistic, noisy, live learning environments, model the activities taking place, and accurately predict outcomes, in order to provide feedback. The challenge becomes increasingly difficult and harder to scale as the classroom size increases and activities become more complex. Currently, automated systems are sufficient for research (such as understanding logistical difficulties systems have to overcome) and knowledge building (such as understanding differences between learners, testing conceptual artifacts) but lack a certain degree of accuracy and reliability to provide automated feedback to teachers and students. Taken together, the current state-of-the-art has demonstrated possibilities of leveraging technologies for face-to-face learning analysis and holds much promise for future development in terms of developing fine-grained analyses, minimising errors, and optimizing performance measures of these automated methods.

4.2 Future Directions

We now turn the discussion to considering potential directions real-life, physical classroom analysis research can take. First, since the inception of 'Internet of Things' technology, new technologies and techniques are being devised to generate and handle volumes of big data from different devices and sources. Similarly, the adoption of devices and sensors for thermal signatures, mobility and location can be utilized in attaining deeper insights in terms of visualizing learner’s and teacher’s movements and actions, especially in ecological studies. Such methods have been preliminarily explored in [9], [18] and [43]. Due to the affordances of mobile devices, many classrooms are integrating digital devices such as tabletop interfaces and tablets, thus, incorporation of such technologies will be useful for collecting data in learning settings which utilize both digital and physical spaces such as blended learning, flipped classroom and team-based learning. In addition, we suggest more research to be geared towards understanding and exploring how adoption of devices and sensors can make learning more accessible and inclusive to people with disabilities. Despite uptake and adoption of devices and sensors in classrooms, creating and understanding learning experiences for the disabled populations remains a largely unexplored area.

Notably, the incorporation of these new technologies should not impede or make learning less accessible. At present, as observed above, the challenges associated with optimal data collection within co-located, real-life learning spaces still impede researchers and much work is required in the field to simplify and customize practical, deployable data collection processes, and methods that answer the specific needs of each type of learning environment. Depending on research and practical objectives, the incorporation of IoT devices may be more suitable for studies which look at creating visualization and feedback systems.

Within the field of online learning analytics, studies which utilize deep learning techniques, though limited, outperform the ones obtained with traditional machine learning algorithms [75]. From our review, there is ostensibly a lack of studies that utilize deep learning algorithms. This is unsurprising as the understanding of how deep learning algorithms can benefit or impact the current field in physical or online spaces has yet to be fully established. Nevertheless, deep learning techniques have been shown to be useful in the research areas of audio processing (speech or non-speech) and computer vision, notably the two modalities which researchers in this field most commonly rely on to collect data from noisy live classroom. Thus, harnessing deep learning methods could offer substantial advancements to the current field of physical classroom assessment.

As observed, the majority of reviewed studies focused on developing and validating predictors or classifiers of learning behaviours or student outcomes. However, of all the studies we reviewed, only one study [50] examined the generalizability of their developed models across different subpopulations of urban large-city and rural small-town classes. This observation is expected as we similarly noted that the small sample sizes in many studies represent the significant impediments faced by researchers. However, for models to be applicable to new, larger and more diverse data, more work has to be directed toward investigating and testing the generalizability of developed models in the future. Predictive features that apply for a certain population may not be applicable to other populations and such variations between populations differentiated by socio-economic status, race, language and region, could cause sufficient deviations from developed models, leading to unfitting conclusions and biases. This could especially be the case for emotion models, as it is well-documented that the interpretation and perception of emotions differ across different populations. We observe an emphasis on participants from Western, industrialised and democratic populations, and acknowledge that there is much room for improvement in terms of setting up more diverse shared data corpus, in order to aid transnational collaborations and cross cultural validation.

5 CONCLUSION

Taken together, our review paints a relatively young field within the EDM and LA community with regard to face-to-face classroom analysis. The case studies reviewed a strong focus on learners and promising results despite the various difficulties in collecting and analyzing noisy live classroom data. Our current review suggests that ongoing and future research in this field has many potential directions, such as moving forward to feedback and visualization systems, utilising multi-modalities in ecological studies, incorporation of mobile technologies and sensors to collect data for learning contexts that take place in digital and physical spaces, creating more deployable and sustainable set-ups for data collection, developing data corpus and models for cross-cultural validation and leveraging the power of deep learning techniques.

ACKNOWLEDGEMENTS

This project is supported by a grant from Centre for Research and Development in Learning (CRADLE@NTU).

REFERENCES

Papers Reviewed:

[1] Alejandro Andrade. 2017. Understanding student learning trajectories using multimodal learning analytics within an embodied-interaction learning environment. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17).* ACM, New York, NY, USA, 70-79. DOI: https://doi.org/10.1145/3027385.3027429

[2] Alejandro Andrade, Ginette Delandshere, and Joshua A. Danish. 2016. Using Multimodal Learning Analytics to Model Student Behavior: A Systematic Analysis of Epistemological Framing. *Journal of Learning Analytics* 3, 2, 282-306.

[3] Areej Babiker, Ibrahima Faye, and Aamir Malik. 2017. A physiological index for situational interest in classroom settings using EEG (preliminary result). In *TENCON 2017*. DOI: 10.1109/TENCON.2017.8228330.

[4] Dana Bevilacqua, Ido Davidesco, Lu Wan, Matthias Oostrik, Kim Chaloner, Jess Rowland, Mingzhou Ding, David Poeppel, and Suzanne Dikker. 2018. Brain-to-Brain Synchrony and Learning Outcomes Vary by Student–Teacher Dynamics: Evidence from a Real-world Classroom Electroencephalography Study. *Journal of Cognitive Neuroscience*, 1-11

[5] Nathaniel Blanchard, Michael Brady, Andrew M. Olney, Marci Glaus, Xiaoyi Sun, Martin Nystrand, Borhan Samei, Sean Kelly, and Sidney D’Mello. 2015. A study of automatic speech recognition in noisy classroom environments for automated dialog analysis. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*. ACM, New York, USA, 557-566. DOI: 10.1007/978-3-319-19773-9_3.

[6] Nathaniel Blanchard., Sidney D’Mello, Martin Nystrand, and Andrew M. Olney. 2015. Automatic Classification of Question & Answer Discourse Segments from Teacher’s Speech in Classrooms. In *8th International Conference on Educational Data Mining*, Madrid, Spain, 282-288.

[7] Nathaniel Blanchard, Patrick J. Donnelly, Andrew M. Olney, Borhan Samei, Brooke Ward, Xiaoyi Sun, Sean Kelly, Martin Nystrand, and Sidney K. D’Mello. 2016. Identifying teacher questions using automatic speech

recognition in live classrooms. In *17th annual meeting of the special interest group on discourse and dialogue*, 191-201.

[8] Nathaniel Blanchard, Patrick J. Donnelly, Andrew M. Olney, Borhan Samei, Brooke Ward, Xiaoyi Sun, Sean Kelly, Martin Nystrand, and Sidney D'Mello. 2016. Semi-Automatic Detection of Teacher Questions from Human-Transcripts of Audio in Live Classrooms. In *The 9th International Conference on Educational Data Mining*, 288-291.

[9] Amelia Brennan, Christina Peace, and Pablo Munguia. 2018. Classroom size, activity and attendance. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK'18)*. ACM, New York, USA, 255-259

[10] Lei Chen, Gary Feng, Chee Wee Leong, and Chong Min Lee. 2014. Using multimodal cues to analyze MLA'14 oral presentation quality corpus: Presentation delivery and slides quality. *In Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge.* DOI: 10.1145/2666633.2666640.

[11] Lei Chen, Gary Feng, Chee Wee Leong, Jillian Joe, Christopher Kitchen, and Chong Min Lee. 2016. Designing an automated assessment of public speaking skills using multimodal cues. *Journal of Learning Analytics* 3, 2, 261-281.

[12] Cortina, Kai S., Kevin F. Miller, Ryan McKenzie, and Alanna Epstein. 2015. Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. *International Journal of Science and Mathematics Education* 13, 2. 389-403.

[13] Sidney D'Mello, Andrew M. Olney, Nathaniel Blanchard, Xiaoyi Sun, Brooke Ward, Borhan Samei, and Sean Kelly. 2015. Multimodal Capture of Teacher-Student Interactions for Automated Dialogic Analysis in Live Classrooms. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15).* ACM, New York, USA, 557-566. DOI: https://doi.org/10.1145/2818346.2830602

[14] Suzanne Dikker, Lu Wan, Ido Davidesco, Lisa Kaggen, Matthias Oostrik, James McClintock, Jess Rowland, Georgios Michalareas, Jay J. Van Bavel, Mingzhou Ding, and David Poeppel. 2017. Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology* 27, 9, 1375-1380.

[15] Patrick J. Donnelly, Nathan Blanchard, Borhan Samei, Andrew M. Olney, Xiaoyi Sun, Brooke Ward, Sean Kelly, Martin Nystran, and Sidney K. D'Mello. 2016. Automatic Teacher Modeling from Live Classroom Audio. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization (UMAP '16).* ACM, New York, USA, 45-53. DOI: https://doi.org/10.1145/2930238.2930250

[16] Patrick J. Donnelly, Nathaniel Blanchard, Borhan Samei, Andrew M. Olney, Xiaoyi Sun, Brooke Ward, Sean Kelly, Martin Nystrand, and Sidney K. D'Mello. 2016. Multi-sensor modeling of teacher instructional segments in live classrooms. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI 2016).* ACM, New York, USA, 177-184. DOI: https://doi.org/10.1145/2993148.2993158

[17] Patrick J. Donnelly, Nathaniel Blanchard, Andrew M. Olney, Sean Kelly, Martin Nystrand, and Sidney K. D'Mello. 2017. Words matter: automatic detection of teacher questions in live classroom discourse using linguistics, acoustics, and context. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17).* ACM, New York, USA, 218-227. DOI: https://doi.org/10.1145/3027385.3027417

[18] Vanessa Echeverria, Roberto Martinez-Maldonado, Tamara Power, Carolyn Hayes and Simon B. Shum. 2018. Where Is the Nurse? Towards Automatically Visualizing Meaningful Team Movement in Healthcare Education. In *International Conference on Artificial Intelligence in Education 2018.* Springer, Cham, 74-78.

[19] Vanessa Echeverría, Allan, Avendano, Katherine Chiluiza, Anibal Vasquez, and Xavier Ochoa. 2014. Presentation skills estimation based on video and kinect data analysis. In *Proceedings of the 2014 ACM workshop on multimodal learning analytics workshop and grand challenge (MLA'14).* 53-60.

[20] Shuchi Grover, Marie Bienkowski, Amir Tamrakar, Behjat Siddiquie, David Salter, and Ajay Divakaran. 2016. Multimodal analytics to study collaborative problem solving in pair programming," In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16).* 516-517, DOI: 10.1145/2883851.2883877.

[21] Arzharkyn Izbassarova, Aidana Irmanova, and Alex P. James. 2017. Automated rating of recorded classroom presentations using speech analysis in kazakh. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI).* IEEE. 393-397

[22] Anusha James, Victoria Y. H. Chua, Tomasz Maszczyk, Ana M. Nunez, Rebecca Bull, Kerry Lee and Justin Dauwels. 2018. Automated Classification of Classroom Climate by Audio Analysis. In *International Workshop on Spoken Dialog System Technology 2018,* Singapore.

[23] Sean Kelly, Andrew M. Olney, Patrick Donnelly, Martin Nystrand, and Sidney. K. D'Mello. 2018. Automatically Measuring Question Authenticity in Real-World Classrooms. *Educational Researcher*, 13189.

[24] Li-Wei. Ko, Oleksii Komarov, David W. Hairston, Tzyy-Ping Jung, and Chin-Teng Lin. 2017. Sustained Attention in Real Classroom Settings: An EEG Study. *Frontiers in Human Neuroscience* 11, 388.

[25] Jian H. Lim, Eng Y. Teh, Ming H. Geh, and Chem H. Lim. 2017. Automated classroom monitoring with connected visioning system. In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC).* DOI: 10.1109/APSIPA.2017.8282063.

[26] Nichola Lubold and Heather Pon-Barry. 2014. Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Chellenge,* 5-12. DOI: 10.1145/2666633.2666635.

[27] Saturnino Luz. 2013. Automatic identification of experts and performance prediction in the multimodal math data corpus through analysis of speech interaction. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI '13).* ACM, New York, USA, 575-582 DOI: 10.1145/2522848.2533788.

[28] Gonzalo Luzardo, Bruno Guamán, Katherine Chiluiza, Jaime Castells, and Xavier Ochoa. Estimation of presentations skills based on slides and audio features. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge,* 37-44

[29] Yunxiang Mao, Haohan Li, and Zhaozheng Yin. 2014. Who missed the class? Unifying Multi-Face Detection, Tracking and Recognition in Videos. In *2014 IEEE International Conference on Multimedia and Expo (ICME).* 1-6.

[30] Roberto Martinez-Maldonado, Tamara Power, Carolyn Hayes, Adrian Abdiprano, Tony Vo, Carmen Axisa, Simon B. Shum. 2017. Analytics meet patient manikins: Challenges in an authentic small-group healthcare simulation classroom. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)*. ACM, New York, USA. DOI: 10.1145/3027385.3027401.

[31] Xavier Ochoa, Katherine Chiluiza, Gonzalo Mendez, Gonzalo Luzardo, Bruno Guaman and James Castells. 2013. Expertise estimation based on simple multimodal features. In *Proceedings of the 2013 ACM on International Conference on Multimodal Interaction (ICMI '13)*. ACM, New York, USA, 583-590. DOI: 10.1145/2522848.2533789.

[32] Xavier Ochoa, Federico Dominguez, Bruno Guaman, Ricardo Maya, Gabriel Falcones and James Castells. 2018. The RAP System: Automatic Feedback of Oral Presentation Skills Using Multimodal Analysis and Low-Cost Sensors. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. ACM, New York, USA, 360-364.

[33] Andrew M. Olney, Borhan Samei, Patrick J. Donnelly and Sidney K. D'Mello. 2018. Assessing the Dialogic Properties of Classroom Discourse: Proportion Models for Imbalanced Classes. In *Proceedings of the 10th International Conference on Educational Data Mining,* 162-167.

[34] Sharon Oviatt. 2013. Problem solving, domain expertise and learning: Ground-truth performance results for math data corpus. In *Proceedings of the 15th ACM on International conference on multimodal interaction (ICMI '13).* ACM, New York, USA, 569-574. DOI: 10.1145/2522848.2533791.

[35] Sharon Oviatt and Adrienne Cohen. 2013. Written and multimodal representations as predictors of expertise and problem-solving success in mathematics. In *Proceedings of the 15th ACM on International conference on multimodal interaction (ICMI '13).* ACM, New York, USA. DOI: 10.1145/2522848.2533793.

[36] Sharon Oviatt and Adrienne Cohen. 2014. Written activity, representations and fluency as predictors of domain expertise in mathematics. In *Proceedings of the 16th ACM on International Conference on Multimodal Interaction (ICMI '14).* ACM, New York, USA, 10-17. DOI: 10.1145/2663204.2663245.

[37] Sharon Oviatt, Kevin Hang, Jianlong Zhou and Fang Chen. 2015. Spoken interruptions signal productive problem solving and domain expertise in mathematics. In *Proceedings of the 16th ACM on International Conference on Multimodal Interaction (ICMI '14)*. ACM, New York, USA, 311-318. DOI: 10.1145/2818346.2820743.

[38] Melissa T. Owens et al. 2017. Classroom sound can be used to classify teaching practices in college science courses. In *Proceedings of the National Academy of Sciences of the United States of America* 114, 12, 3085-3090. DOI: 10.1073/pnas.1618693114

[39] Hector J. Pijeira-Díaz, Hendrik Drachsler, Sanna Jarvela and Paul A. Kirschner. 2016. Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16)*. ACM, New York, USA, 64-73. DOI: 10.1145/2883851.2883897.

[40] Andreas T. Poulsen, Simon Kamronn, Jacek Dmochowski, Lucas C. Parra and Lars K. Hansen. 2017. EEG in the classroom: Synchronised neural recordings during video presentation. Scientific Reports 7, 1, 43916.

[41] Luis P. Prieto, Yun Wen, Daniela Caballero, Kshitij Sharma and Pierre Dillenbourg. 2014. Studying teacher cognitive load in multi-tabletop classrooms using mobile eye-tracking. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14).* ACM, New York, USA, 339-344. DOI: 10.1145/2669485.2669543.

[42] Luis P. Prieto, Kshitij Sharma, Pierre Dillenbourg and Maria J. Rodriguez-Triana. 2016. Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16).* ACM, New York, USA, 148-157. DOI: 10.1145/2883851.2883927.

[43] Luis P. Prieto, Konika Sharma, Łukasz. Kidzinski, Maria J. Rodríguez-Triana, and Pierre Dillenbourg 2018. Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of Computer Assisted Learning* 34, 2, 193-203.

[44] Jinxian Qin, Yaqian Zhou, Hong Lu and Heqing Ya. 2015. Teaching video analytics based on student spatial and temporal behavior mining. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval (ICMR' 15).* ACM, New York, USA, 635-642. DOI: 10.1145/2671188.2749357.

[45] Mirko Raca and Pierre Dillenbourg. 2014. Classroom Social Signal Analysis. *Journal of Learning Analytics* 1, 3, 176-178.

[46] Mirko Raca, Roland Tormey, and Pierre Dillenburg. 2013. Student motion and it's potential as a classroom performance metric. In *3rd International Workshop on Teaching Analytics (ECTEL'13).*

[47] Mirko Raca, Roland Tormey and Pierre Dillenbourg. 2016. Sleepers' lag - study on motion and attention. *Journal of Learning Analytics* 3, 2, 239-260. DOI: 10.1145/2567574.2567581.

[48] Mirko Raca, Lukasz Kidzinski, and Pierre Dillenbourg. 2015. Translating Head Motion into Attention - Towards Processing of Student's Body-Language. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM2015).* International Educational Data Mining Society, 320-326.

[49] Borhan Samei, Andrew M. Olney, Sean Kelly, Martin Nystrand, Sidney K. D'Mello, Nathaniel Blanchard, Xiaoyi Sun, Marci Glaus, and Art Graesser. 2014. Domain Independent Assessment of Dialogic Properties of Classroom Discourse. In *Proceedings of 7th International Conference on Educational Data Mining (EDM2014).* International Educational Data Mining Society, 233-236.

[50] Borhan Samei, Andrew M. Olney, Sean Kelly, Martin Nystrand. 2015. Modeling Classroom Discourse: Do Models that Predict Dialogic Instruction Properties Generalize across Populations? In *Proceedings of 7th International Conference on Educational Data Mining (EDM2015).* International Educational Data Mining Society.

[51] Stefan Scherer, Nadir Weibel, Louis-Philippe Morency, and Sharon Oviatt. 2012. Multimodal prediction of expertise and leadership in learning groups. In *Proceedings of the 1st International Workshop on Multimodal Learning Analytics (MLA'12).* DOI: 10.1145/2389268.2389269.

[52] Bertrand Schneider and Paulo Blikstein. 2015. Unraveling Students' Interaction Around a Tangible Interface Using Gesture Recognition. In *Proceedings of 7th International Conference on Educational Data Mining (EDM2015).* International Educational Data Mining Society, 320-323.

[53] Bertrand Schneider, Kshitij Sharma, Sebastien Cuendet, Guillaume Zufferey, Pierre Dillenbourg, and Roy Pea. 2016. Detecting Collaborative Dynamics Using Mobile Eye-Trackers. In *Proceedings of the 12th International Conference of the Learning Sciences*, 522-529.

[54] Carmen Smith, Barbara King, and Diana Gonzalez. 2016. Using Multimodal Learning Analytics to Identify Patterns of Interactions in a Body-Based Mathematics Activity. *Journal of Interactive Learning Research* 27, 4, 355-379.

[55] Daniel Spikol, Emmanuele Ruffaldi, Lorenzo Landolfi, and Mutlu Cukurova. 2017. Estimation of success in collaborative learning based on multimodal learning analytics features. In *Proceedings of 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*. DOI: 10.1109/ICALT.2017.122.

[56] Daniel Spikol, Emmanuele Ruffaldi, Giacomo Dabisias and Mutlu Cukurova. 2018. Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning* 34, 4, 66-77.

[57] Chinchu Thomas and Dinesh Babu Jayagopi. 2017. Predicting student engagement in classrooms using facial behavioral cues. In *Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education (MIE 2017)*. ACM, New York, USA, 33-40. DOI: 10.1145/3139513.3139514.

[58] Kate Thompson. 2013. Using micro-patterns of speech to predict the correctness of answers to mathematics problems: An exercise in multimodal learning analytics. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI '13).* ACM, New York, USA, 591-598. DOI: 10.1145/2522848.2533792.

[59] Zuowei Wang, Kevin Miller and Kai S. Cortina. 2013. Using the LENA in Teacher Training: Promoting Student Involvement through automated feedback. *Unterrichtswissenschaft* 4, 4.

[60] Zuowei Wang, Xingyu Pan, Kevin Miller and Kai S. Cortina. 2014. Automatic classification of activities in classroom discourse. *Computers & Education 78*, 115-123.

[61] Eiji Watanabe, Takashi Ozeki, and Takeshi Kohama. 2018. Analysis of Interactions Between Lecturers and Students. *In Proceedings of International Conference on Learning Analytics and Knowledge (LAK'18).* ACM, New York, USA, 370-374 https://doi.org/10.1145/3170358.3170360

[62] Andrea S. Won, Jeremy N. Bailenson, and Joris H. Janssen. 2014. Automatic Detection of Nonverbal Behavior Predicts Learning in Dyadic Interactions. In *IEEE Transactions on Affective Computing* 5, 2, 112-125.

[63] Marcelo Worsley and Paulo Blikstein. 2015. Using learning analytics to study cognitive disequilibrium in a complex learning environment. In *Proceedings of the 5th Annual Conference on Learning Analytics and Knowledge (LAK'15)*. ACM, New York, USA, 531-532. DOI: 10.1145/2723576.2723659.

[64] Marcelo Worsley and Paulo Blikstein. 2011. What's an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis. In *Proceedings of 4th International Conference on Educational Data Mining (EDM2011).* International Educational Data Mining Society, 235-240.

[65] Marcelo Worsley, Stefan Scherer, Louis-Philippe Morency, and Paulo Blikstein. 2015. Exploring Behavior Representation for Learning Analytics. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15).* ACM, New York, USA, 251-258.

[66] Marcelo Worsley. 2018. (Dis)engagement matters: identifying efficacious learning practices with multimodal learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18).* ACM, New York, USA, 365-369.

Papers Cited:

[67] Ryan S. Baker, and Paul S. Inventado. 2016. Educational data mining and learning analytics: Potentials and possibilities for online education. In G. Veletsianos (Ed.), *Emergence and Innovation in Digital Learning*, 83–98. doi:10.15215/aupress/9781771991490.01

[68] Matthew Berland, Ryan S. Baker, and Paulo Blikstein. 2014. Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning* 19, 1-2, 205-220.

[69] Paulo Blikstein, and Marcelo Worsley. 2016. Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics* 3, 2, 220-238.

[70] Orlando B. Coelho, and Ismar Silveira. 2017. Deep Learning applied to Learning Analytics and Educational Data Mining: A Systematic Literature Review. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*.

[71] Doug Clow. 2012. The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK'12)*. ACM, New York, USA, 134-138.

[72] Daniele Di Mitri, Jan Schneider, Marcus Specht, Hendrik Drachsler. 2018. From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning* 34, 4, 338-49.

[73] Kenneth R. Koedinger, Sidney K. D'Mello, Elizabeth A. McLaughlin, Zachary A. Pardos, and Carolyn P. Rose. 2015. Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science* 6, 4, 333-353.

[74] Roberto Martinez-Maldonado, Vanessa Echeverria, Olga C. Santos, Augusto D. P. D. Santos, and Kalina Yacef. 2018. Physical learning analytics: a multimodal perspective. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18).* ACM, New York, USA, 375-379.

[75] Xavier Ochoa and Marcelo Worsley. 2016. Augmenting Learning Analytics with Multimodal Sensory Data. *Journal of Learning Analytics* 3, 2, 213-219

[76] Roberto Martinez-Maldonado, Simon B. Shum, Bertrand Schneider, Sven Charleer Joris Klerkx, Erik Duval. 2017. Learning Analytics for Natural User Interfaces. *Journal of Learning Analytics* 4, 1, 24-57.

[77] Alejandro Peña-Ayala. 2014. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications* 41, 4, 1432-1462