Possaghi *et al. Smart Learning Environments* (2025) 12:53
https://doi.org/10.1186/s40561-025-00410-4

Smart Learning Environments

RESEARCH **Open Access**

Integrating multi-modal learning analytics dashboard in K-12 education: insights for enhancing orchestration and teacher decision-making

Isabella Possaghi[1*], Boban Vesin[3], Feiran Zhang[2], Kshitij Sharma[1], Cecilie Knudsen[1], Håkon Bjørkum[1] and Sofia Papavlasopoulou[1]

*Correspondence:
isabella.possaghi@ntnu.no

[1] Universitetet i Trondheim: Norges teknisk-naturvitenskapelige universitet, Trondheim, Norway
[2] School of Design, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR
[3] Department of Business, History and Social Sciences, USN School of Business, Vestfold, Norway

Abstract

Technological advancements are transforming teaching methods while offering wider windows into students' learning journeys. Multi-modal Learning Analytics Dashboards (LADs) are tools that facilitate smart classroom orchestration by aggregating and analyzing students' responses through sensors, such as facial expressions and heart rate, for real-time insights into student engagement and emotional states. In this study, we developed an LAD for open-ended activities in K-12 settings, where orchestration is non-linear and poses challenges for standardized evaluation methods. We engaged end users (e.g., educational researchers) in the process from the early design stages and investigated the feasibility of the LAD when used in the wild. The results show how affective data support greater awareness of students' experiences, improving teachers' orchestration through better decision-making and agency. Roadblocks were also identified regarding data interpretability, students' privacy, and additional teacher workload, which can limit adoption and should be carefully addressed in future implementations. Further research should investigate students' responses more closely and further develop strategies for the responsible, explainable, and unbiased use of student affective data in real classrooms.

Keywords: Smart orchestration, Learning analytics, Teacher-facing dashboard, Multi-modal data

Introduction

Across the span of a decade, technology for educational environments has evolved, transforming teaching approaches and unraveling learning processes. A key example is the rise of smart classroom orchestration, where digital tools, such as dashboards, empower teachers to effectively manage in-classroom experiences. Dashboards usually aggregate and communicate learning indicators to help teachers with high-level classroom monitoring and tailored facilitation. To pinpoint such indicators, dashboards can utilize Learning Analytics (LA) to process sensor data (e.g., face detection (Giannakos et al., 2019; Emerson et al., 2023), gaze direction (Lee-Cultura et al., 2023), and heart rate (Lee-Cultura et al., 2023; Possaghi et al., 2024; Malmberg et al., 2019)) collected from students, providing real-time insights into their learning experiences. When data from multiple streams (or "modalities") are combined in the analysis, the term Multi-Modal (MM) LA is used (Ochoa et al., 2022). Sensor-driven Learning Analytics Dashboards (LADs) can provide highly detailed knowledge acquisition, but also students' affective responses. Emotional data can be used to uncover, and subsequently support, students' Self-Regulated Learning (SRL) processes during activities (Zimmerman, 2013; Tissenbaum And Slotta, 2019).

Shifts in K-12 pedagogy amplify LADs' relevance for classrooms. Specifically, when the call is for flexible support during open-ended learning activities with learning trajectories that are less predictable and more individualized (Reiser, 2013; Van Mechelen et al., 2023). This shift is prominent in STEM subjects, including computer science, where instruction is framed around creative knowledge-building modules that reflect constructionist principles (Ah-Nam And Osman, 2017). While offering many benefits, the non-linear nature of constructionist activities often conflicts with standardized assessment (Giannakos et al., 2020). Moreover, many current evaluation protocols are inherently subjective, frequently relying on self-reported records (e.g., post-activity interviews). In response, LADs offer a less biased alternative by enabling informed decisions based on students' actual experiences rather than self-perceptions (Spikol et al., 2017; Calvo and D'Mello, 2010). Moreover, constructionist activities often intersect with digital tools (e.g., personal computers, tablets) to support artifact building. Aside from fostering students' creative expression, these tools generate a great deal of data (e.g., facial recordings from PC camera) that can be analyzed to gain additional insights into the learning interaction (Ochoa et al., 2022). When extracted and processed through LA, such data, paired with students' affective states, secure a set of MM indicators for gaining unbiased insights into open-ended learning experiences, where traditional methods often fall short (Cukurova et al., 2020). Despite the documented potential, the real-time capturing of student experiences in authentic educational settings remains under-explored (Giannakos et al., 2021; Schwendimann et al., 2016; Järvenoja et al., 2018), most teacher-facing LADs rarely progress beyond the prototyping stage (Susnjak et al., 2022; Alfredo et al., 2024), and they are seldom evaluated through a qualitative lens (Jivet et al., 2018).

Motivated by this gap, we developed and evaluated a teacher-facing LAD enabling the collection, analysis, and signaling of eight affective states (e.g., engagement, stress, happiness) of students. The proposed LAD leverages MM data analytics to automatically monitor students' states in real-time while they engage in collaborative constructionist activity, with programming modules, in a school setting. The LAD development is structured around a user-centric design methodology. In doing so, instructors (i.e., teachers and supporting teachers) and educational researchers actively drafted, designed, implemented, and validated the LAD. The engagement of these practitioners aimed to move beyond purely technical advancements, with the promise of delivering a usability-sound technology along with actionable insights for future developments (Mangaroska and Giannakos, 2018; Verbert et al., 2020). However, introducing LADs to classroom practice has some hurdles, as teachers' beliefs, prior experiences, and comfort with educational technology influence their acceptance and effective use of such tools

(Martinez-Maldonado, 2023). For this study, we sought to understand how instructors experienced incorporating the proposed LAD for affective response in a K-12 classroom setting by addressing the following research questions (RQs):

1. *What benefits or positive impacts did instructors perceive from using the LAD?*
2. *What challenges or drawbacks did instructors face when using the LAD?*
3. *Which design considerations for LADs emerge as critical for informing future, real-world classroom implementations?*

To answer these RQs, we conducted an evaluation of the proposed LAD in two K–12 classrooms. This evaluation focused on its practical implementation from the instructors' perspectives, based on qualitative insights triangulated with students' questionnaire responses. In doing so, we close the user-centric design loop and offer empirical insights to inform the future development of educational technologies aimed at enabling smarter classroom orchestration. We summarize our contribution as follows:

(a) We propose a real-time LAD as an intelligent tool designed to enhance K-12 in-classroom monitoring by automatically analyzing and responding to students' affective states.
(b) We present both qualitative and quantitative insights from an in-the-wild study, collecting students' experiences with data sensors, and teachers' impressions on the LAD.
(c) We provide research and practical implications from our findings, focusing on design-based directions for further improvement and validation of LADs and their in-classroom use.

Background and related work

With Fig. 1 we visually represent the context of our research, positioned as a subset of three concentric areas of existing literature: the field of classroom orchestration with technology, the more specific domain of descriptive dashboards tailored for K-12 teachers, and the challenge of capturing and displaying affective states.

Classroom orchestration with technologies

As described by Dillenbourg, "orchestration" in education is a complex process of real-time process bound by constraints (Dillenbourg, 2013). Those constraints are either "intrinsic" constraints, which refer to the concepts taught, the learners' profile, and the knowledge acquisition workflow, or "extrinsic", which characterize the learning context (e.g., schedules, settings, and assessment techniques). Currently, K-12 curricular guidelines favor subject-specific training within a constructionist framework, which is known for its non-objectivist nature (O'Connor, 2022). This is particularly evident in the field of STEM, where programming and other computational literacies are fostered through a multidisciplinary approach that promotes real-world problem-solving (Johnson et al., 2016). In this context, Dillenbourg's concept of extrinsic constraints becomes more pronounced because of open-ended learning environments typical of collaborative

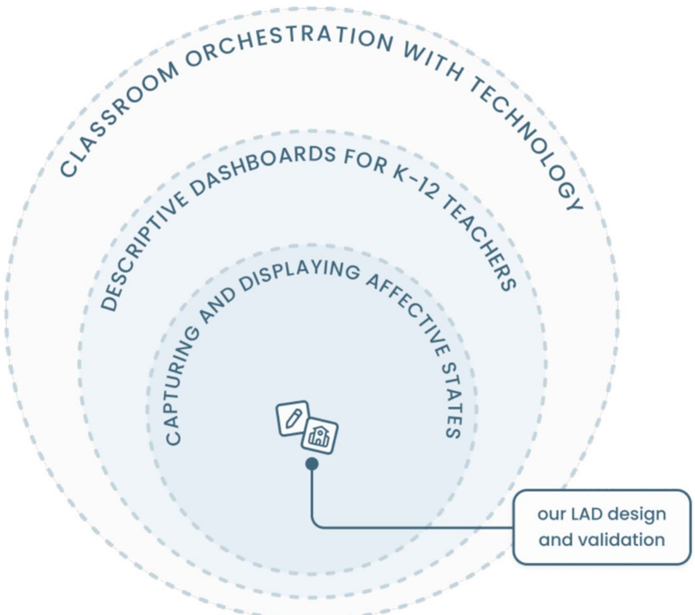


Fig. 1 Our study (LAD design and validation) at the intersection of classroom orchestration, K-12 descriptive dashboards, and the capture of affective states

constructionist settings (Dillenbourg, 2013; Martinez-Maldonado et al., 2014). For instance, constraints in assessment exacerbate challenges in monitoring and evaluation, especially when the process is inherently dynamic and the outcomes are not predetermined. As Amarasinghe et al. (2021) pointed out, mapping open-ended tasks places higher cognitive demands on teachers. As a result, teachers may prioritize immediate educational goals (e.g., artifact creation) over paying attention to students' internal states (e.g., motivation, engagement) that greatly influence their learning experiences (Amarasinghe et al., 2021). Moreover, while constructionism emphasizes student-centered, collaborative learning journeys fostering autonomy, teachers must not be marginalized as secondary figures. Instead, they should shift from being knowledge transmitters to facilitators who promote students' ownership of their learning (Hoover, 1996), leveraging technological insights to inform their decision-making (Martinez-Maldonado, 2019).

To alleviate orchestration burdens, technological efforts have focused on giving teachers access to students' live activity feeds (Holstein et al., 2018), time-reliant automated scoring (Tissenbaum et al., 2016), and tools for mistake tracking (Mangaroska et al., 2021). However, these measures primarily specialize in linear processes and fall short when confronted with open-ended learning dynamics. This poses the risk of forcing activities into overly structured protocols, compromising experiences given inflexible orchestration, and resulting in the limitation of learner-driven processes (Tissenbaum And Slotta, 2019; Dillenbourg et al., 2009). Moreover, the very same introduction of orchestrational technologies and modern data collection protocols considerably tolls teachers, potentially driving them to compromise or reevaluate their pedagogical beliefs (Ertmer et al., 2012). To mitigate this, Shahmoradi et al. (2024) validated an orchestrational dashboard tailored to a robotic constructionism activity in a primary school setting, addressing contextual factors (e.g., student literacy) and its relationship with teachers' perceived usability (Shahmoradi et al., 2024). From the teachers' perspective, Kessel et al. (2025) investigate which background and contextual demands primarily impact dashboard use to update facilitators' curriculum and proficiency programs for long-term LAD implementation in schools (Kessel et al., 2025). While these and many more studies validate the significance of such technologies, a critical gap persists in integrating everyday facilitation approaches with diverse stakeholder perspectives into their design from the outset. As Martinez-Maldonado (2023) reports, underestimating educational stakeholders' expertise can lead to designs unfit for authentic in-class use, or an overemphasis on technical development that discounts user experience (Martinez-Maldonado, 2023).

While a 2018 lature review on LADs identified user experience's assessment lacking in qualitative perspective, but still treated dashboard acceptance as a secondary research focus (Jivet et al., 2018), the field's perspective has significantly evolved. Scholarship now puts human responses upfront from the earliest stages of conceptualization in LAD research (Verbert et al., 2020). For this study, we align with these contemporary views, asserting *"human value"* in design loops for effective digital tool implementation, directly addressing end-users' needs (Mangaroska et al., 2021; Ahn et al., 2019). Consequently, our study employs an iterative user-centered design approach, incorporating qualitative insights through focus groups and semi-structured interviews. This methodology facilitates a continuous dialogue with key stakeholders (e.g., researchers, classroom teachers, and instructors) to iteratively develop orchestrational technologies that genuinely reflect their pedagogical beliefs and the emergent needs of constructionist educational activities, mirroring best practices seen in studies similar in scope (Alfredo et al., 2024; De Vreugd et al., 2024).

Descriptive dashboards for K-12 teachers

Among different technologies for orchestration, teacher-facing LADs peak in popularity with four key aspects (Amarasinghe et al., 2022; Van Leeuwen et al., 2019). First, these LADs mirror the classroom environment by narrating events in a digestible format, thereby providing teachers with straightforward insights for classroom monitoring. For example, complex data are translated into accessible graphical representations, helping teachers pinpoint events that might otherwise remain obscure (Pozdniakov et al., 2023). This can further be done beyond simply displaying performance metrics, such as activity scores. When MM data streams are captured, visual representations highlight patterns in cognitive, behavioral, and affective domains, revealing students' strengths and weaknesses that traditional observation methods may overlook (Cukurova et al., 2020; Giannakos et al., 2019). Second, these LADs, augmented with awareness mechanisms, such as visual or auditory cues, can even offer actionable insights to teachers (Roberts et al., 2017). Aside from achieving better management and less cognitive effort in interpreting depicted information (e.g., which student is challenged), such affordances are particularly useful for teachers with limited digital literacy (Martinez-Maldonado et al., 2020). Research findings support that dashboards with enhanced features, such as data storytelling elements, are particularly beneficial for teachers less acquainted with data visualization, granting better-informed decisions with minimal effort (Pozdniakov et al., 2023). By combining machine-computed cues with a mirrored LAD, teachers gain a deeper understanding of student performance and can pinpoint areas where intervention might be necessary. Third, LADs that incorporate real-time notifications based on student responses can provide teachers with timely insights. For example, Knoop-van Campen et al. (2023) offer a real-time classroom overview covering both punctual tasks and students' overall progress. Results showed teachers moving away from outcome-oriented feedback and instead emphasizing students' learning processes without over-focusing on low-performing students. Knoop-van Campen et al. (2023); d'Anjou et al. (2019). Last, descriptive LADs can facilitate comparisons across different units, such as between individuals or teams, thus providing teachers with valuable insights for those who may need assistance or intervention (Susnjak et al., 2022; Giannakos et al., 2020; Li et al., 2020). As seen, previous research showed that these teacher-facing LADs can inform pedagogical decision-making, granting high-level classroom monitoring and diagnosing (Martinez-Maldonado, 2019; Holstein et al., 2019; Van Leeuwen et al., 2015; Bao et al., 2021), enabling ad hoc guidance for students (Cukurova et al., 2020; Spikol and Cukurova, 2020).

Despite growing interest in LADs and robust research in LA within education, their sustained adoption in real classrooms remains limited. This reluctance stems from an insufficient understanding of their value, persistent issues of trust and ethics, and a critical lack of integrated practitioner perspectives guiding their development (Paulsen and Lindsay, 2024; Martinez-Maldonado, 2023). To counteract this tendency, prior studies have involved teachers in co-developing LADs. For example, *TEADASH*, a LAD integrated into Canvas, was co-designed with teachers to ensure its applicability in university-level engineering and social science courses (Nguyen et al., 2024). Another instance is *TeamSlide*, a LAD based on MM data, which involved senior teaching staff in its design to support teacher-guided reflection in a university nursing course (Echeverria et al., 2024). While most LADs are designed for use in higher education, one recent study by Mohseni et al. (2023) explored conceptual design considerations for LADs in the K-12 context. In this study, teachers were engaged in interviews and co-creation of prototypes; however, a subsequent evaluation of these LADs in an authentic classroom setting was not conducted (Mohseni et al., 2023). As Verbert et al. (2020) advocate, the validation of usability holds equal importance to the fulfillment of functional requirements in order to minimize design pitfalls (Verbert et al., 2020). Building on this principle and addressing the named gaps, LAD's potential can be leveraged to create a versatile and informed learning environment within the K-12 education context by carefully designing and empirically evaluating its effectiveness, explicitly incorporating teachers' needs and voices during open-ended learning activities.

Capturing and displaying affective states

Multiple factors concur in learning processes, going beyond the sole cognitive dimension. SRL theory, for instance, posits that students experience specific affective states as they regulate their practices to achieve learning goals (Zimmerman, 2000). Affective states encompass emotions and feelings, described as *"bodily responses"* (D'mello And Jensen, 2017) while learning. Unlike social and cognitive factors in orchestration guidelines, affective states are often overlooked despite their significance, partly due to the difficulties in identifying them (Bosch et al., 2015). Moreover, Dillenbourg et al. (2009) contend that such emotional dynamics evolve alongside computer-supported collaboration and SRL due to the use of new technologies, generating new needs for both learners and teachers (Dillenbourg et al., 2009). Nonetheless, joint efforts between the learning design and LA communities call for various methods for capturing affective responses, moving away from observation-based mapping (Ocumpaugh, 2015) in favor of more valid protocols that utilize face-based detectors [58] or wearable sensors (Giannakos et al., 2020), among others (Sharma and Giannakos, 2021).

Extrinsic constraints also impact SRL processes. Wolters et al. (2005) state that *"contextual features in the environment"* (Wolters et al., 2005) can significantly modify students' responses, regulating, for example, effort when help is available. Educational technology contributes to this extent, since the growing reliance on digital tools in education creates both a need for their validation and new opportunities to access diverse data channels (Järvelä et al., 2016; Nasir et al., 2021). Converging these multiple data streams in analysis can unravel how pupils' learning journey unfolds and why (Cukurova et al., 2020), enabling a shift from subjective metrics to more germane pieces of evidence on SRL (Järvelä et al., 2021). As mentioned, LA has been widely researched to retrieve clues on regulatory processes from knowledge acquisition and collaboration perspectives. Authors such as Spikol et al. (2017) and Nasir et al. (2021)propose that MMLA can unlock unique information about open-ended activities such as project-based learning. Face, hand tracking, and speech ratio have been utilized to outline components such as the distance between peers, hand gestures, dialogues, and gaze direction, which can describe the inquiry style in teamwork (Spikol et al., 2017; Nasir et al., 2021). Despite promising prospects for uncovering the complexity of learning journeys, particularly in the context of constructionism activities, empirical examples centered on distilling affective responses remain scarce. An exception to this is the work by Sedrakyan et al. (2020), who designed feedback mechanisms embedded in LADs to inform students or teachers through visualizations of process-oriented regulation of learning. This work retraces the importance of the socioemotional context as expressed by Dillenbourg et al. (2009), emphasizing the need to further explore the incorporation of sensor data into LADs to enhance decision-making and testing the quality of biofeedback for emotional monitoring and regulation (Sedrakyan et al., 2020).

Embedding MM data streams in LADs creates powerful touchpoints that transcend the research community and extend to all stakeholders involved in educational experiences. However, the realization of LADs' theoretical potential hinges on their design, with accents on necessary literacy and general sustainability for a long-term deployment (Cukurova et al., 2020). For this reason, research needs to evaluate the design of such dashboards in close consultation with stakeholders to ensure the fulfillment of users' needs and avoid overlooking the human element (Buckingham Shum et al., 2019; Amarasinghe et al., 2021; Cukurova et al., 2020; Yoo And Jin, 2020; Sharma and Giannakos, 2021). Ouhaichi, Holstein, and colleagues emphasize this need, noting a lack of LAD validation in *"authentic settings"* (Ouhaichi et al., 2023) and insufficient value cultivation with real practitioners (Holstein et al., 2018). Despite the theoretical advantages of LADs, MM measurements, and state-of-the-art sensors, bridging the gap between theory and real-world classroom implementation, especially considering teachers' pedagogical beliefs, remains challenging (Wiley et al., 2024; Holstein et al., 2018; Giannakos et al., 2021). Given that the empirical emphasis on this domain concentrates on higher education (Bond et al., 2023), our work contributes by extending this investigation to K-12 settings. This presents an opportunity to provide concrete evidence of affective states' significance in students' learning experiences, which should be explored through in-the-wild interventions (Kaur And Chahal, 2024).

Dashboard design and implementation

Being aware that the creation of LADs entails a multidisciplinary approach encompassing Human-Computer Interaction (HCI), software development, learning, and user experience design (Mangaroska et al., 2021; Xie et al., 2019), we referred to the *LATUX* (Learning Awareness Tools - User eXperience) workflow to integrate stakeholders' perspectives in the development of our LAD. As proposed by Martinez-Maldonado et al. (2015), this workflow is a tailored process for developing awareness and assistance enhancers for educational environments with respect to teachers' needs (Martinez-Maldonado et al., 2015). Moreover, we draw from the steps followed by user-centered methodologies from literature similar in scope (Holstein et al., 2019, 2017; Lee-Cultura et al., 2023; Yoo And Jin, 2020). The following paragraphs describe the design steps enacted to create our LAD.

Initial guidelines for LAD design

As a preliminary step to our workflow (Fig. 2), we conducted a literature review to identify significant research in LA and sensor data usage within educational environments. This review grounded our research gap in empirical implications and solidly framed our contribution. By investigating the state-of-the-art, we were able to develop initial design guidelines for our LAD. We summarized 21 relevant papers, considering the measurements considered, sensor data employed, communication methods (e.g., what to display and how), and ethical recommendations. The resulting guidelines are reported in detail in Table 6 in the Appendix.

After analyzing the selected literature, educational researchers were involved in the process of selecting the most crucial requirements. Efforts were chosen to be directed toward a design for: *a.* accessing a real-time overview of students' affective states at the classroom level via a real-time analysis of their experiences, *b.* intuitive signaling system, *c.* possibility of narrowing down the monitoring at the group or individual level, and *d.* customization of the spatial layout. Moreover, the system's interoperability and usability were valued for creating a dashboard that can easily integrate with other platforms

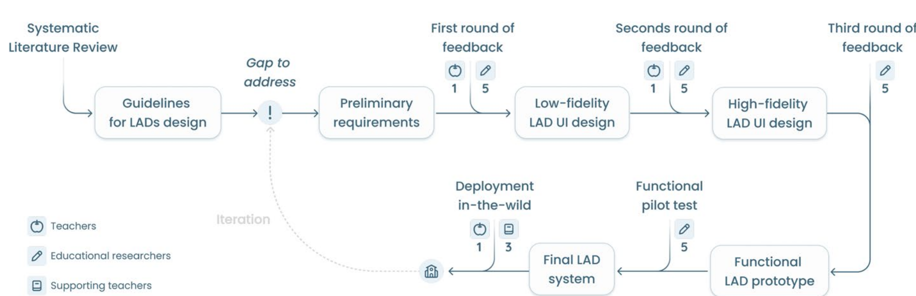


Fig. 2 The LATUX workflow adapted to our research

and does not require technical competence for management beforehand. Finally, the literature review showed that research involving K-12 stakeholders (i.e., students and teachers) remains marginalized, and when this population is addressed, studies focused primarily on students' experiences (Bond et al., 2023). This frames our contribution to providing a more holistic approach, including also the teachers' perspectives, as valuable.

Sensors and measurements chosen

The systematic summarization informed our directions in choosing sensor data and expected measurements. Our LAD integrates different sensors to collect MM data. We employ the Empatica E4 wristband to gather Electrodermal Activity (EDA) via accelerometer (ACC) and Heart Rate Variability (HRV) to retrieve measurements about students' physiological stress and engagement (Lee-Cultura et al., 2021). Along with wristband sensors, we employed the laptop's embedded camera to collect students' facial expressions (Sharma et al., 2022). This was done by running a Python package called PyEmotion to map students' affective states through facial features. Via these data, we extract measurements defining seven emotions: anger, fear, happiness, sadness, surprise, disgust, and a neutral state. The consideration of this batch of emotions is grounded in D'Mello and Gressner's implications on SRL (D'mello and Graesser, 2012), as well as their consistent reference in the literature (Possaghi et al., 2024; Sharma et al., 2022; Pekrun, 2006).

The first round of feedback

To create a usable digital touch point to bridge facilitation with MMLA, we engaged authentic stakeholders from the earliest stages of the LAD design. The session lasted two hours, and we collected notes on the open discussion among the participants and two facilitators. Firstly, we shared system images illustrating the proposed MMLA's interface and functionality with a K-12 teacher and five educational researchers. By presenting these visuals, we aimed to elicit valuable insights regarding educators' specific needs and challenges in their daily practices (First round of feedback in Fig. 2). Participants were then prompted to discuss the requirements drawn from the literature and their preliminary incorporation into the LAD's system mockups. This phase resulted in four design decisions.

- *Real-Time Affective Monitoring* Instead of performance outcomes (e.g., exercise correctness), the decision was to display students' affective states in real-time. The goal is to monitor students' positive or negative responses to the educational experience to prevent lowered performance and support teachers' fast action in adapting facilitation.
- *Reactive Descriptive System* We chose to develop a reactive system. Namely, a descriptive system mirrors in-classroom events and signals to the teacher if some students' status needs attention according to analyzed data in real-time. While developing a predictive system for affective states based on computed data would be beneficial, educational researchers value the opportunity to uncover students' emotional states that often go unrecognized through mere observation. In fact, teachers seek dashboards that collect, organize, and present classroom information clearly and

meaningfully without requiring extensive UI interpretation. Descriptive LADs can assist educators in processing this information.

- *Plain Signaling with visual cues* We explored whether the system should offer specific suggestions or alerts for students in critical states. Ultimately, both the teacher and educational researchers favored a descriptive LAD with plain signaling without prompts, allowing teachers to assess emotional states while still providing the flexibility to determine the most appropriate course of action. For example, pictorial notifications as dashboard feedback redirect the teacher's attention toward students facing challenges without embedding textual information, which could lead to content overload and misinterpretation.
- *Seating Layout Correspondence* Ensure direct correspondence between the class and the displayed layout. This facilitates the teacher's understanding of the mapping between students on the dashboard and their real-world seating arrangements. We opted against embedding data visualizations due to teachers' potential difficulty interpreting charts and graphs (Pozdniakov et al., 2023). Instead, we chose to use pictorial references for more straightforward communication.

Second round of feedback on a low-fidelity prototype

These design decisions guided the development of the low-fidelity LAD prototype, as shown in Fig. 3. Following this, we conducted another session with the same K-12 teacher and educational researchers to facilitate an iterative refinement of the system, referred to as the *Second Round of Feedback* in Fig. 2. This dashboard prototype, developed using *Figma*, was presented to the stakeholders to validate the design choices made thus far. Again, based on notes from the discussion, the insights gathered from this session can be categorized into three main action points:

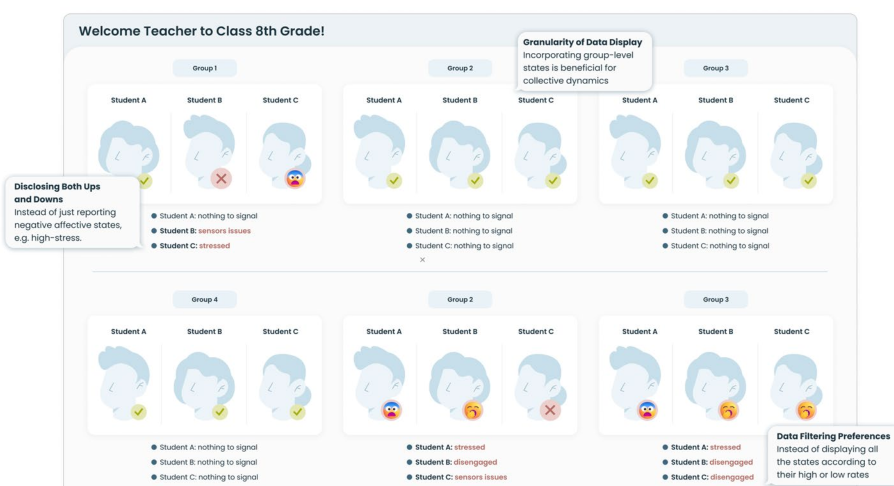


Fig. 3 Low-fidelity prototype with feedback

- *Granularity of Data Display:* The low-fidelity prototype reflects the physical classroom layout, prioritizing the display of affective states at the group level. However, the teacher and educational researchers expressed a desire to navigate between different levels of granularity, allowing them to switch from aggregated data at the group level to more detailed insights at the individual level.
- *Data Filtering Preferences:* Educational researchers emphasized the need to filter which data from the various affective states can be displayed on the LAD. This capability would enable them to concentrate on critical insights without feeling overwhelmed by excessive information.
- *Disclosing Both Ups and Downs:* Our design initially focused solely on signaling changes in negative affective states, specifically highlighting adverse events. However, the teacher also strongly supported disclosing positive affective states, such as high engagement. This addition would enrich the understanding of emotional dynamics, allowing facilitators to recognize when challenges arise and when students are thriving.

Third round of feedback on a high-fidelity prototype

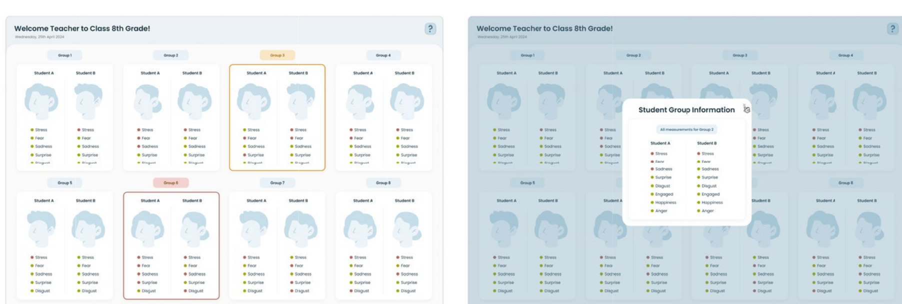
After an agreement on the UI was reached to finalize its high-fidelity, the *Figma* prototype was updated, as shown in Fig. 4.

The front page

The front page enables the teacher to customize the dashboard layout. The number of students and the number of students per group (thus interacting with the same shared device) can be customized. This feature creates a stronger parallelism between the actual classroom and the layout shown, easing the teacher's interpretation of what is displayed on the screen (d'Anjou et al., 2019). The display of measurements allowed or prevented in the dashboard can also be customized to manage content display and prevent information overload (Sedrakyan et al., 2020).

The class view

The teacher can access the customized class view once preferences are submitted. Students are represented as pictorial icons configured accordingly, with the selected



(a) Class view with group signaling (b) Group's measurements check

Fig. 4 The finalized dashboard design measurements displayed below. To access all measurements at any time during the activity, the group can be clicked to open an overlay window displaying the complete set of measurements without going back to the configuration panel Fig. 4b. The red and green icons next to each measurement vary according to the student's response (as calculated in the back end) to signal the teacher that attention is needed. For example, if a student has abnormally high stress values for a specific time window, the icon for the "stress" measurement will turn red. The circle will change to green when the measurement falls within normal values. The signaling is also enabled at the group level. Here, the communication is no longer binary (red and green) but is based on three levels of attention required, determined by the number of measurements exceeding predefined thresholds for the two students combined. Namely, no visual cue (fewer than 4 measurements), an orange visual cue (between 5 and 6 measurements), and a red visual cue (7 or more measurements), as illustrated in Fig. 4a.

The high-fidelity prototype was subjected to a final UI testing by educational researchers for its navigation, which was indicated as the Third round of feedback in our process. The outcome defined the user journey through the various layers of interactions as minimal to serve its purpose Fig. 5. Specifically, users can easily switch between the front page (measurements and group configurations) and the class view. They can access an overlay window to visualize the complete set of measurements from a group's participants.

Technical implementation and preliminary pilot testing

Following the LAD's UI design and user experience consolidation, development proceeded from a technical standpoint. The final system architecture comprises three core components: *a. Frontend*, *b. Backend*, and *c. Database*. The following section provides a detailed account of the LAD's assembly and its technological specifications.

Frontend

The application's frontend serves as a dashboard interface for the end user, typically a teacher, which dynamically adapts based on sensor data stored in the database. The UI design was described in previous sections. *React* with *TypeScript* was utilized to build the framework, facilitating smooth deployment and enhanced readability for developers.

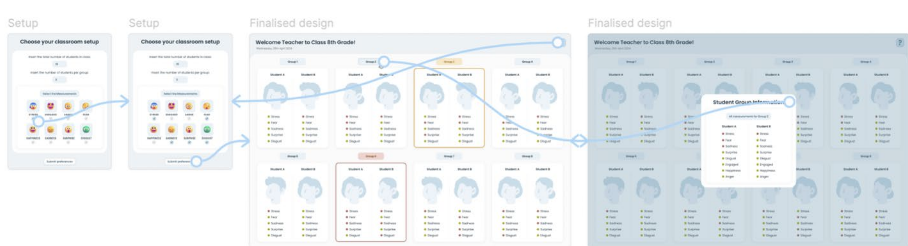


Fig. 5 User interface and user experience flow

Backend

The backend is responsible for real-time data processing, managing sensor connectivity, executing measurement calculations, and handling data aggregation and signaling. The interaction between the backend and the database is illustrated in Fig. 6. *Python* was employed with the *Django* framework due to its ease of use and robust server-side capabilities. A detailed description of how the backend manages sensor data is provided below.

Data from wristbands: The Empatica E4 streaming server was used to process wristband data in real-time. The connection between Empatica wristbands and the laptop running the Django framework was established via Bluetooth Low Energy (BLE). The Django backend maintained an independent socket connection with each wristband, ensuring that any issues encountered with one device did not affect others. The Django application processed signals received from the wristbands, and features such as EDA were used to calculate stress and engagement levels in real-time. Each stress or engagement calculation was sent to the database, allowing for real-time teacher signaling and data storage for later analysis.

Data from the camera Facial recognition was used to map students' emotional states through the PyEmotion package. This process continuously computed the emotions of individuals seated in front of the computer, classifying their emotional state into one of the seven aforementioned emotions (Possaghi et al., 2024; Sharma et al., 2022; Pekrun, 2006). Since emotional mapping happened in real time, video recordings were not stored for later analysis, ensuring anonymity by avoiding the use of participants' facial capture. Like stress and engagement data, emotional state information was sent to the database for storage and immediate signaling purposes. The *PyEmotion* package was modified to return multiple emotions and allow an analysis of emotional feedback from multiple students using a single camera, enabling the system to process emotional data for students working in pairs.

Data processing Besides raw data collection, the system aggregates data in real-time to alert the teacher if any metrics appear unusually high or low. A five-second window was chosen for data retrieval from the database to determine the appropriate signals based on the values collected. *PyEmotion* processed numerical data from the *Empatica E4* wristbands. Namely, the mean and standard deviation were calculated for the entire session to establish a baseline for stress and engagement, which were then compared with the mean of recent measurements. For effective states computed via *PyEmotion*, the analysis focused on data from the previous five minutes (Possaghi et al., 2024; Tisza et al., 2022; Sharma et al., 2021), calculating the fraction of time spent in each emotional

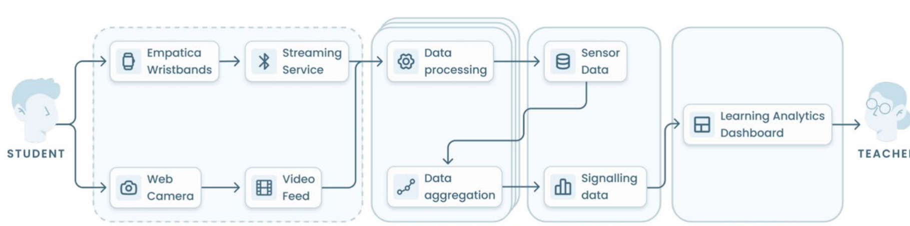


Fig. 6 Architecture overview

state. A signal was sent to the teacher if the fraction of a particular emotion exceeded or fell below a predefined threshold. Each backend instance directly communicated with the database to maintain data integrity, eliminating the need for API endpoints.

Database

System interoperability was achieved through the Firebase database, which served as the central repository for both raw and aggregated data. Each participant was assigned a unique ID to prevent personal data collection and facilitate subsequent analysis. As depicted in Fig. 6 was organized into two main nodes: raw sensor data and aggregated signaling data for each student. Each backend event and its timestamp were saved in the raw data node, enabling the correlation of specific events, such as elevated stress levels, with specific learning activities.

Five educational researchers conducted a pilot evaluation of the functional LAD prototype to ensure system robustness. The session lasted approximately two hours and was designed to recreate an in-school scenario. The evaluation focused on identifying issues related to time management, technology orchestration, and activity delivery, intending to refine the system for future in-class validation.

Validation and performance evaluation in-the-wild

As a final step of our process (Fig. 2), we conducted a case study in a real-world classroom setting to assess the challenges and opportunities of deploying the MMLA dashboard and demonstrate its feasibility. The school was informed beforehand about the nature of the intervention, and participants' approval through parent or guardian consent was requested. Sikt (Norwegian Agency for Shared Services in Education and Research) approved our intervention, data collection, and data handling plan.

Participants

The proposed activity engaged two classes with a total of 44 students who were in the 8th grade from a secondary school in the Vestfold region of Norway. The students were either 13 or 14 years old (M=13.39, SD=0.49), and identified themselves as either boy (27), girl (16), or other (1). The context of the open-ended group activity is described below. It is worth mentioning that although some participants had prior experience with in-school programming, none of them were familiar with the specific platform or activity that we proposed. Given the limited supply of wristbands, 32 of the 44 randomly selected students who were participating in the activity wore them during the activity. Facilitators provided the groups wearing wristbands with a computer equipped with a camera for facial data collection and installed the Empatica Web Application. No employment of school devices was required. The class activity was inclusive, and everyone could participate without being part of the data collection (by not wearing wristbands and avoiding camera capture) if they had not been provided consent by their parents or guardians, without affecting their engagement. Students were free to withdraw their assent to data collection verbally at any time. The facilitation team, consisting of one teacher (who was mainly responsible for the class) and three supporting teachers, was present for both classes. The LAD was presented to all instructors before the lesson started to familiarize them with it.

A design thinking activity on sustainability

The activity was in the form of an activity and lasted six hours, spanning over three days for each of the two classes, and was held in the classroom setting, following the school's regular schedule. Students who participated in the activity were divided into dyads. The topic of this Design Thinking activity covered the importance of recycling and framing it as a sustainability issue for the community. The activity followed the five phases of a Design Thinking process (*empathize, ideate, define, prototype, and test*) and was structured based on the $Exten.(D.T.)^2$ activity plan[1] Using a block-based coding platform, we posed the accent on the prototype phase. Specifically, the platform *GearsBots*[2] for robotic simulations was leveraged to create a parallelism between a gamified and real-life experience to achieve awareness of the recycling challenge. The activity's structure encouraged collaboration among team members and between teams, allowing for peer feedback and prompting iterative improvements on the proposed design solutions in *GearsBots*.

Method

For this in-class intervention, we used a mixed-method research design, combining the elements of quantitative and qualitative research, with emphasis on the latter. The study was qualitatively driven, as our primary aim was to understand the instructors' experience in depth (Hesse-Biber et al., 2015). The whole DT experience was considered for assessment. Still, it was more focused during the *prototype* phase. This phase involved real-time LAD testing as students participated in a programming task, which was at the core of the DT experience and required extended collaborative computer interaction.

Data collection

The quantitative strand comprised a *thirteen-item individual questionnaire* using a Likert scale with values from 1 = strongly disagree to 5 = strongly agree. It was administered post-activity to evaluate students' *(a)* response toward the experience, *(b)* perceptions of wearing the wristband, and *(c)* reactions to the presence of the camera. The questionnaire items are listed in Tables 1, 2, and 3. Since not all students experienced wearing the *Empatica* wristband, the items concerning perceptions of wearing the wristband *(b)* were answered by only 32 students out of 44, while the other questionnaire items were answered by all students.

For the qualitative strand, insights from the facilitation perspective were gathered through in-depth *semi-structured interviews* with the teachers (T1) and supporting teachers (ST1, ST2, ST3) conducted after the three-day intervention. The interviews lasted between 1.5 and 2 h. The semi-structured format allowed additional topics to emerge while maintaining focus on topics relevant to the research scope (Yin, 2009). Additionally, feedback was collected after each day of the intervention through

[1] European Project Extend(DT)2: https://extendt2.eu/.

[2] https://gears.aposteriori.com.sg/

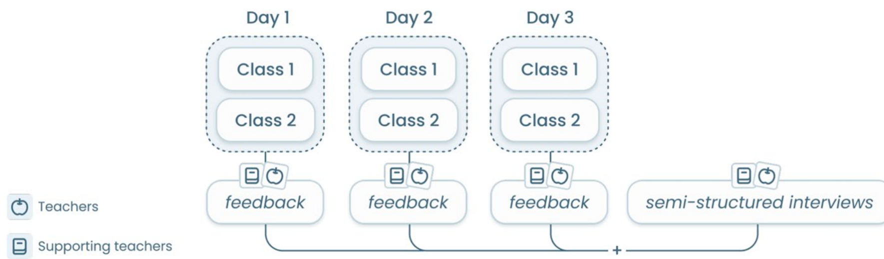


Fig. 7 Data collection for qualitative data

Table 1 Descriptive statistics for the set of items *a*, with a scale from 1 to 5

Post-activity questionnaire: set of items *a*.	Mean (SD)
1. The activity was exciting	3.15 (1.09)
2. The activity was fun	3.06 (1.06)
3. The activity was useful	3.33 (1.14)
4. The activity was difficult	3.44 (1.11)
5. I was stressed during the activities	2.07 (1.18)
6. I was engaged during the activities	3.34 (1.23)

Values in parentheses show the standard deviation (SD) from 44 students

Table 2 Descriptive statistics for the set of items *a*, with a scale from 1 to 5

Post-activity questionnaire: set of items *b*.	Mean (SD)
7. I was bothered by the wristband during task solving	2.13 (1.09)
8. I liked the learning process more with the use of smart wristbands	3.14 (1.05)
9. I could use a smart wristband on more occasions related to learning	3.50 (1.02)
10. I behaved differently because of the smart wristband	2.00 (1.24)

Values in parentheses show the standard deviation (SD) from 32 students

Table 3 Descriptive statistics for the set of items *a*, with a scale from 1 to 5

Post-activity questionnaire: set of items *c*.	textbfMean (SD)
11. I was bothered by the web camera during task solving	1.89 (1.21)
12. I could use a web camera on more occasions related to learning	2.43 (1.25)
13. I was bothered by the thought that the teacher could see what I felt	2.49 (1.19)

Values in parentheses show the standard deviation (SD) from 44 students

1-hour-long joint discussions among the teachers (T1) and supporting teachers (ST1, ST2, ST3), as illustrated in Fig. 7. Such feedback notes were included as part of the data corpus, complementing the insights from the interviews. Combining multiple qualitative data sources helped us address our research questions on the LAD's effectiveness in the learning environment as an auxiliary tool for teaching, and provided a comprehensive perspective on monitoring the process.

Table 4 Spearman correlations between set of items *a.* and set of items *b*

Set of items *a.*	**Item 7**	**Item 8**	**Item 9**	**Item 10**
Exciting (Item 1)	0.25 ($p = 0.16$)	**0.55 (p<0.01)**	**0.41 (p=0.02)**	−0.06 ($p = 0.72$)
Fun (Item 2)	0.28 ($p = 0.12$)	**0.54 (p<0.01)**	**0.37 (p=0.04)**	−0.11 ($p = 0.56$)
Useful (Item 3)	0.21 ($p = 0.25$)	**0.44 (p=0.01)**	0.22 ($p = 0.23$)	−0.04 ($p = 0.83$)
Difficult (Item 4)	−0.02 ($p = 0.92$)	−0.34 ($p = 0.06$)	−0.08 ($p = 0.67$)	−0.08 ($p = 0.66$)
Stressed (Item 5)	−0.21 ($p = 0.25$)	−0.04 ($p = 0.83$)	0.03 ($p = 0.86$)	0.23 ($p = 0.21$)
Engaged (Item 6)	0.24 ($p = 0.18$)	**0.49 (p<0.01)**	0.25 ($p = 0.16$)	−0.14 ($p = 0.45$)

Bold indicates $p < 0.05$

Data analysis

For the students' questionnaires, we analyzed the responses, calculating the means, scoring on a scale ranging from a theoretical maximum of 5 to a minimum of 1. As aforementioned, items related to students' response to the experience *(a)* and reactions to the presence of the camera *(c)* were calculated on the basis of responses from 44 students, while items concerning perceptions of wearing the wristband *(b)* were calculated on 32 students. Next, we used a Spearman correlation analysis to identify relationships between these sets of items. Specifically, we calculated correlations between students' responses to the experience *(a)* and perceptions of wearing the wristband *(b)*, as well as between students' responses to the experience *(a)* and reactions to the presence of the camera *(c)*. We opted for Spearman's rank correlation since the 5-point Likert-scale questionnaire produced ordinal data. Moreover, the Shapiro-Wilk test indicated deviation from normality (see details in the Appendix Tables 7, 8, 9).

Regarding the semi-structured interviews, we approached the data inductively to allow patterns to emerge directly from the data. Recorded dialogues from the teacher (T1) and supporting teachers (ST1, ST2, ST3) were transcribed and subjected to a thematic analysis (Peel, 2020). During this process, meaningful segments of dialogue were organized into preliminary categories through descriptive coding. These categories were then refined iteratively through thematic coding, resulting in a set of eight finalized themes.

Results

Students questionnaires

As a first step, we provide a statistical description of the students' responses to the post-activity individual questionnaire. Results in terms of means are reported in Tables 1, 2, 3. As seen, students wearing the wristband during tasks rated their level of distraction as mild on average. Additionally, they reported a higher level of enjoyment when wearing the wristband. When asked whether their behavior would have changed without the wristbands, students' responses suggested a slight inclination toward affirming that it might not have made a significant difference. Students expressed a mild concern about the mounted camera tracking their facial expressions during their learning experience. Finally, a low mean describes the average response when asked if bothered by the possibility of the teacher getting an insight into their feelings.

In Table 4 , we present Spearman correlation coefficients between students' responses to the experience *(a)* and students' perceptions of wearing the wristband *(b)*. The results show that Item 8 is positively correlated with Item 1 ($\rho = 0.55, p < 0.01$), Item 2

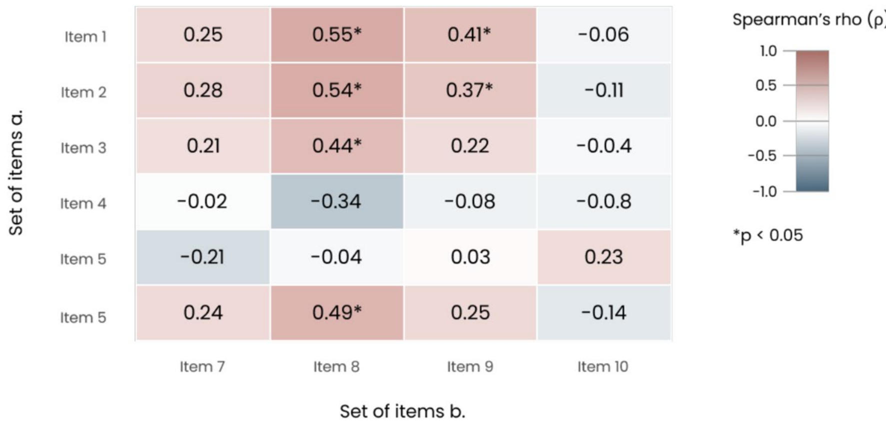


Fig. 8 Heatmap of correlations between students' activity perceptions and wristband-related responses

($\rho = 0.54, p < 0.01$), Item 3 ($\rho = 0.44, p = 0.01$), and Item 6 ($\rho = 0.49, p < 0.01$), indicating that higher scores on these positive experience perceptions tend to align with higher enjoyment of the experience when wearing the wristband. Item 9 shows weaker but still significant positive correlations with Item 1 ($\rho = 0.41, p = 0.02$) and Item 2 ($\rho = 0.37, p = 0.04$), suggesting some relationship between excitement and fun and the willingness to use the wristband in learning activities. No significant correlations were observed between Item 7 and Item 10, despite Item 7 showing a trend toward a negative correlation with the perceived difficulty of the task (Item 4), approaching significance ($\rho = -0.34\ p = 0.06$). The heatmap in Fig. 8 visually summarizes the results.

Next, we report in Table 5 the correlations' results between participants' self-reported experiences during the activity, set of items *(a)*, and their perceptions related to webcam use, set of items *(c)*. Participants who rated high in positive responses (Item 1, Item 2, Item 3, and Item 6) were significantly more prone to use a webcam on future learning occasions (Item 12). For example, fun (Item 2) showed a strong positive association with willingness to use the webcam again ($\rho = 0.64$, $p = 2.9e{-}6$), as did excitement (Item 1; $\rho = 0.54$, $p = 0.00017$), perceived usefulness (Item 3; $\rho = 0.43$, $p = 0.0035$) and engagement (Item 6; $\rho = 0.57$, $p = 4.6e{-}5$). By contrast, no significant correlations were found between positive activity experiences and feeling bothered by the webcam (Item 11) or concerned about being observed emotionally by the teacher (Item 13), except for a weaker positive link between "fun" and Item 13 ($\rho = 0.32$, $p = 0.033$). The heatmap in Fig. 9 visually summarizes the results.

Teachers perspectives

The following sections detail the eight themes from the recorded interviews with the teacher (T1) and supporting teachers (ST1, ST2, ST3), hereafter referred to as either *"teachers"* or *"facilitators"*. The main findings are classified under themes, as seen below.

Reflecting the facilitators' intuitions

During the activity, facilitators could compare their intuitions with the insights displayed by the LAD. One teacher expressed contentment when realizing that the dashboard

Table 5 Spearman correlations between set of items *a.* and set of items *c*

Set of items *a.*	Item 11	Item 12	Item 13
Exciting (Item 1)	0.16 ($p = 0.29$)	**0.54 (p= 0.00017)**	0.19 ($p = 0.23$)
Fun (Item 2)	0.16 ($p = 0.31$)	**0.64 ($p = 2.9e-6$)**	**0.32 ($p = 0.033$)**
Useful (Item 3)	−0.03 ($p = 0.84$)	**0.43 ($p = 0.0035$)**	0.22 ($p = 0.15$)
Difficult (Item 4)	−0.08 ($p = 0.59$)	−0.15 ($p = 0.35$)	0.09 ($p = 0.56$)
Stressed (Item 5)	−0.15 ($p = 0.34$)	−0.05 ($p = 0.75$)	0.21 ($p = 0.16$)
Engaged (Item 6)	0.13 ($p = 0.40$)	**0.57 (p= 4.6e −5)**	0.28 ($p = 0.06$)

Bold indicates $p < 0.05$

allowed for a correct and accurate in-time representation of students' states: *"I wanted to verify with you if the one blinking red on the dashboard was the exact student I was suspecting to be disengaged and the data was correct!"* (T1). Guided by the dashboard, one teacher gave feedback to students flagged as disengaged, even when they denied it if questioned about it. After providing guidance, the dashboard indicated a rise in engagement: *"After providing feedback and instructions, the dashboard data confirmed a rise in his interest in the task! It was super nice to have an input!"* (T1). Moreover, one teacher underscored the value of mapping students' states in real-time, not just after the finished activity, allowing her to access a complete overview of the classroom's experience: *"It was exciting, and [the mapping] was extremely correct!"* (T1). One facilitator also noted that although the LAD signals where attention is needed, *"we still maintain authority over how to intervene if someone is struggling with the task"* (ST2).

Interaction footprints

The interaction with the system granted the facilitators access to students' digital footprints. This affordance prompted reflective dialogues throughout the activity, as facilitators noted: *"We had a starting point to reflect with students on why they were engaged or not, using it as a foundation to start a conversation about their experiences"* (ST1 and ST2). One teacher brought up the increasing information requirement on how students perform and act in the school environment. This would expand the scope of automated

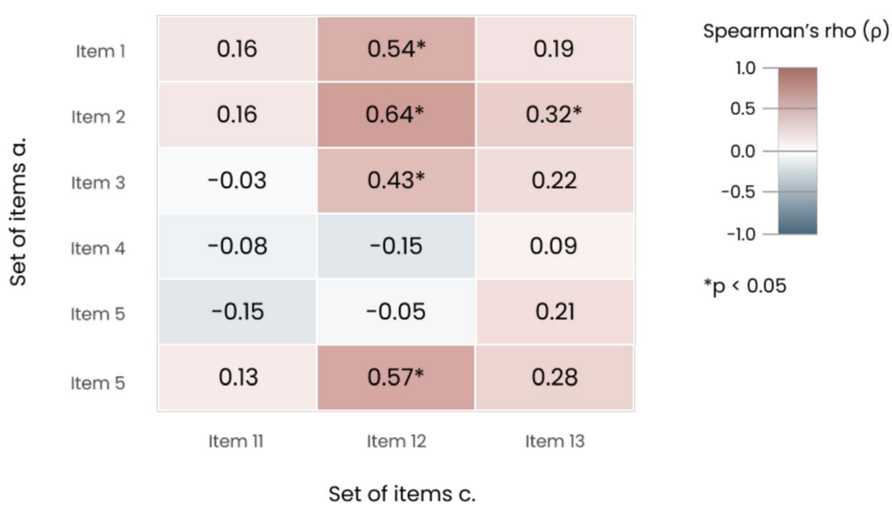


Fig. 9 Heatmap of correlations between students' activity perceptions and webcam-related responses

data collection, making it a possible supplement to existing documentation: *"It would be a good addition for us at the school to see from another perspective and gain an idea of how children possibly work together"* (T1). Moreover, these interaction footprints can be leveraged for student pairing and overall activities facilitation: "*We can keep track of who is excited to participate in the lesson and who is less engaged. This can inspire us to rearrange, for example, the seating layout for better conditions for collaboration"* (T1).

Hidden clues uncovered

When inquired about the range of emotions and states chosen as measurements, the main teacher described them as valuable to have a comprehensive picture of the student's experience and its changes. Among the other measurements, the main teacher indicates engagement as a valuable indicator of students' focus on the programming task. Additionally, the same teacher indicates that the LAD can be a tool for ‘ìnvestigating" students' experience with a tailored approach: *"If I want to know more about good dynamics, I can concentrate on the pupils with positive and good values and try to identify the factors contributing to their success"* (T1) Facilitators point out how the same strategy can be applied to the more disengaged and challenged pupils to pinpoint where the difficulty lies: *"In this way, I understand the problem. I compare their experiences to those of other students and see if other groups share the same issue"* (ST3).

Usability at ease

Overall, facilitators reported low effort in navigating the system and in elaborating information as a foundation for decision-making. Going more in detail, one facilitator pointed out the straightforward relationship between abstract representations in the LAD and the real-world scenario: *"[...] Even if the representation is minimal in details, I can recognize the classroom setting. It is like, you know, when you use the maps app and relate what is on the screen with what you actually see!"* (ST3). Additionally, how affective responses were shown on the LAD left little room for ambiguity, as teachers were already accustomed to managing the range of available emotions: *"Here [on the screen], it is not crowded, and I see that what you mean by engagement is close to what I mean as engagement."* (T1).

Positive change in learning strategies

The facilitators stated that embedding technological support in the classroom can boost the students' motivation and interest. For example, they believe that students are more likely to be attentive and committed to tasks when they perceive that technology plays a crucial and significant role. *"It seems that pupils stay more focused and attentive if they perceive that the activity is regulated by a "high-tech" tool. It makes it more serious for them to attend the task rigorously"* (ST3). Similarly, the main teacher witnessed that the system could inspire their students to increase their efforts, responding positively to support integrating the systems into standard real-life teaching practices: *"I trust that everything new and can help the learning process is super important, as implementing new tools for the scope"* (T1).

Meaningful energy and time allocation

According to our interview, facilitation can not be completely revolutionized. For example, teachers will still need to circulate the classroom to observe student engagement: *"If you are a teacher, you will walk around in the classroom. It is not possible for you to just stand at the front because you need to see what pupils are up to"* (T1). However, both the teacher and supporting teachers commented on how the LAD can provide insights to follow to guide their interventions: *"The dashboard is an effective complement to practice since it can make the teacher have to walk less around because you know right away who you have to assist"* (T1). This enabled facilitators to allocate their time more strategically, focusing on students who needed the most support. Additionally, one facilitator stated how the LAD prevented the same students from being repeatedly approached when they were not struggling or disengaged: *"[...] sometimes students do not want to be checked on, especially if they are in the middle of a "creative" phase and they need room for elaboration"* (ST3).

Teachers can not be substituted

The teacher states that a digital tool can not solely interpret the learning process. The role of the human facilitator in the classroom can not be substituted by a dashboard only, and the teaching strategy can not be fully delegated to the machine: *"Even if it can ease the burden of monitoring the students altogether, the dashboard alone can not substitute the teacher's role"* (ST2). Moreover, the teachers expressed concerns about relying solely on digital tools for mediation, stating: *"It makes me a little worried for the future. Tools should be complementary, but humans must interpret their feedback in the classroom or individual student context"* (T1). At this stage of our technology's maturity, the teacher compared the dashboard to a young teacher assistant: *"[with your dashboard] it is like having a substitute teacher with you in the class. They are really good at reporting how pupils are doing, but the classroom's responsible teacher should double-check their observations to determine how to intervene"* (T1).

Sustainability in deployment

Both the teachers and the educational researchers pointed out the sustainability constraints regarding costs. Moreover, another facilitator expressed how more workforce would be needed in the class: *"Managing the technology and all its components alone is not straightforward"* (ST1). The concern specifically addresses the preliminary preparation (e.g., running the backend), while the sole UI interaction was deemed intuitive and free from interpretative burdens. Moreover, such expensive technology would inevitably burden children with being responsible while wearing them. However, the teacher underscores the potential instructional value in empowering students with wristband management: *"It can be an opportunity for them to learn how to be responsible in borrowing a really important piece of equipment from the school."*. One facilitator echoed this observation by adding, *"If explained how the wristbands work, learners can take ownership of their technology and even troubleshoot minor issues"* (T1), referring to the episode where one student notified the facilitator of a malfunctioning device.

Digital surveillance

Constant monitoring may result in a possible drawback to dashboard employment. Although specifying few in number, the teachers reported that some students were reluctant to wear the E4 *Empatica* wristband: *"They may have felt a little uneasy about privacy, which is why there were one or two who said no to participating because of the data gathering"* (T1). Another facilitator noted this attitude, specifying how: *"[...] children do not feel uneasy in front of the camera, but they dislike the idea of constantly being observed. This fueled higher self-consciousness, stifling, for example, the natural interaction with their team"* (ST3). Moreover, the idea of being always monitored can pressure students to feel perpetually evaluated, according to the teacher: *"Some of them were scared to show that they were not able to perform 100% correctly, even if the dashboard was not showing their performance, but just their feelings"* (T1).

Discussion

This paper presents a teacher-facing LAD that integrates multiple sensor data streams to capture, analyze, and visualize students' affective states in real time, offering a digital "window" into their learning journeys (Giannakos et al., 2019). Findings from its deployment in a classroom setting reveal both encouraging opportunities and important challenges that must be addressed to better support teachers in facilitating open-ended K-12 learning activities. In the following sections, we discuss the key opportunities and concerns identified through our study, followed by implications for future LAD design in response to the three RQs guiding this work (see "Introduction" section).

Coordinate confidence and control: unpacking LAD's opportunities for educator autonomy

To answer RQ1, the LAD raised facilitators' confidence by granting *"superpowers"* to probe into students' experiences from otherwise inaccessible angles (Holstein et al., 2019, 2017; van Leeuwen And Rummel, 2022). These findings align with prior literature, which emphasizes awareness of students' affective states as a key indicator in classroom orchestrations (Lee-Cultura et al., 2023; Papamitsiou et al., 2020). In our case, this awareness stemmed from the pairing of emotion-related data streams with their temporal transparency, reinforcing targeted support (Sung et al., 2023). Teachers involved in in-the-wild evaluation positively commented on sensor data accuracy in measuring affective responses (Giannakos et al., 2019): the LAD's reliability in displaying students' states was unanimously attested, particularly in identifying disengaged learners in a timely manner. Such immediate insights were seen as affirming teachers' intuitions, thereby enhancing their self-confidence and sense of agency in classroom orchestration. Specifically, the alignment between intuitions and LAD's feedback validated the *"effectiveness of their [teachers] own help-giving"* (Holstein et al., 2017), finding their instructional decisions in pieces of evidence. Referring to Holstein et al. (2018) orchestration tutoring agenda (Holstein et al., 2019), the LAD enhanced facilitators' sense of agency, enabling more effective resource allocation for students' management and adjusting their support based on identified moments of challenge. In practice, assistance was evenly distributed throughout the classroom, rather than being directed solely towards low-performing groups (Knoop-van Campen et al., 2023). For instance, facilitators expressed enthusiasm as the engagement icon turned green, following encouragement for students to adopt new programming solutions, substantiating the LAD's utility in informing real-time action (Sedrakyan et al., 2020; Gao et al., 2020; Papamitsiou et al., 2020).

It is well documented that learners voice themselves in diverse ways (Lee-Cultura et al., 2023; Holstein et al., 2019), and our study offers a similar perspective. For example, when challenged, students may struggle to communicate or even miscommunicate through spoken words, leading to a flawed understanding of their emotional state. MM data gathered via physiological and facial decoding can help address the lack of (or inauthentic) verbal and emotional expression without pressuring students to alter their behavior (Giannakos et al., 2020). On the other hand, data must be critically interpreted to promote a more intentional, self-directed approach to both collaboration and individual student progress. Relying exclusively on raw quantitative information, oftentimes, leads to oversimplified conclusions that fail to capture the nuanced realities (Lee-Cultura et al., 2023). On this count, teachers who operated the LAD envisioned a potential in eliciting self-assessment from individual learners or the whole team (Kasepalu et al., 2022; Holstein et al., 2019). For example, the LAD's feedback on students' stress responses can be instrumental as reflection prompts. This meta-exercise can help students unfold the relationship between their emotional states, experiences, and performance. Furthermore, it can clarify whether the stress response was a motivating factor coupled with engagement or a contributor to frustration that slowed down their progress during the activity (d'Anjou et al., 2019). While these visions may appear contradictory, they reflect a complex classroom ecology that is amenable to effective support through the tailored application of the LAD (Graesser, 2020; D'Mello et al., 2014). These affordances resonate with implications from Holstein et al. (2018) on the importance of granting a high degree of information interpretability (Holstein et al., 2019). In our case, teachers could contextualize the LAD's feedback within the classroom culture without interfering with their situational understanding (An et al., 2020). This interpretive flexibility empowered facilitators to decide on the optimal timing and approach for interventions, respecting students' individual pacing in their creative processes, minimizing unnecessary interruptions (Sung et al., 2023). Ultimately, this balance between data-driven insights and contextualized pedagogy can foster a more holistic approach to orchestration.

We posited our design on two premises: first, that feedback-displaying LADs can significantly inform instructional decisions (Martinez-Maldonado, 2019; Holstein et al., 2019; Van Leeuwen et al., 2015; Bao et al., 2021), and second, that pertinent feedback does not inherently depend on traditional performance metrics (Sedrakyan et al., 2020). Our results support that alternative indicators such as biomarkers and facial data can suffice in interpreting students' SRL processes, provided their display is comprehensive (Pozdniakov et al., 2023; Martinez-Maldonado et al., 2020). Specifically, the freedom in customization through on-demand content selection (i.e., which state(s) to display on the LAD) gave teachers a finer-tuned awareness of how specific affective states influence students' learning processes. A set of measurements capturing both positive and negative affective responses afforded this insight, balancing the approach to the diverse emotions that arose during open-ended programming tasks (Sung et al., 2023). Interestingly, a teacher noted that only a few combined measurements might be adequate for this comprehension (Sharma et al., 2022), thereby reducing cognitive effort during the activity's facilitation (Sedrakyan et al., 2020). This closely relates to visualization literacy, which remains challenging to address in the LA community for seamless incorporation into authentic contexts (Pozdniakov et al., 2023). Faulty communication between the system and the user inevitably undermines effective interaction, and the LAD's contextual value becomes difficult to grasp. In our validation, insights from engaged facilitators suggest a promising outlook from a usability perspective, despite their low self-reported knowledge of digital tools. The LAD's minimal visual clutter enabled straightforward interpretation, reducing cognitive load for acting on the information displayed (Sedrakyan et al., 2020). Moreover, employing terminology and visual elements familiar to teachers (e.g., utilizing terms such as "anger" or "sadness" for affective states) proved beneficial, as it ensured a direct correlation between what was being measured and its manifestation in the classroom (e.g., high heart rate visualized as a high-stress indicator) (Kasepalu et al., 2022).

Overall, learning footprints were made visible to teachers through the LAD to counteract a poor awareness of students' workflows (Dillenbourg et al., 2011). Yet, it also prompted changes in teachers' behaviour, promoting the "wandering facilitator" strategy. As described by Tissenbaum and Slotta (2019), a wandering facilitator is a teacher who, instead of leading from the front, supports collaborative learning by moving among individual students or teams. Because these facilitators need to take physical action in the classroom, they may be unable to remain fixed in front of the LAD (Giannakos et al., 2020). Therefore, visualizing the classroom at different granularities, coupled with its simple red-green color coding for signaling, proved instrumental in identifying struggling groups, preventing unsubstantiated assumptions, and guiding interventions, all while leaving room for facilitators' subjective interpretation during intervention (Li et al., 2020; d'Anjou et al., 2019; Schwendimann et al., 2016).

LADs under scrutiny: addressing privacy, control, and over-reliance on data

Despite positive feedback on the LAD's usability, challenges emerged that informed our response to RQ2. Firstly, digital literacy appears to be an unavoidable requirement at this stage to, for instance, verify proper device functioning throughout the activity. Moreover, managing the infrastructure and handling troubleshooting may all require more than one facilitator on-site. Aside from proper technology orchestration, digital literacy is essential to correctly inform students about the procedure (Lee-Cultura et al., 2023). In fact, students may distrust the data collection process if they perceive teachers as lacking confidence, understanding, or clarity in explaining the LAD technology. Effective communication is a crucial enabler for students to become active participants in the learning process rather than merely passive data providers (Giannakos et al., 2020).

The facilitators' acceptance of LAD was high, as it augmented their presence in the educational environment (An et al., 2020). Still, a common skeptical sentiment emerged from the interview insights. Since they know their learners best, facilitators worry that full delegation of monitoring responsibilities to a machine-based monitoring system could be unsuitable. This position aligns with An et al. (2020) work on the delicate balance between teacher autonomy and orchestration automation, where over-reliance on computed outcomes risks reducing a teacher's role to that of a mere executor (An et al., 2020). In our case, it was reassuring that teachers wished to collaborate with technology without being put on the sidelines, since the proposed LAD asks for active mediators who elaborate on the information accessed. We observed positive commentaries, such as the automation of student state detection, which boosted teacher confidence (Bao et al., 2021) and validated their hypotheses in multiple instances. This approach maintained facilitators' freedom to practice according to their pedagogical beliefs without forcing them into practice changes. According to them, while occasional faults may occur in LADs, the "severity" of consequences largely depends on the facilitator's understanding of and reliance on these systems. On this count, a teacher also expressed low concern for data accuracy, explaining that a layer of human mediation should always be present (Sharma and Giannakos, 2021). In fact, inaccuracies would inevitably distort events. Therefore, the teacher's critical thinking should always be maintained in decision-making to mitigate machine errors (Lee-Cultura et al., 2023). Going back to existing research (Holstein et al., 2019; van Leeuwen And Rummel, 2022), this insight of ours corroborates instructors' willingness to remain in charge of the system and not be data-dependent. Further steps to build effective reliance on LADs shouldn't aim to transfer authority from the teacher to the machine for blind trust. Instead, they should seek to achieve stronger alignment with the context and greater clarity in the displayed content to support teachers' professional judgment. This concept is recurrent in the learning design and HCI fields and closely relates to the importance of situational factors when inquiring about SRL (Järvenoja et al., 2018). For example, phases involving affective states often perceived as negative (e.g., stress and confusion) are expected during SRL processes (Pekrun, 2011). However, they might lead to unnecessary assistance if feedback is poorly contextualized, especially in open-ended and collaborative activities that follow constructivist principles (Possaghi et al., 2024; Sung et al., 2023).

As observed in previous studies (Jones et al., 2020; Gao et al., 2020; Sung et al., 2023; Mangaroska et al., 2021), "data surveillance" is a concern for both facilitators and students' perspectives alike. In limited instances, students reported feeling ever monitored, leading to increased pressure and reduced sense of autonomy. While some students were uncomfortable with the presence of cameras mounted on the PC, this discomfort stemmed not from image acquisition itself, but from their use in the classroom setting (Sharma et al., 2022). This point is critical to elaborate on, as the organic nature of in-team dynamics may diminish if students feel compelled to uphold a facade of "seriousness", concerned that their informal interactions will be scrutinized or judged (Giannakos et al., 2020; de Arriba-Pérez et al., 2017). The correlation between students' questionnaire responses regarding their feelings and their perceptions of wearing a wristband or being recorded by the camera did not reveal any remarkable patterns of negative experiences. Interestingly, a link exists between perceiving the activity as "fun" and feeling bothered by teachers' access to their emotions via LAD. This suggests that "having fun" may have made students more conscious of the teacher's observational capabilities, thereby connecting stronger emotional expression with greater awareness of surveillance, accompanied by a slight sense of discomfort. Our system, however, did not capture student images or broadcast any footage to the facilitator as it focused solely on mapping emotional states. In retrospect, clarifying this distinction would have alleviated concerns and dispelled misconceptions about intrusive technology (Sharma and Giannakos, 2021;

Sharma et al., 2022). This might also explain why we observed differences in students' acceptance of cameras versus wristbands from the questionnaire's results. In fact, wristbands were highly tolerated, likely because teachers spent more time explaining their purpose and data collection, given their novelty as classroom tools. As such, transparent communication about the role and function of orchestration technologies should parallel efforts in technology integration, thereby preserving authentic learning dynamics (Giannakos et al., 2020).

Finally, the employment of LADs can be argued to contribute to the further centralization of the teacher's role (Morrison, 2014). This inherently creates friction with constructivist learning approaches that posit learners as the core of educational experiences. We aimed to balance the "sage on the stage" archetype by keeping the facilitator unobtrusive (intervening only when necessary) while also preventing a detached monitoring role (An et al., 2020). Therefore, it's crucial to conceptualize LAD use as a touchpoint: aligning the facilitator's perspective with students' experiences, enhancing proximity to their learning journeys, and dispelling ambiguity about their learning. Overcoming this tension is essential in building a connection founded on learning process transparency, rather than fostering detachment.

Considerations for LAD design

From our study, we applied an MMD approach that reinforces models of learning (Cukurova et al., 2020) based on children's responses. Ethical concerns, technical competence, and usability are the three main factors drawn from our results that overlap with existing literature. As pointed out by Shibani et al., the engagement of teachers in the loop can result in design materials that are potentially applicable to future projects (Shibani et al., 2020). To answer our RQ3, we now present several implications for overcoming scale-down hurdles when designing LAD systems for authentic environments:

1. Greater practitioner involvement in the design process can enhance the ability of the LAD system to be authorable (Martinez-Maldonado, 2023; Verbert et al., 2020).
2. Prioritizing customizable affordances in the LAD system to mitigate the sense of technology overshadowing the human practice, fostering a feeling of teacher agency (Shibani et al., 2020). For example, define the level of LAD's automation for the degree of human and machine-driven assessment and facilitation decisions (Sailer et al., 2024).
3. Leveraging real-world interventions as a context for validating and generating empirically-driven design insights relevant to the classroom ecology (Susnjak et al., 2022).
4. Cultivating usability enhances the effectiveness of LADs as a touchpoint between stakeholders and as a tool for transversal communication and interpretation (Verbert et al., 2020; De Vreugd et al., 2024).
5. Given the increasing involvement of LADs in Child-Computer Interaction, ethical considerations on data handling and privacy should be prioritized as part of the user-centered design process (Crescenzi-Lanna, 2020; D'mello and Graesser, 2010; Sharma and Giannakos, 2021).

Limitations and future work

While opportunities for in-class LAD use are promising, it is important to acknowledge and elaborate on our study's limitations to guide future research. We focused on the iterative design and validation of a teacher-facing LAD, specifically from the facilitators' perspective, within an authentic environment. However, we recognize the deployment as confined to a single school setting and a narrow group of teachers. For a more nuanced account of the LAD's effectiveness in classrooms, future interventions should broaden participant cohorts, expand the context of use, and extend the temporal scope to organize an extensive longitudinal investigation. This would facilitate deeper case analysis (e.g., interaction patterns over time). Despite the availability of sensor data, our analysis was based solely on qualitative impressions from teachers and quantitative questionnaires from students. This decision aligns with our RQs, prioritizing human-centered feedback for our LAD's first validation in-the-wild. Nevertheless, valuable insights remain derivable from the collected MM data. For instance, it would be of particular interest to achieve a comparative understanding between MM data with self-assessment measures (e.g., self-reported stress) provided by students regarding their learning experience (Giannakos et al., 2020). Shortcomings of Spearman's rank correlations are also worth addressing. First, the correlations found indicate associations without implying causal relationships. Second, our small sample size introduces statistical power caveats (e.g., limited generalizability), making it harder to confirm the stability of findings. To counteract this, future studies can address larger student samples and further supplement correlations with qualitative insights.

As reported as a logical implication of our user-centered workflow, we built a descriptive system complete with signaling. There is potential for further automation, such as implementing time-stamped emotional indicators to access a high-level view of students' progression (Sedrakyan et al., 2020; Gao et al., 2020; Papamitsiou et al., 2020). Moreover, future research direction could entail designing a prescriptive system, namely LADs, capable of giving inputs based on computed data or a predictive system able to foresee students' states based on the collected information. This would grant timely intervention before students reach, for example, a state of high stress. Closely related, the embedment of Artificial Intelligence (AI) is the next step to allow further automatization of data computation, potentially decreasing human bias by relying on automated threshold settings, depending, for example, on students' affective patterns. Even if AI can communicate with teachers proactively, foreseeing students' emotional states, it is not a choice exempt from challenges. Among the others, lack of transparency can fuel further mistrust if LA is deployed as black boxes (Susnjak et al., 2022; Spikol and Cukurova, 2020). Finally, our in-the-wild intervention employed a project-based, open-ended learning activity, contrasting with the predominantly linear approaches that can be easily observed in the literature for LAD validation (Giannakos et al., 2020). This contribution can be a foundational perspective for further developing monitoring techniques tailored to constructionist experiences within a real classroom, especially if augmented by learner input. Furthermore, drawing upon relevant work in the field (e.g., Silvola et al. in higher education (Silvola et al., 2021)), actively engaging students as equitable collaborators in the design process would ensure comprehensive consideration of all stakeholder perspectives. Moreover, including students "in the loop" would raise their ownership over learning journeys in light of, for example, data privacy (Sharma and Giannakos, 2021; Sharma et al., 2022).

Conclusion

The progressive pervasiveness of technology opens up exciting opportunities for the digitalization of supportive systems. Inspired by the feasibility of MM measurements and sensors and their underrepresentation in empirical K-12 research (Holstein et al., 2018; Susnjak et al., 2022), we developed a real-time LAD system designed to help teachers reduce their orchestrational cognitive load. We followed a user-centered approach, involving stakeholders since the early phases of our design journey. Our contribution validates the feasibility of a descriptive LAD in an authentic in-school environment, grounding this position in teachers' and educational researchers' insights to advance educational ecosystems toward greater "smartness" (Giannakos et al., 2020). However, despite the overall acceptance of the system, concerns regarding privacy, teacher digital literacy, and technology management highlight areas for improvement. We can further advance in the field, granting an even better experience to students if practitioners, particularly teachers, are positioned as key facilitators for the broader and more effective deployment of LADs.

Appendix A

See Tables 6, 7, 8, and 9.

Table 6 Guidelines derived from reviewing the literature

Papers	Guideline
Giannakos et al. (2020, 2019); Gao et al. (2020); Mangaroska et al. (2020); Giannakos et al. (2022)	MMLA offers a valuable opportunity to explore students' learning experiences, enhance their SRL, and assist teachers in this endeavor
Lee-Cultura et al. (2021); Papamitsiou et al. (2020); Giannakos et al. (2022); de Arriba-Pérez et al. (2017); Gao et al. (2022)	LA addressing physiological and affective metrics can assess learners' experience more holistically, inform tailored guidance, and increase awareness of students' response to the environment they are immersed in
Gao et al. (2020); Papamitsiou et al. (2020); Giannakos et al. (2022); de Arriba-Pérez et al. (2017); Giannakos et al. (2020) Sharma et al. (2022); Fortenbacher and Yun (2020); Sharma et al. (2022)	LA analyzing heart rate and skin response data is effective in mapping stress and emotional engagement as measurements
Emerson et al. (2023); Papamitsiou et al. (2020); Liu et al. (2018); Giannakos et al. (2022); de Arriba-Pérez et al. (2017)	Facial recognition technology using cameras is effective in capturing certain affective states, such as boredom, by decoding facial expressions
Sung et al. (2023)	Negative affective states, as well as positive ones, should not be overlooked in data gathering to capture a more comprehensive spectrum of student experiences
Sedrakyan et al. (2020); Gao et al. (2020); Papamitsiou et al. (2020); d'Anjou et al. (2019); Giannakos et al. (2022, 2020); Mangaroska et al. (2021)	Data gathering, analysis, and feedback from orchestration technologies (e.g., LADs) should be time-reliant to ensure clear pinpointing of events and timely facilitators' intervention
Gao et al. (2020); Sung et al. (2023); de Arriba-Pérez et al. (2017)	Students' monitoring and data collection via sensors should not be perceived as limiting or intrusive to their experience
Sedrakyan et al. (2020); Giannakos et al. (2020); Mangaroska et al. (2021); Martinez-Maldonado et al. (2020)	End-user acceptance towards orchestration technologies (e.g., LADs) should be iteratively cultivated and informed by stakeholders' insights
Sedrakyan et al. (2020)	Avoid providing excessive feedback (e.g., too frequent or overly detailed), as it may lead to content overload. Mitigate this risk by introducing user control
Li et al. (2020); Giannakos et al. (2020)	The possibility of changing the granularity of information displayed by the LAD (e.g., individual level, class level) improves educators awareness
d'Anjou et al. (2019); Giannakos et al. (2020)	Orchestrational technologies (e.g., LADs) should leverage facilitators' awareness and decision-making without burdening cognitively
d'Anjou et al. (2019); Holstein et al. (2019); Martinez-Maldonado et al. (2020)	Orchestrational technologies (e.g., LADs) should detain a high degree of interpretability to encourage facilitators' own intuitions
Giannakos et al. (2022); d'Anjou et al. (2019)	Deploying orchestrational technologies (e.g., LADs) in real-world settings is essential to ensure their convenience, accuracy, and seamless usability within the context
Giannakos et al. (2022); Fortenbacher and Yun (2020); Mangaroska et al. (2021); Martinez-Maldonado et al. (2020)	Orchestration technologies (e.g., LADs) should be designed to maintain data privacy at the forefront to achieve accountability in educational venues
Yan et al. (2023); Giannakos et al. (2020); d'Anjou et al. (2019)	The design method for Orchestrational technologies (e.g., LADs) should be context dependent from the early stages

Table 7 Shapiro–Wilk normality test results for the set of items *(a)*

Shapiro–Wilk test	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
W	0.88	0.90	0.86	0.88	0.85	0.90
p-value	0.0003*	0.0012*	7.1e−5*	0.0002*	4.3e−5*	0.0009*
$\mathcal{N}$?	$\neg\mathcal{N}$	$\neg\mathcal{N}$	$\neg\mathcal{N}$	$\neg\mathcal{N}$	$\neg\mathcal{N}$	$\neg\mathcal{N}$
N of answers	44	44	44	44	44	44

W statistics and p-values are reported, indicating significant departures from normality ($p < 0.05$, marked with *)

Table 8 Shapiro–Wilk normality test results for the questionnaire's set of items *(b)*

Shapiro–Wilk test	Item 7	Item 8	Item 9	Item 10
W	0.85	0.92	0.88	0.78
p-value	0.0005*	0.017*	0.0022*	2.1e−5*
$\mathcal{N}$?	$\neg\mathcal{N}$	$\neg\mathcal{N}$	$\neg\mathcal{N}$	$\neg\mathcal{N}$
N of answers	32	32	32	32

W statistics and p-values are reported, showing significant departures from normality ($p < 0.05$, marked with *)

Table 9 Shapiro–Wilk normality test results for the questionnaire's set of items *(c)*

Shapiro–Wilk test	Item 11	Item 12	Item 13
W	0.75	0.88	0.90
p-value	2.3e−7*	0.0003*	0.0010*
$\mathcal{N}$?	$\neg\mathcal{N}$	$\neg\mathcal{N}$	$\neg\mathcal{N}$
N of answers	44	44	44

W statistics and p-values are reported, showing significant departures from normality ($p < 0.05$, marked with *)

Abbreviations

SRL Self-regulated learning
LA Learning analytics
MM Multi-modal
LAD Learning analytics dashboard

Acknowledgements

Our many thanks go out to every student and teacher who dedicated their time and effort to this research. Making this study a reality would have been impossible without their openness to participate and share their experiences. The authors acknowledge the use of OpenAI's ChatGPT and Google Gemini, which were employed under human supervision to support language editing, phrasing refinements, and clarity improvements in the preparation of this manuscript. Some vector icons and illustrations used in figures were provided by DrawKit (© Designstripe). Used under license; no endorsement implied.

Author contributions

Isabella Possaghi wrote most of the paper, participated in the LAD's development, and conducted the experiment. Boban Vesin followed the whole LAD's design and implementation process, contributed to the Dashboard Design and Implementation section, edited the overall paper, and provided valuable insights for improvements. Feiran Zhang followed the whole LAD's design and implementation process, contributed to the Background and Dashboard Design and Implementation section, edited the overall paper, and provided valuable insights for improvements. Kshitij Sharma followed the whole LAD's design and implementation process, edited the overall paper, and provided valuable insights for improvements. Cecilie Knudsen and Håkon Bjørkum developed the LAD, designed the in-the-wild validation, and participated in conducting the experiment. Sofia Papavlasopoulou followed the whole LAD's design and implementation process and provided overall supervision for the paper's writing.

Funding

Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital). Funded by the European Union. Views and opinions expressed are, however, those of the author only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This research has received funding from the European Union's Horizon

Europe Framework Programme for Research and Innovation under Grant Agreement No. 101060231, Exten.D.T.2-Extending Design Thinking with Emerging Digital Technologies.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. All data have been fully anonymized in compliance with Sikt (Norwegian Agency for Shared Services in Education and Research) regulations governing research with children, and can only be shared in this anonymized format.

Declarations

Competing interests

The authors declare that they have no Conflict of interest.

Received: 26 November 2024 Accepted: 1 August 2025
Published online: 21 August 2025

References

Ah-Nam, L., & Osman, K. (2017). Developing 21st century skills through a constructivist-constructionist learning environment. *K-12 Stem Education, 3*(2), 205–216.

Ahn, J., Campos, F., Hays, M., & DiGiacomo, D. (2019). Designing in context: Reaching beyond usability in learning analytics dashboard design. *Journal of Learning Analytics, 6*(2), 70–85.

Alfredo, R., Echeverria, V., Zhao, L., Lawrence, L., Fan, J. X., Yan, L., Li, X., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Designing a human-centred learning analytics dashboard in-use. *Journal of Learning Analytics, 11*(3), 62–81.

Amarasinghe, I., Michos, K., Crespi, F. & Hernández-Leo, D. (2022). Learning analytics support to teachers' design and orchestrating tasks. *Journal of Computer Assisted Learning*.

Amarasinghe, I., Hernández-Leo, D., & Ulrich Hoppe, H. (2021). Deconstructing orchestration load: Comparing teacher support through mirroring and guiding. *International Journal of Computer-Supported Collaborative Learning, 16*(3), 307–338.

An, P., Holstein, K., d'Anjou, B., Eggen, B. & Bakker, S. (2020). The ta framework: Designing real-time teaching augmentation for k-12 classrooms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–17).

Arriba-Pérez, F., Caeiro-Rodríguez, M. & Santos-Gago, J.M. (2017). Towards the use of commercial wrist wearables in education. In *2017 4th Experiment@ International Conference* (exp. At'17) (pp. 323–328). IEEE.

Bao, H., Li, Y., Su, Y., Xing, S., Chen, N.-S., & Rosé, C. P. (2021). The effects of a learning analytics dashboard on teachers' diagnosis and intervention in computer-supported collaborative learning. *Technology, Pedagogy and Education, 30*(2), 287–303.

Bond, M., Viberg, O. & Bergdahl, N. (2023). The current state of using learning analytics to measure and support k-12 student engagement: A scoping review. In *LAK23: 13th International Learning Analytics and Knowledge Conference* (pp. 240–249).

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L. & Zhao, W. (2015). Automatic detection of learning-centered affective states in the wild. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 379–388).

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. *Journal of Learning Analytics, 6*(2), 1–9.

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.

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

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology, 51*(5), 1441–1449.

d'Anjou, B., Bakker, S., An, P. & Bekker, T. (2019). How peripheral data visualisation systems support secondary school teachers during vle-supported lessons. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (pp. 859–870). New York: Association for Computing Machinery.

De Vreugd, L., Van Leeuwen, A., Jansen, R., & Schaaf, M. (2024). Learning analytics dashboard design and evaluation to support student self-regulation of study behaviour. *Journal of Learning Analytics, 11*(3), 249–262.

Dillenbourg, P., Zufferey, G., Alavi, H., Jermann, P., Do-Lenh, S., Bonnard, Q., Cuendet, S. & Kaplan, F. (2011). *Classroom orchestration: The third circle of usability*.

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education, 69*, 485–492.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). *The Evolution of Research on Computer-supported Collaborative Learning: From Design to Orchestration*. Springer.

D'Mello, S., Jensen, E. (2017). Emotional learning analytics. In Lang, C., Siemens, G., Wise, A.F., Gašević, D., Merceron, A. (Eds.) *The Handbook of Learning Analytics* (2nd ed., pp. 120–129). SoLAR. Section: 12. https://www.solaresearch.org/publications/hla-22/hla22-chapter12/

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*, 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., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153–170.
Echeverria, V., Yan, L., Zhao, L., Abel, S., Alfredo, R., Dix, S., Jaggard, H., Wotherspoon, R., Osborne, A., Buckingham Shum, S., et al. (2024). Teamslides: a multimodal teamwork analytics dashboard for teacher-guided reflection in a physical learning space. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp 112–122).
Emerson, A., Min, W., Rowe, J., Azevedo, R. & Lester, J. (2023). Multimodal predictive student modeling with multi-task transfer learning. In *LAK23: 13th International Learning Analytics and Knowledge Conference* (pp. 333–344).
Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education, 59*(2), 423–435.
Fortenbacher, A. & Yun, H. (2020). Can sensors effectively support learning? *Artificial Intelligence Supported Educational Technologies*, 93–114.
Gao, N., Shao, W., Rahaman, M. S., & Salim, F. D. (2020). n-gage: Predicting in-class emotional, behavioural and cognitive engagement in the wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4*(3), 1–26.
Gao, W., Wei, T., Huang, H., Chen, X., & Li, Q. (2022). Toward a systematic survey on wearable computing for education applications. *IEEE Internet of Things Journal, 9*(15), 12901–12915.
Giannakos, M., Cukurova, M. & Papavlasopoulou, S. (2022). Sensor-based analytics in education: Lessons learned from research in multimodal learning analytics. In *The Multimodal Learning Analytics Handbook* (pp. 329–358). Springer.
Giannakos, M. N., Lee-Cultura, S., & Sharma, K. (2021). Sensing-based analytics in education: The rise of multimodal data enabled learning systems. *IT Professional, 23*(6), 31–38.
Giannakos, M. N., Sharma, K., Papavlasopoulou, S., Pappas, I. O., & Kostakos, V. (2020). Fitbit for learning: Towards capturing the learning experience using wearable sensing. *International Journal of Human-Computer Studies, 136*, Article 102384.
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.
Graesser, A. C. (2020). Emotions are the experiential glue of learning environments in the 21st century. *Learning and Instruction, 70*, Article 101212.
Hesse-Biber, S.N., Rodriguez, D. & Frost, N.A. (2015). A qualitatively driven approach to multimethod and mixed methods research.
Holstein, K., Hong, G., Tegene, M., McLaren, B.M. & Aleven, V. (2018). The classroom as a dashboard: Co-designing wearable cognitive augmentation for k-12 teachers. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 79–88).
Holstein, K., McLaren, B.M. & Aleven, V. (2017). Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 257–266).
Holstein, K., McLaren, B.M. & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher-ai complementarity. *Grantee Submission*.
Hoover, W. A. (1996). The practice implications of constructivism. *SEDL Letter, 9*(3), 1–2.
Järvelä, S., Järvenoja, H., Malmberg, J., Isohätälä, J., & Sobocinski, M. (2016). How do types of interaction and phases of self-regulated learning set a stage for collaborative engagement? *Learning and Instruction, 43*, 39–51.
Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction, 72*, Article 101203.
Järvenoja, H., Järvelä, S., Törmänen, T., Näykki, P., Malmberg, J., Kurki, K., Mykkänen, A., & Isohätälä, J. (2018). Capturing motivation and emotion regulation during a learning process. *Frontline Learning Research, 6*(3), 85–104.
Jivet, I., Scheffel, M., Specht, M. & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge, pp. 31–40.
Johnson, A.M., Jacovina, M.E., Russell, D.G. & Soto, C.M. (2016). Challenges and solutions when using technologies in the classroom. In *Adaptive Educational Technologies for Literacy Instruction* (pp. 13–30). Routledge.
Jones, K. M., Asher, A., Goben, A., Perry, M. R., Salo, D., Briney, K. A., & Robertshaw, M. B. (2020). "we're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology, 71*(9), 1044–1059.
Kasepalu, R., Chejara, P., Prieto, L. P., & Ley, T. (2022). Do teachers find dashboards trustworthy, actionable and useful? a vignette study using a logs and audio dashboard. *Technology, Knowledge and Learning, 27*(3), 971–989.
Kaur, A., & Chahal, K. K. (2024). A learning analytics dashboard for data-driven recommendations on influences of non-cognitive factors in introductory programming. *Education and Information Technologies, 29*(8), 9221–9256.
Kessel, M.v., Molenaar, I., Knoop-van Campen, C.A., Jonge, M.d. & Saab, N. (2025). Primary school teacher perspectives on effective dashboard use in the classroom: Skills, knowledge, and contextual conditions.
Knoop-van Campen, C. A., Wise, A., & Molenaar, I. (2023). The equalizing effect of teacher dashboards on feedback in k-12 classrooms. *Interactive Learning Environments, 31*(6), 3447–3463.
Lee-Cultura, S., Sharma, K., Cosentino, G., Papavlasopoulou, S. & Giannakos, M. (2021). Children's play and problem solving in motion-based educational games: Synergies between human annotations and multi-modal data. In *Interaction Design and Children* (pp. 408–420).
Lee-Cultura, S., Sharma, K., & Giannakos, M. N. (2023). Multimodal teacher dashboards: Challenges and opportunities of enhancing teacher insights through a case study. *IEEE Transactions on Learning Technologies, 17*, 181–201.
Leeuwen, A., Rummel, N., et al. (2019). Orchestration tools to support the teacher during student collaboration: A review. *Unterrichtswissenschaft, 47*(2), 143–158.
Leeuwen, A., & Rummel, N. (2022). The function of teacher dashboards depends on the amount of time pressure in the classroom situation: Results from teacher interviews and an experimental study. *Unterrichtswissenschaft, 50*(4), 561–588.

Li, Q., Ren, Y., Wei, T., Wang, C., Liu, Z., & Yue, J. (2020). A learning attention monitoring system via photoplethysmogram using wearable wrist devices. *Artificial Intelligence Supported Educational Technologies, 133*, 150.

Liu, S., Chen, Y., Huang, H., Xiao, L. & Hei, X. (2018). Towards smart educational recommendations with reinforcement learning in classroom. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering* (TALE) (pp. 1079–1084). IEEE.

Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X., & Siipo, A. (2019). Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior, 96*, 235–245.

Mangaroska, K., & Giannakos, M. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies, 12*(4), 516–534.

Mangaroska, K., Martinez-Maldonado, R., Vesin, B., & Gašević, D. (2021). Challenges and opportunities of multimodal data in human learning: The computer science students' perspective. *Journal of Computer Assisted Learning, 37*(4), 1030–1047.

Mangaroska, K., Sharma, K., Gaševic, D., & Giannakos, M. (2020). Multimodal learning analytics to inform learning design: Lessons learned from computing education. *Journal of Learning Analytics, 7*(3), 79–97.

Martinez-Maldonado, R., Echeverria, V., Fernandez Nieto, G., Buckingham Shum, S. (2020). From data to insights: A layered storytelling approach for multimodal learning analytics. In *Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems* (pp. 1–15).

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J. & Clayphan, A. (2015). The latux workflow: designing and deploying awareness tools in technology-enabled learning settings. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 1–10).

Martinez-Maldonado, R. (2019). A handheld classroom dashboard: Teachers' perspectives on the use of real-time collaborative learning analytics. *International Journal of Computer-Supported Collaborative Learning, 14*(3), 383–411.

Martinez-Maldonado, R. (2023). Human-centred learning analytics: Four challenges in realising the potential. *Learning Letters, 1*, 6.

Martinez-Maldonado, R., Clayphan, A., Yacef, K., & Kay, J. (2014). Mtfeedback: Providing notifications to enhance teacher awareness of small group work in the classroom. *IEEE Transactions on Learning Technologies, 8*(2), 187–200.

Mohseni, Z., Masiello, I., & Martins, R. M. (2023). Co-developing an easy-to-use learning analytics dashboard for teachers in primary/secondary education: A human-centered design approach. *Education Sciences, 13*(12), 1190.

Morrison, C.D. (2014). From 'sage on the stage'to' guide on the side': A good start.

Nasir, J., Kothiyal, A., Bruno, B., & Dillenbourg, P. (2021). Many are the ways to learn identifying multi-modal behavioral profiles of collaborative learning in constructivist activities. *International Journal of Computer-Supported Collaborative Learning, 16*(4), 485–523.

Nguyen, N.B.C., Lithander, M., Östlund, C.M., Karunaratne, T. & Jobe, W. (2024). Teadash: Implementing and evaluating a teacher-facing dashboard using design science research. In: Informatics, vol. 11, p. 61. MDPI.

Ochoa, X., Lang, C., Siemens, G., Wise, A., Gasevic, D., & Merceron, A. (2022). Multimodal learning analytics-rationale, process, examples, and direction. *The Handbook of Learning Analytics, 2*, 54–65.

O'Connor, K. (2022). Constructivism, curriculum and the knowledge question: Tensions and challenges for higher education. *Studies in Higher Education, 47*(2), 412–422.

Ocumpaugh, J. (2015). Baker rodrigo ocumpaugh monitoring protocol (bromp) 2.0 technical and training manual. New York, NY and Manila, Philippines: Teachers College, Columbia University and Ateneo Laboratory for the Learning Sciences *60*.

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

Papamitsiou, Z., Pappas, I. O., Sharma, K., & Giannakos, M. N. (2020). Utilizing multimodal data through fsqca to explain engagement in adaptive learning. *IEEE Transactions on Learning Technologies, 13*(4), 689–703.

Paulsen, L. & Lindsay, E. (2024). Learning analytics dashboards are increasingly becoming about learning and not just analytics-a systematic review. *Education and Information Technologies*, 1–30.

Peel, K. L. (2020). A beginner's guide to applied educational research using thematic analysis. *Practical Assessment Research and Evaluation, 25*, 1.

Pekrun, R. (2011). Emotions as drivers of learning and cognitive development. In *New Perspectives on Affect and Learning Technologies* (pp. 23–39). Springer.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.

Possaghi, I., Zhang, F., Sharma, K. & Papavlasopoulou, S. (2024). Design thinking activities for k-12 students: Multi-modal data explanations on coding performance. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference* (pp. 290–306).

Pozdniakov, S., Martinez-Maldonado, R., Tsai, Y.-S., Echeverria, V., Srivastava, N. & Gasevic, D. (2023). How do teachers use dashboards enhanced with data storytelling elements according to their data visualisation literacy skills? In *LAK23: 13th International Learning Analytics and Knowledge Conference* (pp. 89–99).

Reiser, B.J. (2013). What professional development strategies are needed for successful implementation of the next generation science standards. In *Invitational Research Symposium on Science Assessment* (pp. 1–22). ETS Washington, DC, USA.

Roberts, L. D., Howell, J. A., & Seaman, K. (2017). Give me a customizable dashboard: Personalized learning analytics dashboards in higher education. *Technology, Knowledge and Learning, 22*, 317–333.

Sailer, M., Ninaus, M., Huber, S.E., Bauer, E. & Greiff, S. (2024). The end is the beginning is the end: The closed-loop learning analytics framework. *Computers in Human Behavior*, 108305.

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies, 10*(1), 30–41.

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior, 107*, Article 105512.
Shahmoradi, S., Kothiyal, A., Bruno, B., & Dillenbourg, P. (2024). Evaluation of teachers' orchestration tools usage in robotic classrooms. *Education and Information Technologies, 29*(3), 3219–3256.
Sharma, K., Papavlasopoulou, S. & Giannakos, M. (2021). Faces don't lie: Analysis of children's facial expressions during collaborative coding. In *FabLearn Europe/MakeEd 2021-An International Conference on Computing, Design and Making in Education* (pp. 1–10).
Sharma, K., Pappas, I., Papavlasopoulou, S. & Giannakos, M. (2022). Wearable sensing and quantified-self to explain learning experience. In *2022 International Conference on Advanced Learning Technologies* (ICALT) (pp. 136–138). IEEE.
Sharma, K., & Giannakos, M. (2021). Sensing technologies and child-computer interaction: Opportunities, challenges and ethical considerations. *International Journal of Child-Computer Interaction, 30*, Article 100331.
Sharma, K., Lee-Cultura, S., & Giannakos, M. (2022). Keep calm and do not carry-forward: Toward sensor-data driven ai agent to enhance human learning. *Frontiers in Artificial Intelligence, 4*, Article 713176.
Sharma, K., Papavlasopoulou, S., & Giannakos, M. (2022). Children's facial expressions during collaborative coding: Objective versus subjective performances. *International Journal of Child-Computer Interaction, 34*, Article 100536.
Shibani, A., Knight, S., & Shum, S. B. (2020). Educator perspectives on learning analytics in classroom practice. *The Internet and Higher Education, 46*, Article 100730.
Silvola, A., Näykki, P., Kaveri, A., & Muukkonen, H. (2021). Expectations for supporting student engagement with learning analytics: An academic path perspective. *Computers & Education, 168*, Article 104192.
Spikol, D. & Cukurova, M. (2020). Multimodal learning analytics. In *Encyclopedia of Education and Information Technologies* (pp. 1221–1228). Springer.
Spikol, D., Ruffaldi, E., & Cukurova, M. (2017). *Using multimodal learning analytics to identify aspects of collaboration in project-based learning*. International Society of the Learning Sciences.
Sung, G., Bhinder, H., Feng, T., & Schneider, B. (2023). Stressed or engaged? addressing the mixed significance of physiological activity during constructivist learning. *Computers & Education, 199*, Article 104784.
Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: A tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education, 19*(1), 12.
Tissenbaum, M., Matuk, C., Berland, M., Lyons, L., Cocco, F., Linn, M., Plass, J. L., Hajny, N., Olsen, A., Schwendimann, B., et al. (2016). *Real-time visualization of student activities to support classroom orchestration*. International Society of the Learning Sciences.
Tissenbaum, M., & Slotta, J. (2019). Supporting classroom orchestration with real-time feedback: A role for teacher dashboards and real-time agents. *International Journal of Computer-Supported Collaborative Learning, 14*, 325–351.
Tisza, G., Sharma, K., Papavlasopoulou, S., Markopoulos, P. & Giannakos, M. (2022). Understanding fun in learning to code: A multi-modal data approach. In *Interaction Design and Children* (pp. 274–287).
Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of multiple computer-supported collaborating groups. *Computers in Human Behavior, 52*, 233–242.
Van Mechelen, M., Smith, R. C., Schaper, M.-M., Tamashiro, M., Bilstrup, K.-E., Lunding, M., Graves Petersen, M., & Sejer Iversen, O. (2023). Emerging technologies in k-12 education: A future hci research agenda. *ACM Transactions on Computer-Human Interaction, 30*(3), 1–40.
Verbert, K., Ochoa, X., De Croon, R., Dourado, R.A. & De Laet, T. (2020). Learning analytics dashboards: The past, the present and the future. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 35–40).
Wiley, K., Dimitriadis, Y., & Linn, M. (2024). A human-centred learning analytics approach for developing contextually scalable k-12 teacher dashboards. *British Journal of Educational Technology, 55*(3), 845–885.
Wolters, C.A., Pintrich, P.R. & Karabenick, S.A. (2005). Assessing academic self-regulated learning. What do children need to flourish? Conceptualizing and measuring indicators of positive development, 251–270.
Xie, B., Harpstead, E., DiSalvo, B., Slovak, P., Kharrufa, A., Lee, M.J., Pammer-Schindler, V., Ogan, A. & Williams, J.J. (2019). Learning, education, and hci. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI EA '19, (pp. 1–4). https://doi.org/10.1145/3290607.3311761
Yan, L., Martinez-Maldonado, R., Zhao, L., Li, X. & Gašević, D. (2023). Physiological synchrony and arousal as indicators of stress and learning performance in embodied collaborative learning. In *International Conference on Artificial Intelligence in Education* (pp. 602–614). Springer.
Yin, R.K. (2009). *Case Study Research: Design and Methods* (vol. bseriesno5). Sage.
Yoo, M., & Jin, S.-H. (2020). Development and evaluation of learning analytics dashboards to support online discussion activities. *Educational Technology & Society, 23*(2), 1–18.
Zimmerman, B.J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of Self-regulation* (pp 13–39). Elsevier.
Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist, 48*(3), 135–147.

Publisher's Note