

Article

Transforming Virtual Reality Simulation Logs into Pedagogically Interpretable Evidence

Ibis M. Alvarez ^{1,*}, Sergi Robles ², Alejandro Romero-Hernández ³, Andy Morodo ¹ and Borja Manero ³

¹ Department of Basic, Evolutionary and Education Psychology, Universitat Autònoma de Barcelona, Bellaterra, 08193 Cerdanyola del Vallès, Spain; andy.morodo@uab.cat

² Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, Bellaterra, 08193 Cerdanyola del Vallès, Spain; sergi.robles@uab.cat

³ Software Engineering and Artificial Intelligence Department, Universidad Complutense de Madrid, 28040 Madrid, Spain; alerom02@ucm.es (A.R.-H.); bmanero@ucm.es (B.M.)

* Correspondence: ibismarlene.alvarez@uab.es

Abstract

The digital transformation of higher education requires assessment approaches capable of capturing complex learning processes. In teacher education, immersive virtual reality (VR) simulations generate multimodal records, yet transforming them into pedagogically meaningful evidence for assessing classroom-management competence remains challenging. This methodological pilot study examines Mission Control–Didascalías VC, an analytics infrastructure for classroom-management simulations. Of 626 sessions, 595 were retained in the cleaned analytical corpus. Data sources comprised session metadata, verbal interactions, programmed challenging events, vocal and movement signals, dashboard exports and, for 17 focal cases, video recordings and transcripts. The analysis examined data quality, traceability and interpretability, developed and refined a multimodal coding framework, and included a worked example of sequence-based coding. Results show that simulation records do not constitute assessment evidence in themselves. Their formative value emerges through cleaning, temporal alignment, multimodal triangulation and contextual interpretation, which transform isolated technical traces into pedagogically meaningful interactional sequences. Critical verbal interactions, participant responses, emergent classroom situations, and vocal and movement patterns provided complementary evidence, including in sessions without programmed challenging events. The study concludes that VR analytics can support sustainable formative assessment when technical records are transformed into pedagogically interpretable interactional evidence for debriefing, feedback and evidence reuse over time.

Keywords: virtual reality; teacher education; classroom management; multimodal learning analytics; formative assessment; assessment infrastructure



Academic Editor: Eleanor Dommett

Received: 20 July 2026

Revised: 31 August 2026

Accepted: 2 September 2026

Published: 4 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

Digital transformation occupies a central place in the evolution of higher education. Its educational value depends on technologies being integrated through pedagogically meaningful approaches, rather than simply incorporated into teaching (Haleem et al., 2022). From a sustainability perspective, such integration should support evidence-informed practices that can be sustained, scaled and improved over time (Markauskaite et al., 2023).

Assessment is a key dimension of these practices because it enables pedagogical meaning to be attributed to the information generated during learning. In digital environments, data acquire formative value when teachers can retrieve, interpret and transform them into useful feedback that supports reflection and improvement. Process-oriented approaches to assessment therefore emphasise the importance of making learning trajectories visible, together with evidence of how learners act, revise decisions and respond to difficulties, beyond the final products of learning (Winstone et al., 2026).

This need is particularly relevant in teacher education, where the development of professional competence requires opportunities for situated practice, reflection and formative feedback. Managing classroom interaction, conflict and participation involves making pedagogical decisions in complex situations. Digital technologies are therefore especially valuable when embedded in designs that enable such practices to be rehearsed and generate evidence of how they unfold (Garzón Artacho et al., 2020; Haleem et al., 2022).

Among these technologies, immersive virtual reality (VR) offers particular value for teacher education because it can recreate realistic, interactive and repeatable professional situations. Simulations enable pre-service teachers to practise managing classroom interaction, conflict and group dynamics before encountering comparable situations in authentic educational settings (Han et al., 2025; Li et al., 2024). At the same time, they generate detailed records of professional performance, creating the possibility of examining how participants act, respond and make decisions during the simulation.

Realising this assessment potential depends on infrastructures that make such records observable, traceable and open to pedagogical interpretation. From the perspective of technological mediation, digital systems do more than capture activity: they shape how professional practice is represented and made available as evidence (Voordijk & Vahdatikhaki, 2022). Assessment infrastructures therefore provide the conditions through which simulation records can be organised, contextualised and transformed into information that supports teacher interpretation and feedback.

This process lies within the field of learning analytics, which examines how data generated during learning activities can be collected, analysed and interpreted to support learning and educational decision-making. From an implementation perspective, this process extends from the learning environment and raw data capture through data preparation and analysis to educational application (Qazi & Pachler, 2025; Ruipérez-Valiente, 2020). Digital records acquire assessment value only when they are organised and connected to relevant constructs, learning objectives and pedagogical criteria (Gašević et al., 2015, 2017). Once these relationships have been established, learning analytics can make patterns of activity visible and provide information that informs teachers' decisions and feedback practices (Banihashem et al., 2022).

In immersive simulations, this analytical process must account for the multimodal nature of professional performance. Classroom management is expressed through verbal exchanges, behaviour, bodily movement, timing, audiovisual interaction and responses to unfolding events. Multimodal learning analytics provides a suitable approach for integrating these sources and interpreting processes that cannot readily be represented through a single indicator (Blikstein & Worsley, 2016). This integration nevertheless requires careful contextual interpretation: vocal signals, movements, events and audiovisual records do not correspond directly to competence, learning or decision-making. Their meaning depends on sequence, verbal context, participant responses and the pedagogical situation, as well as on ethical and privacy considerations (Sharma & Giannakos, 2020). Such an interpretative foundation is also necessary for future automated or semi-automated procedures, since the patterns they identify become educationally meaningful only when they can be explained in pedagogical terms (Gašević et al., 2017; Zhang et al., 2025).

Turning this interpretative potential into sustainable assessment practice requires infrastructures capable of preserving the relationships among data sources, context and pedagogical meaning. These infrastructures must support the capture, organisation and retrieval of multimodal records, while enabling them to be revisited and transformed into evidence for feedback, teacher decision-making and continuous improvement. Their role therefore extends beyond data storage or visualisation: they provide the conditions under which evidence generated during immersive experiences can remain interpretable, reusable and educationally valuable over time.

Recent reviews of VR in teacher education have mainly examined educational effects, implementation and user experience, while research on learning analytics in immersive environments remains underdeveloped in terms of theoretical grounding, multimodal integration and support for educators' decision-making (Han et al., 2025; Tao et al., 2025). Taken together, these findings point to a relative lack of attention to the assessment infrastructures needed to transform immersive multimodal records into pedagogically actionable evidence. A related shortcoming can be observed in teacher-facing learning analytics dashboards, which often enhance awareness by visualising activity but provide less support for interpreting data and translating them into well-founded pedagogical action (Kaliisa et al., 2023). A gap therefore remains between the capacity of immersive simulations to generate rich multimodal records and the availability of validated infrastructures capable of transforming them into interpretable, actionable and reusable assessment evidence.

To clarify the mechanism underlying this transformation, the study proposes the Model of Transformation of Multimodal Traces into Pedagogical Evidence in Simulations (MTE-Sim) (Figure 1). The model articulates learning analytics, multimodal learning analytics and formative assessment as complementary methodological principles. It assumes that immersive VR simulations generate situated traces of professional activity, but that these traces do not constitute assessment evidence by themselves. Their educational value emerges from a transformational trajectory. In this process, technical records are prepared and integrated analytically, reconstructed through multimodal triangulation, interpreted using pedagogical constructs, and ultimately consolidated for formative use. Contextual-pedagogical mediation is central to this process because it links digital records to learning objectives, classroom management constructs, ethical safeguards and the situated meaning of participants' responses. Assessment infrastructure operates transversally by preserving traceability, contextual relationships and long-term reusability across the phases. Thus, the model frames assessment evidence as the result of an interpretive and infrastructural process that connects technological records, pedagogical meaning and formative feedback.

Didascalias VC provides a pertinent context in which to examine this gap. It is an immersive VR platform designed for teacher education in classroom management, whose techno-pedagogical design, implementation and acceptance among pre-service teachers have been explored in previous studies, with favourable findings regarding user acceptance (Alvarez et al., 2024). The platform operates in conjunction with Mission Control, a dashboard and data infrastructure that manages simulation sessions and records different dimensions of participants' activity. Mission Control therefore provides the means through which performance in the simulated classroom can be retrieved, traced and made available for pedagogical interpretation.

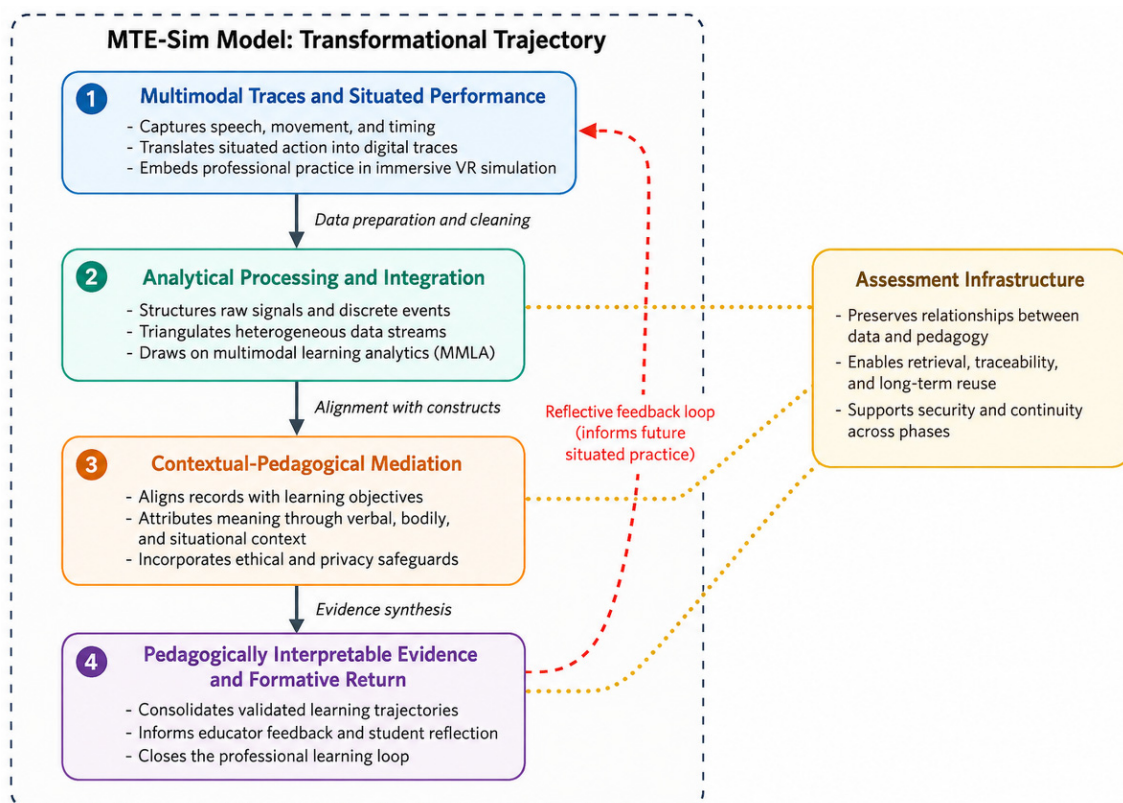


Figure 1. Model of Transformation of Multimodal Traces into Pedagogical Evidence in Simulations (MTE-Sim). The model represents the conceptual trajectory through which multimodal traces generated in immersive VR simulations are transformed into pedagogically interpretable evidence for formative assessment.

Before the formative impact of Didascalias VC can be examined at scale, the assessment potential of this infrastructure must first be established. Specifically, it is necessary to determine whether the records generated during the simulations and made available through Mission Control meet the requirements of quality, traceability and interpretability needed to support formative assessment. This validation is particularly timely at the platform's current stage of development, as it can provide a foundation for evidence-based feedback, broader implementation and future analytical or AI-supported developments.

The present study is a methodological pilot examining the assessment infrastructure of Mission Control–Didascalias VC. It investigates whether the technical and multimodal records generated during the simulations are sufficiently complete, traceable and interpretable to support formative assessment. It also proposes a preliminary multimodal coding framework for converting these records into pedagogically meaningful evidence of classroom-management performance. Rather than evaluating the formative impact of the VR platform, the study addresses a prior condition: establishing how simulation data can support teacher interpretation and evidence-based feedback.

The study is guided by the following research questions (RQs):

- RQ1. What types of data are generated by Mission Control–Didascalias VC, and what are their main affordances and limitations for formative assessment?
- RQ2. How can discrete events, continuous voice and non-verbal signals, and audiovisual records of the simulations be transformed into multimodal assessment evidence?
- RQ3. How can a sequence-based multimodal coding framework support the formative assessment of classroom-management competence in initial teacher education?

2. Materials and Methods

2.1. Study Design

This study adopted a methodological pilot design to examine whether records generated during immersive VR simulations could be transformed into traceable and pedagogically interpretable evidence for formative assessment. Such designs are appropriate for examining the feasibility of an emerging procedure, identifying the conditions required for its implementation and informing subsequent refinement (Bowen et al., 2009). The study did not evaluate the impact of Didascalias VC on classroom-management competence; instead, it addressed the prior methodological question of whether the records generated during the simulations and made available through Mission Control could support a viable assessment procedure.

Mission Control, developed by the research team at the Universidad Complutense de Madrid and the Universitat Autònoma de Barcelona (Madrid and Barcelona, Spain), was examined as an emerging digital assessment infrastructure that records and organises session metadata, facilitator-mediated verbal interactions, programmed challenging events and continuous voice and movement signals. Complementary audiovisual records and transcripts were incorporated into the focal analysis. In this study, the methodological examination focused on the systematic analysis of data availability, quality, traceability and interpretability, together with the development and refinement of a preliminary multimodal coding framework. This examination did not include an independent reliability assessment of the coding framework.

The design combined corpus-level analysis with focal multimodal case analysis. The methodological process was iterative and comprised four interconnected phases: (a) corpus review, description and cleaning; (b) exploratory analysis of discrete events and continuous signals; (c) development and analytical refinement of the coding framework; and (d) focal analysis and multimodal triangulation of selected cases. These phases were operationalised through the six procedural stages described in Section 2.6. The researchers adopted an iterative approach, alternating between exploratory coding and corpus-level analysis. By systematically interpreting the 17 focal cases, the team progressively refined coding categories and decision rules. This process allowed for the identification of data limitations and interpretative ambiguities while surfacing critical interactional patterns, reflecting the established value of pilot work in refining qualitative procedures (Malmqvist et al., 2019).

2.2. Didascalias VC

Didascalias VC, developed by the Universidad Complutense de Madrid and the Universitat Autònoma de Barcelona (Madrid and Barcelona, Spain), was the immersive simulation environment in which the sessions analysed in this study were conducted. As introduced above, it is a VR platform for teacher education focused on classroom management. In each session, a pre-service teacher wearing a VR headset interacted with a classroom populated by virtual students and conducted a short pedagogical activity while responding to classroom situations as they unfolded. The virtual students' behaviour evolved according to both predefined pedagogical scripts and facilitator-mediated interventions introduced through a web-based interface. This adaptive design allowed each simulation to follow a different interactional trajectory according to the participant's decisions and the unfolding classroom situation, while maintaining pedagogical consistency and traceability across sessions.

Didascalias VC was also designed to generate structured multimodal records for educational analysis. This design responds to a recognised gap in immersive learning analytics: the need to move from rich traces of activity to pedagogically interpretable evidence for feedback and teacher decision-making. Each simulation produced a structured record

comprising session metadata, facilitator-mediated TALK events, programmed challenging events and continuous voice and movement signals. These records were managed by Mission Control, the analytics infrastructure described in the following section, which made simulated professional activity traceable, retrievable and pedagogically interpretable for formative assessment and debriefing.

2.3. Mission Control

Mission Control is the technological infrastructure responsible for configuring simulation sessions, coordinating communication between the different system components and collecting the multimodal data generated during each simulation. In the context of this study, it is examined as the infrastructure that makes Didascalias VC sessions traceable, retrievable and available for pedagogical interpretation. Rather than being a standalone web application, Mission Control was designed as a modular architecture that emphasises secure access, reliable data collection and long-term preservation of simulation records.

The infrastructure is organised around two logically separated servers. The public-facing front-end server constitutes the only component directly accessible from the Internet. Besides hosting the facilitator dashboard, it operates as a reverse HTTP proxy that forwards authenticated requests to the internal back-end server. This architecture isolates the core services from direct external access, reducing the attack surface while simplifying network administration through a single public endpoint.

The back-end server manages session creation, real-time communication with the VR headsets, event processing and persistent data storage. Throughout each simulation, facilitators monitor the session through the dashboard and can introduce facilitator-mediated TALK events or launch programmed challenging events in real time while the participant interacts with the virtual environment. Although TALK events are verbal interventions rather than predefined behavioural incidents, some may function as non-programmed critical moments when they introduce pedagogical, relational or ethical demands requiring participant response or teacher intervention.

To support the long-term availability and integrity of the collected data, both servers use RAID-5 storage arrays and regular automated backups. Security is maintained using HTTPS connections protected by TLS certificates, alongside JWT-based authentication and user-level authorisation. Restricted administrative access is managed via SSH, employing public-key authentication over an institutional VPN. Database operations employ parameterised queries and validated inputs to prevent SQL injection attacks.

The modular architecture also facilitates future extensions, allowing additional services, such as live headset streaming, automatic transcription or AI-assisted learning analytics, to be incorporated without modifying the public interface. These extensions are intended to support the organisation and interpretation of evidence rather than to automate pedagogical judgement. This architectural flexibility supports the long-term preservation and accessibility of simulation data, ensuring that records remain usable and interpretable as the platform evolves over time.

Each simulation session is instantiated when the VR application requests a new session from the server, which generates a unique, ephemeral session identifier valid only for the duration of that session. Facilitators join the same session through the web dashboard by entering this identifier, without needing to be physically co-located with the VR participant. This mechanism links the VR headset, the facilitator dashboard and the session record within a single traceable interaction, while the persistent bidirectional connection between the headset and the server enables TALK events and programmed challenging events to be transmitted to the virtual classroom in real time.

Once a simulation session has been established, the VR headset periodically transmits behavioural data to the Mission Control back-end at fixed short intervals (Figure 2). Each transmitted record is classified according to its content. Records containing an explicit event type correspond to discrete classroom events, including facilitator-generated TALK events and programmed challenging events. Records indicating session termination mark the end of the simulation, whereas all remaining records are interpreted as continuous behavioural signals, including vocal activity and head and hand movement measurements. This early classification preserves the distinction between discrete pedagogical events and continuous behavioural traces while maintaining a common temporal reference for subsequent multimodal analysis. Once a session concludes, its records are stored persistently in the database, providing the durable analytical corpus used throughout this study. Sessions interrupted before proper closure may generate incomplete records, an issue addressed during the data-cleaning procedure described in Section 2.6.

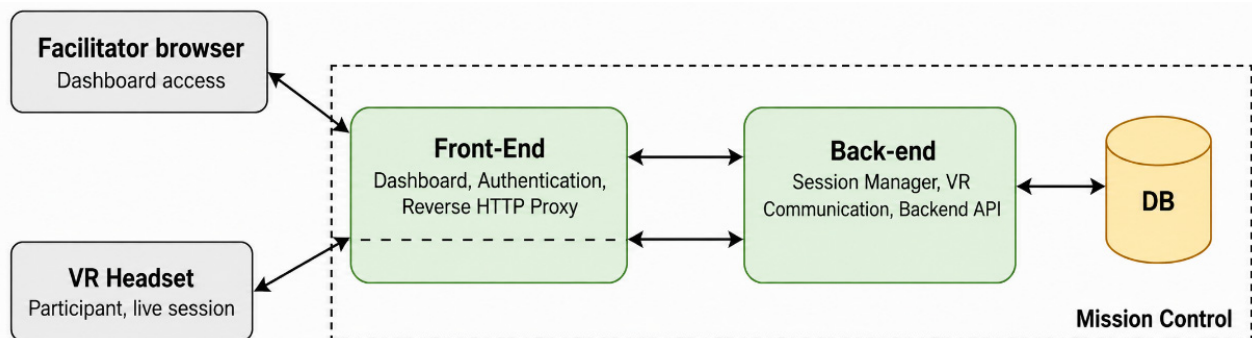


Figure 2. Didascalias VC and Mission Control architecture.

The main programmed challenging events identified in the corpus were STAND_UP, SEPARATE and INSULT. Unlike TALK events, these appeared as predefined uppercase dashboard entries and represented classroom incidents intentionally triggered by the facilitator during the simulation. Specifically, STAND_UP involved a student leaving their seat and moving around the classroom, SEPARATE involved a student attempting to sit next to another classmate despite previous arrangements, and INSULT involved a student producing a potentially inappropriate or offensive comment.

TALK events corresponded to brief verbal interventions introduced by the facilitator through Mission Control and delivered through one or more avatars during the simulation. These interventions could support the flow of classroom interaction, introduce pedagogical, relational or ethical demands, or respond to the participant's actions. Some TALK events were subsequently coded as critical TALK when they generated situations requiring pedagogical judgement or teacher intervention.

Continuous voice and movement signals were analysed as exploratory indicators of vocal and bodily activity. Because these signals are generated automatically at regular intervals, independently of discrete events, they were not interpreted as direct measures of emotion, stress, teaching competence or professional performance, but as complementary traces that could help identify potentially relevant moments for further qualitative examination.

Figure 3 illustrates how Mission Control represents an individual simulation session by integrating discrete events and continuous signals within a single temporal display. The upper graph shows the session timeline, vocal volume, head and hand movement signals, and discrete events such as TALK and INSULT. The lower section provides session metadata and timestamped events. This information was used to inspect vocal dynamics, event sequences and exploratory movement indicators.



Figure 3. Exportable analytics graph of a session in the Mission Control dashboard. The session statistics dashboard shows movement-derived variables and behavioural events. Non-English interface labels appearing in the dashboard correspond to standard interface terms: “Estadísticas de la sesión” = session statistics; “Gráfico de Estadísticas” = statistics graph; “Información” = information; “Eventos” = events; “Fecha de inicio” = start date; “Fecha de fin” = end date; “Volver a Analytics” = back to analytics. Speech-related events (TALK) are marked by vertical annotations. The corresponding English translations are: TALK (17:46:22): ‘I feel very bad.’ TALK (17:46:30): ‘They hit me.’ TALK (17:46:42): ‘No, because I feel unwell.’ TALK (17:46:55): ‘I hit them.’ INSULT (17:47:10). TALK (17:47:20): ‘It wasn’t for you, teacher, sorry.’

2.4. Corpus and Unit of Analysis

The initial corpus comprised 626 closed simulation sessions registered in Mission Control between April 2025 and February 2026. Together, these sessions accounted for 1929.83 min of simulation activity, equivalent to 32.16 h. Session duration varied considerably, with a mean of 3.08 min and a median of 3.00 min.

During corpus cleaning, 31 sessions were flagged because they were incomplete, anomalous, unusually short or potentially associated with technical testing. These sessions were excluded from the cleaned analytical corpus, which comprised 595 closed sessions, representing 95.0% of the initial corpus. Unless otherwise stated, the analyses reported in this article are based on these 595 sessions.

Each session corresponded to one application of the simulation in a class-group setting and was identified in Mission Control by its date and time of completion. The cleaned corpus was estimated to correspond to 558 unique participants. In most applications, each participant completed a single simulation, so that one session corresponded to one participant’s simulation experience. The documented exceptions concerned two Latin American university applications in which participants completed two simulations: 31 pre-service teachers in Chile generated 62 sessions, and 6 psychopedagogy students in Cuba generated 12 sessions. These 37 repeated participants therefore accounted for 74 sessions in the cleaned corpus.

As an example of cohort/session distribution, the Chilean implementation involved two Bachelor of Primary Education class groups that participated in two simulation moments approximately one week apart: the first group on 28 April and 9 May 2025, and the second group on 12 May and 16 May 2025. This example is reported because these project-specific applications included additional sociodemographic and implementation records not captured by Mission Control. As an example of a Spanish class-group application, the 28 November 2025 block corresponded to a Bachelor of Primary Education group from which the 17 focal multimodal cases with video recordings and transcripts were selected; this block involved one simulation per participant.

Mission Control did not include participant-level demographic fields or cohort identifiers in the version analysed in this study, such as degree programme, class group, university site or facilitator. For this reason, cohort/site distribution was reconstructed from implementation records and date-based session blocks rather than from dashboard metadata.

Accordingly, the session was retained as the primary unit of analysis because the purpose of the study was to examine the traceability, quality and interpretability of simulation records rather than to estimate individual learning trajectories.

Corpus cleaning was conducted to distinguish pedagogically interpretable simulation records from records that mainly reflected technical testing, familiarisation with the environment or incomplete use of the simulation. Sessions were flagged for review if they met one or more of the following criteria: (a) incomplete participation, such as failing to introduce oneself, conduct the class, or engage with the scenario; (b) insufficient or unreliable multimodal signals, such as missing voice data; (c) an unusually short duration of under one minute; or (d) evidence of technical testing rather than a complete pedagogical simulation.

A review of the excluded records showed that 19 sessions were flagged for incomplete participation or absence of an interpretable classroom sequence, 9 for insufficient or unreliable multimodal signal registration, 5 for an unusually short duration of under one minute, and 3 for evidence of technical testing or exploratory use. Because the exclusion criteria were not mutually exclusive, these counts exceed the 31 excluded sessions: three sessions were both incomplete and under one minute, and two sessions were both incomplete and had insufficient or unreliable signal registration. Some excluded records were consecutive or near-consecutive sessions registered on the same date and appeared to correspond to preliminary testing or familiarisation with the system before its use with teachers or volunteers.

The valid analytical window was defined as the interval within a session in which the participant was engaged in the pedagogical simulation and the recorded signals could reasonably be interpreted in relation to classroom interaction. Initial and final moments that appeared to correspond to equipment adjustment, controller manipulation, headset removal or other technical actions were not interpreted as evidence of classroom-management performance.

For descriptive characterisation of the initial corpus, 537 of the 626 sessions contained at least one registered dashboard event, whereas 89 contained none. A total of 3511 discrete events were recorded: 3314 TALK events and 197 programmed challenging events. The latter comprised 60 STAND_UP, 99 SEPARATE and 38 INSULT events.

The primary unit of analysis was therefore the closed simulation session. Nested analytical units were also examined during data exploration and framework development, including individual discrete events, TALK sequences, programmed challenging events and continuous-signal segments. In addition, a focal multimodal subsample of 17 sessions was analysed in greater depth using complementary records, including dashboard captures, CSV exports, video recordings and transcripts.

2.5. Data Sources

The study combined six types of data generated by or associated with Mission Control. First, session-level metadata were used to characterise the corpus and identify doubtful or atypical records. These included session identifiers, date, start time, end time, duration and session status.

Second, the analysis used discrete dashboard events, including TALK events and programmed challenging events. These records provided time-stamped markers of facilitator-mediated verbal interventions and dashboard-launched classroom incidents.

Third, the study examined continuous signals associated with voice and body movement. The vocal signal was operationalised as a relative measure of voice volume or vocal intensity, while movement-related signals included HeadVelocity, LeftHandVelocity and RightHandVelocity. These variables were not counted as events, but were analysed as exploratory indicators of activity, variation, peaks, phase changes and closing patterns.

Fourth, exportable dashboard visualisations were used when the quality of the graph was sufficient. These images allowed the research team to inspect the temporal distribution of events, the vocal curve and the movement signals.

Fifth, CSV files were used when valid exports were available. These files provided numerical records for some continuous signals, although the current export format presented limitations regarding temporal synchronisation, precision and integration between continuous signals and discrete events.

Finally, for a reduced set of cases, complementary audiovisual records were available. These videos were transcribed and used to contrast the dashboard records with the actual interactional sequence. The transcripts enriched the analysis by recovering participant responses, conversational turns, hesitations, interruptions, laughter, pauses and non-verbal actions that were not fully captured by dashboard TALK events alone.

2.6. Procedure

The methodological procedure was organised in six stages.

1. **Exploratory Coding:** Evaluating 178 sessions to establish baseline data quality and vocal indicators. This phase was used to examine the structure and quality of the available records and to develop initial categories related to discrete events, continuous vocal indicators, session phases, closing patterns and feedback potential.
2. **Corpus Refinement:** Cleaning the dataset to 595 valid sessions by excluding technical tests and incomplete records. Closed sessions were reviewed to identify records that could not be interpreted as complete pedagogical simulations. The cleaning procedure considered four operational criteria: incomplete participation in the expected simulation sequence, insufficient or unreliable multimodal signal registration, unusually short duration, defined as less than one minute, and evidence of technical testing or exploratory use of the VR environment. Because these criteria could overlap within the same record, they were used as exclusion indicators rather than mutually exclusive categories.
3. **Window Definition:** Isolating the ‘pedagogically interpretable’ interval to exclude technical setup and equipment removal. The full technical duration of each session was distinguished from the interval considered pedagogically interpretable, defined as the period in which the participant was engaged in the simulated classroom activity and the recorded signals could reasonably be related to classroom interaction. Initial and final segments associated with entering the VR environment, adjusting the headset, manipulating controllers, ending the session or removing the device were treated as potentially technical artefacts and were not interpreted as evidence of classroom-management performance.

4. Framework Development: Designing a coding structure across six thematic blocks (descriptive, quality, event, signal, sequential, and interpretive). These covered session context and usability; TALK and programmed challenging events; vocal and movement patterns; session phases, sequence patterns and intervention windows; and pedagogical or relational tensions, participant responses, emergent critical situations and feedback potential.
5. Category Testing: Refining rules through a consensus-based review of contrasting session records. The records represented different event configurations and interactional patterns, including STAND_UP, SEPARATE and INSULT events; critical TALK occurring before or after a programmed event; abrupt incidents without identifiable preceding verbal signals; and critical TALK without a programmed challenging event. These records were reviewed by members of the research team to identify ambiguities in category definitions, differences in interpretation and cases requiring clearer decision rules. Discrepancies were discussed until consensus was reached, and the coding framework was refined accordingly. This preliminary exercise supported the refinement of the categories and the reconceptualisation of feedback potential as a sequence-level rather than an event-level variable. It did not form part of the subsequent focal analysis.
6. Focal Multimodal Analysis: Triangulating the 17 purposively selected cases with video recordings and transcripts to examine and refine the framework. All cases were drawn from the same cohort of third-year Primary Education students. These participants had received identical prior instruction and practical guidance on inclusive education and classroom management. Furthermore, they completed the simulations as part of the practical component of a shared Learning and Development course, facilitated by the same teacher educator. This shared instructional context reduced variation attributable to differences in preparation, curriculum or facilitation.

The focal cases were selected purposively according to four inclusion criteria: availability of complementary records, including dashboard captures, CSV signal exports, video recordings and transcripts; variation in interactional trajectories; presence or absence of programmed challenging events; and pedagogical relevance for testing the coding framework. The selection therefore included sessions with programmed challenging events and sessions without such events but containing sensitive or pedagogically relevant interactions, such as help-seeking, discrimination, vulnerability, task resistance, bodily discomfort, role challenge and disengagement. The absence of programmed events in several cases was consistent with the personalised and pedagogically responsive approach of Didascalias VC, in which challenges are introduced according to the participant's unfolding performance and the learning opportunities emerging during the simulation.

The number of focal cases was determined by the methodological purpose of the pilot study. The aim was not to obtain a statistically representative subsample of the 595 retained sessions, but to apply, examine and refine the framework across sufficiently varied multimodal sequences. The 17 cases provided a manageable corpus for detailed sequence reconstruction while preserving variation in data availability, event configuration and pedagogical demand. These materials enabled dashboard events and continuous signals to be triangulated with participant responses and interactional details not consistently represented in the event logs. The 17 focal cases were used to apply and refine the framework across varied sequences and to examine feedback potential as a property of the interaction as a whole rather than of isolated events.

ChatGPT (GPT-5.6 Thinking, accessed through a ChatGPT Plus subscription; OpenAI, San Francisco, CA, USA) was used as an auxiliary tool to support the organisation and preliminary exploration of the large and heterogeneous multimodal corpus, including dashboard records, exported files, transcripts and evolving coding documents. Consistent with Morgan (2023) and recommendations for transparent, iterative and human-supervised AI-assisted qualitative analysis (Zhang et al., 2025), its outputs were restricted to organisational and exploratory tasks and were checked systematically against the original records. All category definitions, coding decisions, case interpretations and pedagogical judgements were developed and reviewed by the research team; no analytical result or conclusion was accepted without human verification.

2.7. Data Analysis

The analysis combined descriptive, sequential and multimodal approaches. Figure 4 summarises the analytical pathway through which Mission Control records were transformed into multimodal assessment evidence.

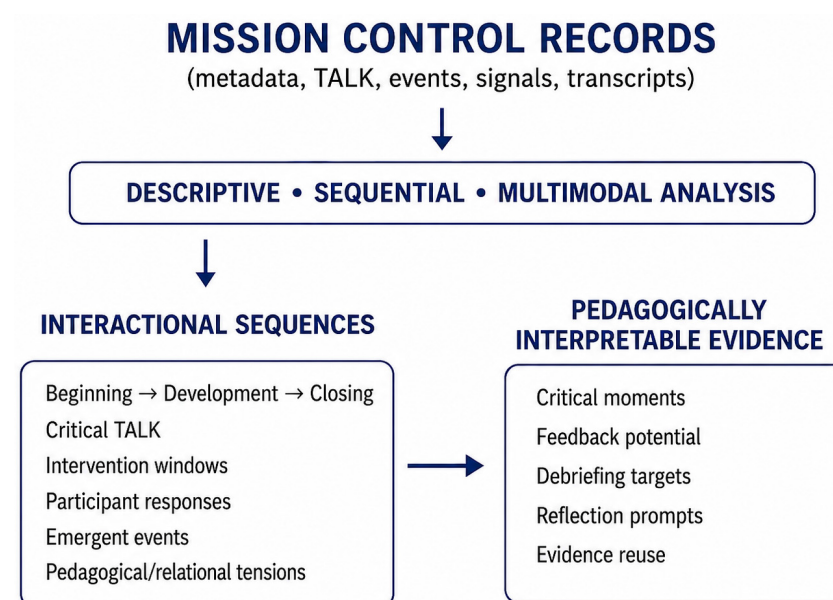


Figure 4. Analytical pathway for transforming Mission Control records into multimodal assessment evidence.

The descriptive analysis was used to characterise the corpus in terms of session frequency, session duration, event distribution, occurrence of TALK events, programmed challenging events and data quality issues identified during corpus cleaning. This level of analysis made it possible to organise the reference corpus, distinguish between closed and doubtful records, and identify the availability of different data layers across sessions.

The sequential analysis focused on the temporal organisation of simulation sessions. Each session was examined as an interactional sequence comprising a beginning, development and closing phase. Events were analysed according to their order, timing and relationship with surrounding events. Particular attention was given to the relationship between critical TALK and programmed challenging events, including the identification of possible intervention opportunities and different sequence configurations, such as abrupt, progressive or interactionally embedded incidents.

For this purpose, an intervention window was defined as the time interval between the last critical TALK preceding a programmed challenging event and the event itself. In this study, TALK refers to short natural-language utterances introduced by the facilitator through the Mission Control chat during the simulation and reproduced by one or more

avatars in the virtual classroom. These utterances were designed to maintain interactional realism, adapt to participants' decisions or discourse, and provide personalised prompts during the simulated lesson. They differ from programmed challenging events, which are predefined dashboard-triggered incidents such as STAND_UP, SEPARATE or INSULT.

Critical TALK refers to facilitator-generated avatar utterances that introduced relevant pedagogical, relational or ethical demands, such as task resistance, peer conflict, help-seeking, discomfort, exclusion or potential discrimination. Intervention windows were classified according to their duration, coded as no identifiable window when a programmed challenging event occurred without an identifiable preceding critical TALK, and coded as not applicable when no programmed challenging event occurred in the session.

The multimodal analysis integrated dashboard events, continuous signals, CSV exports, graphical representations, video recordings and transcripts. As illustrated in Figure 3, rather than treating events as isolated dashboard entries, the analysis examined how different forms of evidence combined within broader interactional sequences. This sequence-based perspective guided the interpretation of classroom-management situations throughout the study and supported the identification of critical verbal interactions, participant responses, sensitive tensions and emergent events that were not necessarily represented as programmed dashboard incidents.

Continuous voice and movement signals were analysed as exploratory indicators of activity rather than direct measures of competence or emotional states. Signal analysis focused on indicators such as coverage, variability, range, continuity, peaks and closing patterns. Exported graphs were used to examine temporal patterns and discontinuities, while CSV files provided numerical values for estimating signal availability and variation. These signals were interpreted in relation to the interactional sequence and not as stand-alone evidence of performance, stress or professional competence.

In the version of Mission Control analysed in this study, exports did not provide automated synchronisation across event logs, CSV signal files, visual graphs and audiovisual recordings. Temporal alignment was therefore performed manually. Time-stamped events were used as reference points to compare dashboard records with exported signal patterns and, where available, video recordings and transcripts. This procedure allowed sequence reconstruction, but it also required interpretive verification when sampling intervals, graph visualisation and audiovisual timing could not be assumed to align exactly across sources.

Finally, feedback potential was coded as a sequence-level property indicating the extent to which a record could support meaningful debriefing and formative feedback. It did not assess the quality of the participant's performance, but the pedagogical usefulness of the available evidence for teacher education. This coding considered the richness of the interactional sequence, the presence of critical TALK or challenging events, participant responses, multimodal evidence and the pedagogical significance of the situation.

Feedback potential was considered high when a sequence included: (a) an identifiable pedagogical, relational or ethical tension; (b) a participant response that could be discussed in relation to classroom management; (c) sufficient contextual or multimodal evidence to reconstruct how the situation unfolded; and (d) a clear opportunity for formative questioning, reflection or alternative action planning. Sequences were considered to have lower feedback potential when records were too fragmented, lacked participant response evidence, contained only isolated technical traces, or did not provide a pedagogically meaningful focus for debriefing.

Because the framework was being developed and refined during this methodological pilot, coding validation followed a consensus-based rather than a statistical reliability procedure. Selected contrasting records and focal cases were reviewed by four members of the research team, combining psychopedagogical and technological perspectives. Two

researchers focused on the pedagogical, relational and formative-assessment interpretation of the sequences, while two researchers examined the technical structure of the records, event registration and signal availability. The consensus process was used to examine whether the categories were conceptually clear, applicable to different sequence types and useful for identifying possible focuses for formative feedback. Differences in interpretation were discussed until consensus was reached, and category definitions and decision rules were adjusted accordingly.

This process was conducted iteratively and built on the stages described in Section 2.6. Initial category development was led by the principal researcher and the lead engineer responsible for Mission Control data capture and signal management, resulting in a preliminary illustrated codebook organised around thematic categories. The framework was subsequently reviewed by specialists in pedagogical design and simulation technology, whose complementary expertise contributed to refining category boundaries, signal definitions and coding rules. The incorporation of video transcripts was particularly relevant, as it revealed interactional and contextual elements not fully represented in dashboard records alone and supported the refinement of sequence interpretation, tension typologies and feedback-potential coding. The framework was then jointly applied and reviewed across the focal cases by the psychopedagogical researchers, leading to further refinement of category definitions, tension typologies and feedback-potential classifications.

Several consensus decisions illustrate how the framework was refined. First, the team agreed to distinguish discrete events from continuous signals because they required different analytical treatment: TALK events and programmed challenging events were treated as discrete temporal markers within the sequence, whereas vocal volume and movement traces were analysed as continuous exploratory indicators. Second, the team refined the distinction between TALK events and programmed challenging events: TALK referred to facilitator-mediated verbal interventions registered during personalised monitoring of the participant's performance, whereas programmed challenging events referred to predefined dashboard-launched incidents, such as STAND_UP, SEPARATE and INSULT. Third, the valid analytical window was limited to the period in which classroom activity was pedagogically interpretable, excluding initial or final segments mainly associated with headset adjustment, controller manipulation or equipment removal.

Formal inter-rater reliability statistics, such as Cohen's Kappa or percentage agreement, were not calculated at this stage because the primary objective of the pilot was the iterative development and refinement of the coding framework rather than the assessment of reliability of a fixed instrument. Such statistics would be more appropriate in subsequent studies once the framework reaches greater conceptual and operational stability.

The resulting framework should therefore be understood as a preliminary, consensus-refined analytical structure. Its use in this study supports the reconstruction and pedagogical interpretation of the focal cases, but does not provide evidence of independent inter-rater reliability. Accordingly, the framework should not be interpreted as a validated coding instrument at this stage. Overall, the proposed framework was examined according to its capacity to organise heterogeneous simulation records, support sequence reconstruction and identify possible focuses for formative feedback.

2.8. Ethical and Data Considerations

This study examined an assessment infrastructure rather than individual participants. The simulations were conducted within regular teacher education courses, but participation in the research component was voluntary and had no consequences for academic assessment. Multimodal records were analysed as contextual evidence within interactional

sequences, with voice and movement data treated as exploratory indicators rather than direct measures of emotion, stress, competence or teaching effectiveness.

The coding framework was designed for formative feedback and methodological examination, not summative assessment or participant ranking. To protect privacy, the sample anonymised data deposited in the Zenodo repository include only anonymised records, transcript excerpts and illustrative coding materials; identifiable audiovisual data are not publicly shared.

3. Results

3.1. Mission Control Data: Affordances and Limitations for Formative Assessment

The first result of the methodological examination was the identification of what Mission Control records make visible—and what they leave only partially visible—for formative assessment. The analysis showed that Mission Control generates a multilayered record of immersive classroom simulations, combining session metadata, discrete event logs, programmed challenging events, continuous voice and movement signals, visual analytics and, in selected cases, exportable CSV files. Complementary audiovisual records were also available for selected cases. Together, these sources preserved traces of the temporal organisation of the sessions and of specific moments of interaction.

The different data layers varied in assessment value. Session metadata supported corpus organisation, cleaning and traceability, but did not provide direct evidence of pedagogical action. Discrete dashboard events offered entry points into the interactional sequence, although they did not capture the full dialogue or the participant's complete response. Programmed challenging events marked planned classroom disruptions, but their occurrence alone did not determine formative relevance. Continuous voice and movement signals added complementary information, but functioned only as exploratory indicators requiring contextual interpretation.

The event distribution reinforced this distinction. For descriptive characterisation of the initial corpus, 537 of the 626 closed sessions contained at least one registered dashboard event. The 3511 recorded events were strongly dominated by TALK events: 3314 were TALK records, whereas 197 corresponded to programmed challenging events. Most of the interactional activity represented in the dashboard therefore concerned verbal interventions rather than predefined incidents. Restricting the analysis to programmed challenging events would have excluded a substantial part of the potentially relevant information. Verbal interactions included demands related to task resistance, help-seeking, relational tension, discrimination, vulnerability and disengagement.

The 89 sessions without registered dashboard events further showed that event absence could not, by itself, characterise the activity developed during a session. Some of these sessions still displayed vocal activity or potentially relevant changes in the continuous signal. Conversely, sessions containing programmed challenging events varied in the richness and interpretability of the surrounding records. The assessment value of the data therefore depended on reconstructing the broader interactional sequence within which programmed challenging events, verbal interventions, participant responses and continuous signals acquired meaning.

Across the corpus, Mission Control offered three main affordances for formative assessment: traceability, temporal organisation and multimodal triangulation. Time-stamped records allowed sessions, events and signals to be located and compared. The temporal structure of the dashboard supported examination of the beginning, development and closing phases of each session. In addition, dashboard records could be related to exported signals, visual graphs and, where available, audiovisual records and transcripts.

The analysis also identified four principal limitations. The TALK register was selective and did not constitute a full transcript of the simulation. Continuous voice and movement signals were not self-explanatory and could not be treated as direct indicators of emotion, stress, competence or performance. The alignment of discrete events, continuous signals, visual graphs, CSV exports and audiovisual records required manual verification. Some sessions also required cleaning because their records were doubtful, atypical, very short or potentially affected by technical actions at the beginning or end of the simulation. Table 1 summarises the assessment affordances and limitations associated with each data layer. It therefore characterises the data infrastructure itself, before the subsequent tables address analytical transformation and coding structure.

Table 1. Data layers generated or connected through Mission Control: assessment affordances and limitations.

Data Layer	Assessment Affordance	Main Limitation
Session metadata	Supports corpus organisation, traceability and cleaning	Does not provide direct evidence of pedagogical action
TALK events	Identifies facilitator-mediated verbal interaction points and possible pedagogical, relational or ethical demands	Does not capture the full dialogue or participant response
Programmed challenging events	Marks predefined dashboard-launched incidents such as STAND_UP, SEPARATE and INSULT	Event occurrence does not determine formative value without sequence-level interpretation
Continuous voice signals	Helps locate changes, discontinuities, peaks and patterns of vocal activity	Cannot be interpreted as emotion, stress or competence
Movement signals	Provides exploratory information about bodily activity	Requires contextual validation and may include technical artefacts
Exportable graphs and CSV files	Enable inspection and partial quantification of signals	Require alignment with events, video and transcripts
Video and transcripts	Recover participant responses, conversational developments, non-verbal actions and interactional details not fully represented in dashboard records	Video is not currently integrated into Mission Control or automatically linked to session identifiers, requiring manual matching and verification

3.2. Transforming Technical Records into Multimodal Assessment Evidence

The second result was the development and refinement of a procedure for transforming technical records into multimodal assessment evidence. As summarised in Table 2, Mission Control data were converted into analytically meaningful indicators and interpreted within broader interactional sequences. Rather than analysing events and signals in isolation, the procedure combined session descriptors, discrete events, continuous signals and complementary records—exported graphs, CSV files, videos and transcripts—to reconstruct pedagogically meaningful classroom situations.

Table 2. Analytical transformation of Mission Control records into multimodal assessment evidence.

Mission Control Record	Analytical Transformation	Analytical Contribution
Session metadata	Duration, event density and session type	Corpus cleaning, contextualisation and case selection
TALK events	TALK type, criticality, sensitivity, pedagogical demand and participant response	Identification of pedagogical, relational and ethical demands and feedback focus
Programmed challenging events: STAND_UP, SEPARATE, INSULT	Event type, sequence position and intervention window	Examination of escalation, response patterns and feedback focus
Vocal volume curve	Signal continuity, variability, visual peaks, phase changes and closing pattern	Identification of vocal activity patterns and temporally salient moments
Movement signals	Head and hand activity, movement profile and peaks	Complementary interpretation of bodily activity within the interaction
Exportable graphs	Temporal distribution of events, vocal and movement signals	Inspection of sequence organisation, peaks and data quality
CSV exports	Numerical signal values and coverage indicators	Estimation of signal availability, variability and range
Video recordings and transcripts	Participant responses, conversational turns, pauses, hesitations, interruptions, laughter, non-verbal actions and emergent critical situations	Verification of event interpretation, reconstruction of interactional sequences and identification of feedback focus

Session descriptors provided the contextual structure of the corpus. They supported data cleaning, the identification of incomplete, atypical or potentially unreliable records, and the contextualisation and selection of sessions for further analysis.

Discrete events marked identifiable moments within the simulations. The analysis distinguished between facilitator-mediated verbal interventions registered as TALK and programmed challenging events, allowing verbal demands and interactional tensions to be examined independently of predefined incidents. This distinction also helped identify escalation patterns, participant responses and intervention windows informing the potential and focus of formative feedback. Continuous signals provided complementary evidence of vocal and bodily activity. They were treated not as direct measures of competence or emotional states, but as exploratory markers of continuity, variability, discontinuity and change. Exported graphs supported the identification of temporal patterns, peaks and phase transitions, while CSV files provided numerical information about signal coverage, availability and variation.

The analytical transformation involved moving from isolated records to interactional sequences. TALK events were analysed in relation to their content, timing, pedagogical significance and connection to subsequent actions. Programmed challenging events were interpreted in relation to the verbal interactions that preceded or followed them. The

intervention window between a critical verbal interaction and a subsequent programmed event was also coded, enabling distinctions between immediate disruption, brief opportunities for response and sequences in which no programmed event followed. This sequence-based perspective supported the identification of critical moments, escalation or interruption patterns, participant responses, intervention windows and potential targets for formative feedback.

Table 2 summarises how each record type was transformed and the analytical contribution it made to the reconstruction of multimodal assessment evidence. Unlike Table 1, which describes the available data layers, Table 2 focuses on the analytical operations through which these records were converted into interpretable evidence.

The focal multimodal analysis of 17 selected cases was used to apply, examine and refine the procedure in sessions with richer complementary records. Case selection considered variation in interactional trajectories and the availability of CSV exports, dashboard captures, video recordings and transcripts. Most focal cases did not contain programmed challenging events, a distribution consistent with the personalised and pedagogically responsive approach of Didascalias VC, in which such events are introduced according to participants' unfolding responses and emerging learning opportunities.

Video recordings allowed classroom interactions to be transcribed and compared with dashboard records, showing that the TALK register represented only part of the interactional sequence. Transcripts and video material revealed participant responses, conversational turns, pauses, hesitations, interruptions, laughter and non-verbal actions that were not consistently visible in the dashboard. They also made it possible to examine whether participants responded during available intervention windows and to identify the pedagogical or relational focus around which feedback could be organised.

The focal analysis therefore showed that dashboard records were insufficient as stand-alone sources for reconstructing the interaction in full. Their interpretation was strengthened through triangulation with audiovisual records and transcripts. The combined dataset supported the verification of event interpretations, the reconstruction of broader interactional sequences, the identification of pedagogical, relational and ethical demands only partially represented in the technical records, and the classification of sequences according to their feedback potential and possible debriefing focus.

3.3. Structure and Analytical Refinement of the Multimodal Coding Framework

The third result was the development of a preliminary multimodal coding framework for organising technical and complementary records and supporting their pedagogical interpretation. The framework comprised six blocks of variables: descriptive variables, cleaning and data quality variables, event variables, continuous signal variables, sequential variables and interpretive variables.

Descriptive variables identified and contextualised each session. Cleaning and data quality variables documented whether the available record could be included in the cleaned corpus or required further review. Event variables classified discrete dashboard records, including TALK, critical TALK and programmed challenging events. Continuous signal variables described exploratory patterns of voice and movement activity. Sequential variables reconstructed the temporal organisation of the session and the relationships among verbal interactions, programmed events, participant responses and intervention windows. Interpretive variables connected these reconstructed sequences with pedagogical and relational demands, including type of tension, participant response, emergent critical situations and feedback potential.

Table 3 summarises the six blocks, their main functions, and representative variables. Whereas Table 2 presents the transformation from records to evidence, Table 3 presents the resulting coding architecture used to organise and interpret that evidence.

Table 3. Coding architecture of the multimodal framework: blocks, functions and representative variables.

Block of Variables	Main Function	Examples of Variables
Descriptive variables	Identify and contextualise the session	Session ID, date, duration, start time, end time
Cleaning and data quality variables	Assess the completeness and usability of the available record	Doubtful session, reason for review, inclusion decision
Event variables	Describe and classify discrete dashboard events	TALK, critical TALK, STAND_UP, SEPARATE, INSULT, event density
Continuous signal variables	Describe exploratory patterns of voice and movement activity	Vocal coverage, vocal variability, range, peaks, closing pattern, movement profile
Sequential variables	Reconstruct temporal and interactional organisation	Beginning, development, closing, sequence pattern, event position, intervention window
Interpretive variables	Relate reconstructed sequences to pedagogical interpretation	Type of tension, participant response, emergent critical situation, feedback potential, possible feedback focus

A central outcome of the framework was the distinction between event-level and sequence-level interpretation. During the preliminary exploratory phase, feedback potential was considered in relation to individual events. The subsequent focal analysis showed, however, that this potential was more consistently determined by the broader interactional sequence. Sessions containing the same programmed event differed in their pedagogical significance, while some sessions without programmed challenging events were coded as having high feedback potential because they contained sensitive TALK, participant responses, discrimination, exclusion, help-seeking or relational tension. Feedback potential was therefore redefined as a sequence-level variable rather than as a property of an isolated event.

The framework also refined the coding of intervention windows. A distinction was introduced between no identifiable window and not applicable. No identifiable window was assigned when a programmed challenging event occurred without an identifiable preceding critical TALK from which an intervention interval could be calculated. Not applicable was used when no programmed challenging event occurred in the session. This distinction prevented sessions without programmed events from being treated as equivalent to sequences in which an event occurred without a detectable pre-incident opportunity for intervention.

The focal cases also showed that pedagogical situations frequently involved overlapping demands. Task resistance could coincide with peer conflict, exclusion, discrimination, help-seeking or group-management difficulties within the same sequence. Type of tension was therefore coded as a multi-label variable, allowing more than one pedagogical or relational demand to be assigned when supported by the available records.

Together, these refinements shifted the framework away from counting isolated events and towards reconstructing how tensions, responses and opportunities for intervention developed over time. They also enabled feedback potential to be coded in relation to the sequence as a whole and to the specific aspects of participant performance that could warrant closer examination.

3.4. From Events to Interactional Sequences: Contributions of the Focal Case Analysis

The focal case analysis contributed to refining the coding framework and establishing how it could support the pedagogical interpretation of classroom-management performance. Before this analysis, the emerging categories and coding rules had been explored using a small set of contrasting session records representing different event configurations and interactional patterns. This preliminary exercise indicated that analytically relevant material was not confined to programmed challenging events. Critical TALK preceding an event could provide information about emerging tension and possible opportunities for intervention, whereas verbal interactions following an event could reveal exclusion, discrimination, vulnerability, help-seeking or relational conflict. These observations supported the inclusion of critical TALK occurring both before and after programmed events and reinforced the need to interpret events within their broader interactional sequence.

The systematic focal analysis was then conducted on 17 selected cases for which richer complementary records were available. Video recordings and transcripts made it possible to reconstruct participant responses and interactional developments that were only partially represented in dashboard records. These complementary sources showed how participants acknowledged, redirected, questioned, ignored or addressed emerging classroom demands and enabled the interpretation of pauses, hesitations, interruptions, laughter and non-verbal actions within the unfolding interaction.

As shown in Table 4, only two focal cases were primarily organised around programmed dashboard incidents. Most derived their analytical contribution from sensitive verbal interactions, help-seeking, discrimination, vulnerability, task resistance or ambiguous situations requiring pedagogical judgement. The categories in the table represent the dominant contribution of each case rather than mutually exclusive classifications, since several sequences involved overlapping tensions. Appendix A summarises the 17 focal cases, including their main sequence types, dashboard-launched programmed events, dominant pedagogical or relational tensions, feedback potential and associated non-exclusive typologies. Appendix B presents a worked example of how a sensitive TALK-only sequence was transformed into pedagogically interpretable evidence.

Table 4. Analytical contribution of the 17 focal multimodal cases.

Type of Focal Case	Sessions	Main Analytical Contribution
Programmed-event sessions	2	Examination of STAND_UP, SEPARATE and INSULT within their surrounding interactional sequences
Sensitive sequences without programmed events	9	Identification of discrimination, vulnerability, help-seeking and identity-related tensions through TALK and participant responses
Resistance, disengagement or role-challenge sequences	4	Analysis of task resistance, engagement difficulties and challenges to the participant's instructional role
Bodily discomfort or exit-from-class sequences	2	Interpretation of ambiguity among avoidance, physical need, discomfort and classroom management demands

Note: The categories indicate the dominant analytical contribution used to summarise the 17 focal cases at an aggregate level. The case-by-case assignment underlying these counts is reported in Table A1, while additional non-exclusive typologies are presented in Table A2. Several cases involved more than one pedagogical or relational tension.

Across the focal cases, sessions without programmed challenging events frequently contained sequences coded as having substantial feedback potential, whereas the occurrence of a programmed event did not by itself determine the richness or interpretability of the available evidence. The formative relevance of a record depended on the relationship among verbal demands, participant responses, timing, subsequent developments and the surrounding pedagogical situation.

These findings led to the principal refinement of the framework: the interactional sequence, rather than the isolated event, was adopted as the main unit of interpretation. This made it possible to identify emerging tensions, anticipation cues, participant responses, intervention windows and possible focuses for feedback. It also allowed programmed challenging events, TALK-only sequences and emergent critical situations to be examined through a common analytical structure.

The focal analysis therefore clarified how the coding framework can support formative assessment of classroom-management competence. It organises heterogeneous records, reconstructs how pedagogical and relational demands develop over time and identifies aspects of participants' responses that may warrant closer examination during debriefing. Its contribution at this stage lies in making potential evidence more traceable and interpretable; its use by teacher educators and its effects on feedback and subsequent professional learning remain to be examined.

4. Discussion

This methodological pilot examined how technical and multimodal records generated during Didascalias VC simulations can be transformed into evidence for formative assessment. The findings show that Mission Control provides traceable but unevenly complete records whose pedagogical significance depends on systematic cleaning, temporal alignment, multimodal triangulation and contextual interpretation. Their educational value does not derive from the volume of data collected or from any single event or signal, but from the reconstruction of interactional sequences in which verbal interventions, programmed events, participant responses and continuous signals acquire meaning. The preliminary coding framework supports this transformation by linking these traces to pedagogically relevant tensions, intervention windows and possible focuses for feedback. These findings are consistent with learning analytics research showing that digital traces acquire educational meaning only when they are connected to relevant constructs, learning objectives and contexts of use (Gašević et al., 2015, 2017).

The results address the three research questions in complementary ways. First, Mission Control makes several types of data available, including session metadata, discrete events, verbal interactions and continuous voice and movement signals, although their availability, level of detail and interpretability vary across sessions. Audiovisual recordings and transcripts provide complementary information, even though they are not currently integrated into the dashboard. Second, transforming these records into assessment evidence requires data cleaning, temporal alignment, multimodal triangulation and sequence-based interpretation. Third, the preliminary coding framework supports formative assessment by shifting the unit of interpretation from isolated technical markers to contextually reconstructed interactional sequences and by identifying tensions, participant responses, intervention windows and possible focuses for feedback.

The principal contribution of the study is therefore methodological. It shows that the assessment value of immersive simulation data lies neither in the accumulation of records nor in the detection of isolated events, but in a pedagogically grounded process through which multimodal traces are organised and interpreted as evidence. In this respect, the study extends current research on immersive learning analytics, which remains

comparatively underdeveloped in its theoretical grounding, multimodal integration and support for educators' decision-making (Tao et al., 2025).

4.1. From Raw Simulation Data to Assessment Evidence

The findings reinforce a central premise of learning analytics: educational value does not arise from data capture alone, but from the processes through which records are prepared, interpreted and connected to pedagogical purposes. Learning analytics can support assessment by making patterns of activity visible, helping educators examine learning processes and providing information that can inform feedback and decision-making (Banihashem et al., 2022; Qazi & Pachler, 2025). Achieving this potential requires reliable data, interpretable features and a clear relationship between the available records and the educational constructs of interest (Gašević et al., 2015, 2017).

The present findings illustrate these requirements in an immersive professional simulation. Mission Control captures session metadata, verbal interactions, discrete events and continuous signals, but these raw records offer only partial representations of what occurred. TALK records had to be interpreted in relation to their content and temporal position; programmed events had to be situated within preceding and subsequent interactions; and voice and movement signals had to be aligned with participant responses and the broader classroom situation. Data preparation, temporal alignment and contextual interpretation were therefore integral to the construction of assessment evidence rather than preliminary technical operations.

This process also required pedagogical decisions about what should count as relevant evidence. The coding framework did not simply extract variables from the available records. It organised them around classroom management constructs, pedagogical and relational tensions, temporal sequences and opportunities for feedback. In this respect, the findings support the argument that learning analytics should remain centred on learning and educational action rather than on the availability of increasingly large or technically sophisticated datasets (Gašević et al., 2015).

The findings consequently define both the potential and the current limits of Mission Control. The infrastructure can make aspects of professional action available for analysis, but it does not yet convert them automatically into formative assessment. Its records become useful when they are selected, combined and interpreted in relation to pedagogical questions. From a technological mediation perspective, these records are not direct representations of classroom-management competence; they are technologically produced traces whose visibility and meaning are shaped by the system through which they are captured and represented (Voordijk & Vahdatikhaki, 2022).

The present study therefore advances from raw simulation records to pedagogically interpretable evidence, while the educational application of these analytics remains to be tested. Ruipérez-Valiente (2020) emphasises that the learning analytics cycle is completed when analytical outputs return to the educational context and are used to support feedback, intervention and improvement. In this pilot, the coding framework identifies sequences with potential value for debriefing and formative feedback, but it does not yet establish how teacher educators and participants use this evidence or whether such use influences subsequent performance.

4.2. Interactional Sequences as the Unit of Assessment

A central finding is that isolated events are insufficient as units of assessment. TALK events and programmed challenging events such as STAND_UP, SEPARATE and INSULT provide useful temporal markers, but their pedagogical significance depends on how they function within the broader interaction. Sessions containing the same programmed incident

differed in interpretability, complexity and feedback potential, while some sessions without programmed challenging events contained highly relevant material for pedagogical interpretation.

These differences reinforce the argument that behavioural traces and system events cannot be treated as transparent indicators of learning or competence (Gašević et al., 2017). An event becomes meaningful only when interpreted in relation to what preceded it, how the participant responded and how the simulated situation subsequently developed. This is particularly important in classroom management, a complex area for pre-service teachers in which professional performance unfolds over time through the interpretation of emerging situations, verbal and non-verbal interaction, and the adaptation of pedagogical responses as events develop (Blikstein & Worsley, 2016; Han et al., 2025; Li et al., 2024; Street et al., 2025). The same programmed incident may therefore reflect delayed intervention, appropriate restraint, relational sensitivity, escalation or successful de-escalation, depending on the surrounding sequence. Assessing classroom-management competence consequently requires attention to the temporal and relational organisation of participants' actions rather than to individual markers considered in isolation.

The focal analysis provided empirical support for adopting the interactional sequence as the main unit of interpretation. Critical TALK occurring before an event revealed emerging tensions and possible intervention windows, whereas interactions following an event helped reconstruct participant responses and subsequent developments. The distinction between an identifiable intervention window, the absence of a preceding critical TALK and cases in which no programmed event occurred also prevented different temporal configurations from being treated as equivalent.

The cases further showed that pedagogical situations frequently combined several demands. Task resistance could coincide with peer conflict, exclusion, discrimination, help-seeking or group-management difficulties within the same sequence. Treating the type of tension as a multi-label variable therefore preserved the complexity of classroom situations more adequately than assigning each case to a single category.

The sequence-based analysis also revealed that formative value may emerge beyond the incidents anticipated by the simulation design. In several focal cases, requests for help, resistance to a task, discriminatory comments, expressions of vulnerability, exclusion and hesitation yielded richer material for pedagogical interpretation than visible disruptive behaviour. This finding is especially relevant because most focal cases did not contain programmed challenging events. Their distribution reflects the personalised and pedagogically responsive approach underpinning Didascalias VC, in which challenges are introduced according to the participant's unfolding performance and the learning opportunities emerging during the simulation.

Video recordings and transcripts were essential for reconstructing interactional processes only partially represented in the dashboard. They revealed participant responses, conversational developments, pauses, hesitations, interruptions, laughter and non-verbal actions specific to each simulation trajectory. Their triangulation with dashboard records helped verify event interpretations and identify pedagogical, relational and ethical demands that would otherwise have remained only partially visible.

The analysis therefore supports a shift from isolated event detection to sequence-based assessment. This is consistent with process-focused approaches that direct attention towards the traceable development of actions, decisions, difficulties and revisions rather than towards static outcomes or isolated products (Winstone et al., 2026). Assessing classroom-management competence consequently requires attention to the relational and temporal organisation of participants' actions.

4.3. Teacher Interpretation and Evidence-Based Feedback

The findings reinforce the distinction between making activity visible and making it pedagogically actionable. Teacher-facing dashboards often increase awareness of what has occurred but provide more limited support for understanding why it matters or determining how teacher educators might use this information (Kaliisa et al., 2023). Mission Control currently shares some of these characteristics: it provides access to temporal markers, events and signals, but their pedagogical significance is not self-evident.

The coding framework helps address this limitation by connecting dashboard traces with interactional processes and potential questions for debriefing. Its purpose is not to produce automated evaluations of performance, but to support teacher educators in identifying episodes that warrant closer examination and interpreting them in relation to pedagogical purposes. This is consistent with research showing that learning analytics can strengthen feedback when analytical information is designed around educational goals and supports teachers' interpretative and decision-making processes (Banihashem et al., 2022).

This pedagogically responsive approach also has implications for feedback. Rather than organising debriefing solely around predefined incidents, teacher educators need to attend to the particular trajectory of each simulation and to the learning opportunities emerging from participants' responses, hesitations and decisions. This is especially relevant in a field where validation studies have often focused more strongly on implementation, educational effects and user experience than on the learning principles shaping immersive practice (Han et al., 2025; Tao et al., 2025).

The findings show that Mission Control records are valuable because they make aspects of unfolding professional action available for review, rather than because they generate definitive judgements about competence. The reconstructed sequences can help examine how participants interpret situations, respond to pedagogical and relational demands and adapt their actions over time. In this respect, the study extends to immersive professional simulations the process-oriented assessment principle proposed by Winstone et al. (2026), which emphasises making learning processes visible beyond static outcomes or final products.

Continuous voice and movement signals must be interpreted within this process-oriented logic. In the analysed sessions, they helped locate changes, interruptions, peaks and discontinuities in activity, but did not function as direct indicators of emotion, stress, competence or teaching quality. Their meaning depended on their temporal relationship with verbal exchanges, participant responses and the broader interactional situation. This interpretative caution is central to responsible multimodal learning analytics, since sensor-derived and behavioural signals require contextual triangulation and cannot be equated directly with internal states or educational constructs (Blikstein & Worsley, 2016; Sharma & Giannakos, 2020).

This interpretative caution also has practical implications for debriefing. For teacher educators, the framework offers a way to operationalise multimodal evidence in debriefing without requiring exhaustive inspection of all dashboard data. Episodes with high feedback potential can be selected when they include an identifiable pedagogical, relational or ethical tension, a discussable participant response, sufficient contextual or multimodal evidence to reconstruct the sequence and a clear opportunity for formative questioning or alternative action planning. In fast-paced debriefing contexts, this sequence-based selection may reduce cognitive load by focusing attention on a small number of interpretable classroom situations, discouraging isolated readings of voice or movement signals and organising feedback around what happened, how the participant responded and which pedagogical alternatives could be considered.

Even with this sequence-based support, teacher educators remain central to the use and further refinement of the framework. Their professional judgement is required to determine which evidence is relevant, interpret it in relation to intended competences and formulate feedback appropriate to the participant and the situation. Their involvement is also needed to refine the analytical categories and ensure that the information produced responds to the practical demands of debriefing and formative assessment. The framework supports this work by making evidence more accessible, traceable and manageable, thereby reducing the effort required to locate and reconstruct relevant episodes without predetermining the evaluation of performance.

The distinction between supporting evidence management and automating pedagogical judgement is also important for future AI-assisted functions. Automatic transcription, event detection, temporal alignment and sequence segmentation could reduce the effort required to retrieve and organise relevant evidence, while automated systems might suggest possible prompts for debriefing. Such functions should support pedagogical inquiry rather than generate decontextualised judgements of competence. The present pilot offers an initial methodological basis for these developments by identifying which sources need to be aligned, which temporal and contextual relationships must be preserved and why technical indicators require pedagogical interpretation before they can inform feedback or assessment (Gašević et al., 2017; Zhang et al., 2025).

4.4. Implications for Sustainable and Responsible Assessment Infrastructures

The findings also clarify how sustainability can be understood in relation to immersive assessment. Here, sustainability refers to the capacity to preserve, retrieve and reuse evidence in ways that support feedback, reflection and improvement over time. This interpretation is consistent with postdigital perspectives that locate sustainability in the coherent articulation of technology, pedagogy, institutional practices and educational purposes (Markauskaite et al., 2023).

From a technological mediation perspective, a VR environment does not simply reproduce professional practice; it shapes how that practice is experienced, represented and made available for interpretation (Voordijk & Vahdatikhaki, 2022). The examination of Mission Control must therefore address more than the technical availability of records. It must examine whether the traces produced by the system preserve the temporal, interactional and contextual relationships needed to reconstruct participants' performance in pedagogically meaningful ways. The variation observed in record availability and detail, together with the manual alignment required across data sources, indicates that this interpretative mediation depends on how effectively events, signals, verbal interactions and audiovisual materials can be connected.

Learning analytics can contribute to the optimisation of VR learning environments by making these mediated traces available for systematic review and by revealing which dimensions of the simulated experience are captured, obscured or insufficiently connected. When records remain traceable and the relationships among events, signals, verbal interactions and context are preserved, they can support debriefing, teacher feedback, participant reflection and comparisons across successive simulations. Evidence reuse should therefore be understood as part of a recursive process in which the analysis of participants' activity informs both formative assessment and the continued refinement of the learning environment.

Several priorities follow for the further development of Mission Control. Improving data integration, temporal synchronisation and export consistency would facilitate the reconstruction of multimodal sequences across a larger number of sessions. Integrating audiovisual records—or establishing a more direct and reliable link between these records

and their corresponding sessions—would reduce the manual effort required to recover interactional context. These improvements would also help teacher educators revisit relevant moments, compare successive performances and use the available evidence more efficiently during debriefing.

The findings also suggest that future dashboard functions should be organised around pedagogically relevant sequences rather than isolated events. Supporting the alignment of TALK records, programmed challenging events, continuous signals and participant responses would make it easier to identify intervention windows, emerging tensions and possible focuses for feedback. Involving teacher educators in the refinement of these functions will be essential to ensure that technical developments respond to the practical demands of formative assessment rather than merely increasing the volume of information displayed.

The transferability of the study lies less in the specific event labels and data structures of Didascalias VC than in the methodological procedure developed: examining the quality and accessibility of records, distinguishing technical traces from assessment evidence, triangulating multimodal sources, reconstructing interactional sequences and relating them to pedagogically relevant questions. This procedure may inform the evaluation and refinement of other immersive simulations in teacher education and related professional fields. Its transfer would require the coding categories and interpretative criteria to be adapted to the competences, scenarios and pedagogical purposes of each context. This responds to a broader limitation in immersive learning analytics, where multimodal integration and support for educators' interpretation remain underdeveloped (Tao et al., 2025).

Responsible implementation is inseparable from this sustainability. Voice, movement and audiovisual records concern forms of personal and professional behaviour that can easily be overinterpreted. Institutions must therefore define legitimate purposes for data collection and use, maintain informed consent, protect privacy, document analytical procedures and establish clear limits on what may be inferred from technical signals (Pardo & Siemens, 2014; Sharma & Giannakos, 2020). In particular, exploratory indicators should not be used to assign emotions, stress levels, competence or performance without validated interpretative models and contextual evidence.

Transparency is equally important. Teacher educators and participants should be able to understand what is recorded, how records are transformed and how they contribute to feedback. Sustainable implementation depends on trust in the infrastructure and on the capacity of its users to question, contextualise and interpret the evidence it produces.

4.5. Limitations and Future Work

The findings should be interpreted in relation to the methodological scope of the study. First, the pilot focused on establishing whether simulation records could be transformed into traceable and pedagogically interpretable evidence rather than on evaluating formative outcomes. It therefore does not establish learning gains, changes in classroom-management competence or the effectiveness of feedback derived from the framework. This delimitation identifies the next empirical step: integrating sequence-based evidence into debriefing and feedback and examining how its use supports reflection, subsequent performance and professional learning.

Second, the focal multimodal analysis was conducted on 17 purposively selected cases for which richer complementary records were available. These cases enabled the framework to be applied, examined and refined across varied interactional trajectories, but they were not intended to represent the full corpus statistically. This selection strategy may have increased the visibility of pedagogically rich or more interpretable sequences, while sessions with fewer complementary records may have offered less accessible evidence

for sequence reconstruction. The findings derived from the focal cases should therefore be understood as demonstrating the analytical potential of the framework rather than estimating the prevalence of specific patterns in the full corpus. Applying the framework to a larger and more diverse set of sessions, scenarios, programmed challenging events and emergent tensions would allow the stability, range and transferability of the identified sequence patterns to be examined more systematically.

Third, the availability and level of detail of the records varied across sessions, reflecting the current stage of development of Mission Control and the heterogeneous sources included in the analysis. The dashboard provided useful temporal and event-based traces, while fuller interactional detail was recovered through audiovisual records and transcripts. Voice and movement signals were not available in all sessions, and aligning dashboard events, exported data, graphs and audiovisual materials still required manual processing. This uneven availability of records may affect the credibility of sequence-level interpretations when complementary sources are missing or incomplete, because some pedagogically relevant responses may remain only partially visible in the dashboard data. For this reason, the study treats Mission Control records as partial traces requiring triangulation, rather than as complete representations of classroom-management performance.

A related technical limitation concerns temporal synchronisation across data sources in the version of Mission Control analysed in this study. CSV exports, dashboard graphs, discrete event logs and audiovisual recordings were not automatically aligned within a unified timeline. As a result, sequence reconstruction required manual matching and verification, and minor discrepancies in sampling intervals or visual representation may have affected the precision with which events and continuous signals were related. Future versions of the infrastructure should incorporate automated timestamp harmonisation, unified session identifiers and synchronised multimodal exports to strengthen reproducibility, reduce manual workload and improve data integration across sources.

Fourth, the coding framework remains preliminary. Its categories were developed and refined through the analysed corpus and require further examination of conceptual clarity, applicability across scenarios and consistency between coders. Although the consensus-based refinement procedure was appropriate for the methodological purpose of this pilot study, it does not provide independent evidence of inter-rater reliability. In particular, the classification of critical TALK, pedagogical tensions, intervention windows and feedback potential involves interpretative judgement. This may affect the framework's application in teaching practice if categories are used without shared definitions, examples and opportunities for professional discussion. In the present study, coding was used to organise evidence and identify possible focuses for debriefing, not to generate summative judgements or automated ratings of competence. Future work should include independent coding of a larger sample, agreement analysis and expert review with teacher educators to assess whether the categories are sufficiently clear, reliable, useful and feasible for formative practice.

These considerations also clarify the transferable contribution of the study. Although the event labels, data structures and technical configurations are specific to Didascalias VC, the methodological procedure has broader applicability: examining the quality and accessibility of available records, distinguishing technical traces from assessment evidence, triangulating multimodal sources, reconstructing interactional sequences and relating them to pedagogically relevant questions. Its transfer to other immersive professional simulations would require the coding categories and interpretative criteria to be adapted to the competences, scenarios and pedagogical purposes of each context.

The next stage should examine the educational use of the evidence produced through the framework. Future studies should investigate how teacher educators identify and interpret relevant sequences, how they translate them into debriefing questions and feed-

back, and how participants understand, evaluate and use that feedback, since feedback effectiveness also depends on learners' capacity to engage with and act on the information provided (Liu & Deris, 2025). This would make it possible to determine whether sequence-based evidence contributes to reflection, subsequent performance and the development of classroom-management competence.

Once the framework has been further examined with teacher educators, future analytical functions could support automatic transcription, temporal alignment, event detection, sequence segmentation and the identification of potentially relevant episodes. They might also suggest possible prompts for debriefing. Such functions should reduce the effort required to retrieve and organise evidence while preserving contextual relationships and leaving the interpretation of performance and competence to professional judgement.

Overall, the study demonstrates that the transition from immersive experience to formative assessment requires an explicit interpretative infrastructure. Its original contribution is a preliminary and potentially transferable procedure for transforming heterogeneous multimodal traces into contextually reconstructed evidence of professional action. This contribution provides a methodological basis for teacher interpretation, evidence-based feedback and the responsible development of future learning analytics in immersive professional education.

5. Conclusions

This methodological pilot examined how technical and multimodal records generated during Didascalias VC simulations can be transformed into pedagogically interpretable evidence for formative assessment. In response to the research questions, the study characterised the data made available through Mission Control, examined their affordances and limitations, and developed a preliminary sequence-based multimodal coding framework to explore how multimodal records may support the formative assessment of classroom-management competence in initial teacher education.

The findings show that simulation records do not constitute assessment evidence in themselves. Session metadata, TALK events, programmed challenging events, continuous voice and movement signals, visual graphs, CSV exports, video recordings and transcripts each provided partial representations of the simulated activity. Their pedagogical significance emerged through data cleaning, temporal alignment, multimodal triangulation and contextual interpretation. The interactional sequence, rather than the isolated event or signal, therefore proved to be the most informative unit for reconstructing how classroom situations developed and how participants responded to them.

The study makes three main contributions. First, it identifies the affordances and current scope of Mission Control as an emerging assessment infrastructure. Its time-stamped records support traceability and temporal analysis, while variation in data availability, detail and integration indicates priorities for further technical refinement. Second, it proposes a preliminary coding framework that organises descriptive, data-quality, event, continuous-signal, sequential and interpretive variables. The framework links technical traces with pedagogical and relational tensions, participant responses, intervention windows and possible focuses for feedback. Third, the focal analysis shows that relevant assessment evidence may emerge beyond programmed dashboard incidents. Sensitive verbal interactions, help-seeking, discrimination, vulnerability, resistance and other evolving classroom demands may offer substantial material for pedagogical interpretation, including in sessions without programmed challenging events.

These contributions help narrow the gap between the capacity of immersive simulations to generate rich records and the availability of procedures for transforming them into interpretable assessment evidence. The educational value of the infrastructure depends

on preserving the temporal, interactional and contextual relationships through which professional action becomes understandable. Learning analytics can therefore contribute to formative assessment when technical traces support pedagogical inquiry rather than function as direct indicators of competence. From this perspective, responsible and sustainable implementation requires assessment infrastructures that preserve evidence traceability and interpretability to support feedback, reflection and improvement over time.

The findings also identify priorities for the continued development of Mission Control. Better integration of dashboard events, continuous signals, exported data and audiovisual records, together with improved temporal synchronisation, export consistency and links between recordings and session identifiers, would facilitate multimodal sequence reconstruction and reduce the manual effort required to retrieve relevant evidence. On this basis, future AI-supported functions could assist teacher educators through automatic transcription, temporal alignment, sequence segmentation, identification of pedagogically relevant episodes and possible debriefing prompts. These functions should be designed around educators' interpretative and decision-making needs, supporting evidence retrieval and pedagogical inquiry rather than automated judgements of participant competence. Their development should also include safeguards concerning privacy, consent, transparency and the limits of inference from voice and movement signals.

For practitioners, the findings suggest three recommendations. First, teacher educators should use dashboard records as entry points for pedagogical inquiry rather than as direct indicators of competence. Second, debriefing should focus on a small number of sequence-based episodes with high feedback potential, particularly those involving identifiable pedagogical, relational or ethical tensions, participant responses and opportunities for alternative action planning. Third, continuous voice and movement signals should be interpreted only in relation to the broader interactional sequence and, where possible, triangulated with video, transcripts or other contextual evidence. These recommendations can help practitioners use VR analytics to support focused feedback and reflection while avoiding overinterpretation of isolated technical traces.

Future research should apply the framework to a larger and more varied set of sessions, examine its consistency and usefulness with teacher educators, and incorporate sequence-based evidence into debriefing and feedback practices. This next stage is needed to determine how educators and participants interpret and use the evidence, and whether such use contributes to reflection, subsequent performance and classroom-management competence.

Overall, the study provides a preliminary and potentially transferable procedure for transforming heterogeneous multimodal traces into contextually reconstructed evidence of professional action. Its contribution lies in clarifying the analytical and pedagogical conditions under which immersive simulation records can support teacher interpretation, evidence-based feedback and the responsible development of learning analytics in initial teacher education.

Author Contributions: Conceptualisation, I.M.A., S.R. and B.M.; Methodology, I.M.A. and A.M.; Software, S.R. and A.R.-H.; Validation, A.M., S.R., B.M. and A.R.-H.; Formal Analysis, I.M.A. and A.M.; Data Curation, A.M., I.M.A. and S.R.; Writing—Original Draft Preparation, I.M.A.; Writing—Review and Editing, all authors; Project Administration, B.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science, Innovation and Universities (Spain), grant PID2024-156187OB-I00.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Universitat Autònoma de Barcelona (protocol code CERec374, date of approval: 21 January 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Sample anonymised data are openly available in Zenodo at <https://doi.org/10.5281/zenodo.21282825>. Video recordings cannot be shared due to ethical and privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
API	Application programming interface
CSV	Comma-separated values
DB	Database
HTTPS	Hypertext Transfer Protocol Secure
JWT	JSON Web Token
RAID-5	Redundant Array of Independent Disks, level 5
RQ	Research question
SSH	Secure shell
SSL/TLS	Secure Sockets Layer/Transport Layer Security
TALK	Facilitator-mediated verbal event introduced through Mission Control
VPN	Virtual private network
VR	Virtual reality

Appendix A. Focal Cases Selected for Sequence-Based Multimodal Analysis

Table A1 presents the 17 focal cases selected for sequence-based multimodal analysis, summarising their main sequence type, dominant pedagogical or relational tension and feedback potential. Table A2 reports the non-exclusive typologies assigned to each case. Dashboard-launched programmed challenging events are specified in the main sequence type only for the cases in which they occurred: C16 included SEPARATE and INSULT, and C17 included STAND_UP. Original Mission Control session identifiers and, where applicable, participant names were replaced with anonymised case IDs to protect participant confidentiality.

‘Main sequence type’ refers to the dominant analytical configuration of the case, while ‘non-exclusive typologies’ indicates additional analytical categories that may coexist within the same sequence. Cases without dashboard-launched programmed challenging events may still contain sensitive interactions, emergent tensions or non-programmed critical moments. Feedback potential refers to the formative value of the sequence for debriefing and not to participant performance; it was coded as High when the sequence provided rich, interpretable evidence for discussing pedagogical, relational or ethical decision-making, and as Medium when the evidence was relevant but more limited, brief or less fully contextualised. Case C03 is further developed in Appendix B as a worked example of sequence-based multimodal coding.

Table A1. Focal cases selected for sequence-based multimodal analysis.

Case	Dominant Analytical Contribution in Table 4	Main Sequence Type	Dominant Pedagogical/Relational Tension	Feedback Potential
C01	Resistance, disengagement or role-challenge sequences	Contained verbal resistance/initial role challenge	Challenge to the teacher role	Medium
C02	Sensitive sequences without programmed events	Sensitive TALK-only sequence without dashboard-launched event	Discrimination/identity	High
C03	Sensitive sequences without programmed events	Sensitive TALK-only sequence with pedagogical elaboration	Vulnerability + body-related discrimination + normalisation of aggression	High
C04	Sensitive sequences without programmed events	Sensitive TALK-only sequence without dashboard-launched event	Discrimination/identity	High
C05	Sensitive sequences without programmed events	Sensitive TALK-only sequence without dashboard-launched event	Discrimination/identity	High
C06	Sensitive sequences without programmed events	Sensitive TALK-only sequence without dashboard-launched event	Discrimination/identity	High
C07	Sensitive sequences without programmed events	Sensitive TALK-only sequence without dashboard-launched event	Discrimination /identity	High
C08	Sensitive sequences without programmed events	Sensitive TALK-only sequence without dashboard-launched event	Discrimination/identity	High
C09	Bodily discomfort or exit-from-class sequences	Contained verbal resistance/initial role challenge involving bodily discomfort/exit-from-class demand	Physical discomfort/leaving the classroom	Medium
C10	Resistance, disengagement or role-challenge sequences	Sensitive TALK-only sequence involving help-seeking and disengagement from the classroom task	Help-seeking/disengagement	High
C11	Sensitive sequences without programmed events	Sensitive TALK-only sequence without dashboard-launched event	Discrimination/identity	High
C12	Resistance, disengagement or role-challenge sequences	Contained verbal resistance/initial role challenge	Challenge to the teacher role	Medium

Table A1. *Cont.*

Case	Dominant Analytical Contribution in Table 4	Main Sequence Type	Dominant Pedagogical/Relational Tension	Feedback Potential
C13	Sensitive sequences without programmed events	Sensitive TALK-only sequence without dashboard-launched event	Discrimination/identity	High
C14	Bodily discomfort or exit-from-class sequences	Sensitive TALK-only sequence without dashboard-launched event involving bodily discomfort/exit-from-class demand	Physical discomfort/help-seeking/leaving the classroom	High
C15	Resistance, disengagement or role-challenge sequences	Emergent non-programmed critical sequence with disengagement and relational conflict	Relational conflict/classroom coexistence	High
C16	Programmed-event sessions	Complex mixed sequence with multiple programmed events: SEPARATE and INSULT	Justification or normalisation of aggression	High
C17	Programmed-event sessions	Abrupt disruption with programmed challenging event: STAND_UP	Neutral interaction/classroom disruption	High

Note: ‘Dominant analytical contribution in Table 4’ identifies the category used for the aggregate counts reported in Table 4. ‘Main sequence type’ and ‘Dominant pedagogical/relational tension’ provide a finer-grained description of each case and may include overlapping demands. Each case was assigned to one dominant Table 4 category for summary purposes, while Table A2 reports additional non-exclusive analytical typologies that may coexist within the same sequence.

Table A2. Non-exclusive typologies assigned to the focal cases.

Case	Additional Non-Exclusive Analytical Typologies
C01	Contained verbal resistance
C02	Explicit help-seeking; sensitive relational conflict; sensitive sequence without programmed challenging event; participant hesitation/nervous laughter
C03	Explicit help-seeking; sensitive relational conflict; sensitive sequence without programmed challenging event; critical TALK; participant response to discrimination
C04	Explicit help-seeking; sensitive relational conflict; sensitive sequence without programmed challenging event
C05	Sensitive relational conflict; sensitive sequence without programmed challenging event
C06	Explicit help-seeking; sensitive relational conflict; sensitive sequence without programmed challenging event
C07	Explicit help-seeking; sensitive relational conflict; sensitive sequence without programmed challenging event
C08	Sensitive relational conflict; sensitive sequence without programmed challenging event; participant hesitation/nervous laughter

Table A2. *Cont.*

Case	Additional Non-Exclusive Analytical Typologies
C09	Explicit help-seeking; bodily discomfort/exit-from-class sequence; participant hesitation/nervous laughter
C10	Explicit help-seeking; task resistance/disengagement
C11	Sensitive relational conflict; sensitive sequence without programmed challenging event
C12	Contained verbal resistance; participant hesitation/nervous laughter
C13	Explicit help-seeking; sensitive relational conflict; sensitive sequence without programmed challenging event
C14	Explicit help-seeking; bodily discomfort/exit-from-class sequence; physical discomfort-related request
C15	Sensitive relational conflict; sensitive sequence without programmed challenging event; emergent non-programmed critical event; task resistance/disengagement; participant hesitation/nervous laughter
C16	Sensitive relational conflict; abrupt disruption; progressive escalation; programmed event with subsequent TALK; participant hesitation/nervous laughter
C17	Neutral/opening TALK events; isolated programmed challenging event; abrupt disruption

Note: Additional typologies are non-exclusive and describe analytical features that may coexist within the same focal case, such as help-seeking, sensitive relational conflict, disengagement, bodily discomfort, programmed events or participant hesitation. These typologies complement the dominant analytical contribution reported in Table A1 and do not replace the category used for the aggregate counts in Table 4.

Appendix B. Worked Example of Sequence-Based Multimodal Coding

Appendix B.1. Purpose of the Worked Example

This appendix presents a worked example of the analytical pathway described in Figure 3, showing how Mission Control records can be transformed into pedagogically interpretable assessment evidence through sequence-based multimodal coding. The selected case illustrates a sensitive interactional sequence without programmed challenging events and was chosen because it shows that the absence of a dashboard-launched incident does not imply the absence of formative evidence. For ethical reasons, the video recording is not shared. The example includes only anonymised Mission Control records, selected anonymised transcript excerpts, the coding matrix and the pedagogical interpretation derived from the analysis.

Appendix B.2. Case Overview

Case C03 is a sensitive TALK-only sequence in which no programmed challenging event was launched from the dashboard. It represents a frequent configuration in the focal sample: a sensitive verbal sequence whose formative value emerges from avatar TALK, participant responses and exploratory dashboard signals rather than from a programmed incident. The interaction centred on vulnerability, body-related discrimination and the normalisation of aggression. The participant recognised the discriminatory nature of the situation and attempted to transform it into a classroom discussion, making the sequence useful for analysing pedagogical elaboration, ethical sensitivity and opportunities for debriefing. Table A3 summarises the session metadata and available records. Voice and movement signals were inspected as contextual traces of vocal and bodily activity during the critical TALK sequence and near the closing phase. These indicators were interpreted

in relation to the transcript and dashboard record, not as direct measures of emotion, stress, competence or performance.

Table A3. Session metadata and available records.

Field	Description
Case identifier	Case C03
Duration	Approximately 3 min
Session type	Sensitive sequence without programmed challenging events
Programmed challenging events	None
Main data sources	Mission Control event log, dashboard timeline, CSV voice and movement signals, transcript excerpt and coding matrix
Main analytical focus	Vulnerability, body-related discrimination, normalisation of aggression, pedagogical elaboration and session closure
Exploratory signal focus	Inspection of vocal activity, head and hand movement traces, and closing pattern in relation to critical TALK moments
Main type of tension	Bullying/vulnerability
Secondary tensions	Discrimination/identity; relational conflict; mild challenge to the teacher role
Feedback potential	High

Note. This worked example uses one anonymised focal case from Appendix A; original Mission Control session identifiers are not reported. Voice and movement signals were treated as exploratory contextual indicators, not as direct indicators of emotion, stress, competence or performance.

Appendix B.3. Exploratory Interpretation of Dashboard Signals

The dashboard signals were inspected to contextualise the TALK sequence within the temporal organisation of the session. Table A4 summarises how the vocal and movement traces were used as exploratory evidence for feedback interpretation, while Figure A1 provides the corresponding visual dashboard evidence.

Table A4. Exploratory interpretation of dashboard signals in Case C03.

Indicator	Observation in Case C03	Use for Feedback Interpretation
TALK sequence	The critical sequence developed between approximately 19:25:48 and 19:26:52, moving from help-seeking and body-related discrimination to the normalisation of aggression.	Identifies the main pedagogical moment for debriefing and supports discussion of how the participant responded to vulnerability and discriminatory language.
Vocal signal	The volume curve showed sustained activity and marked variation during the critical TALK sequence.	Helps locate communicative activity and possible moments of tension for review, but does not indicate emotion, stress or competence by itself.

Table A4. *Cont.*

Indicator	Observation in Case C03	Use for Feedback Interpretation
Head movement	Head velocity remained mostly low during the central critical sequence, with isolated peaks at the beginning and near the end of the session.	Provides contextual information on bodily activity, but requires triangulation with video and transcript before pedagogical interpretation.
Hand movement	Left- and right-hand velocity showed intermittent activity, with stronger peaks toward the final part of the record.	May indicate embodied activity or technical manipulation, especially near closure, and should not be interpreted independently.
Closing pattern	A marked increase in movement appeared near the end of the record, after the final sensitive TALK and participant response.	Supports identifying the closing phase as a relevant moment for review, while distinguishing pedagogical closure from possible technical artefacts.
Multimodal interpretation	The voice and movement signals were interpreted in relation to the TALK sequence, transcript and dashboard timeline.	The feedback value emerges from triangulation rather than from any isolated metric.

Note: Voice and movement indicators were treated as exploratory contextual traces for locating communicative activity, tension and closure in relation to the TALK sequence, not as direct indicators of emotion, stress, competence or performance. Movement peaks near the beginning or end of the record may reflect technical actions, such as adjusting or removing equipment.

Table A5. Relevant Mission Control TALK events.

Order	Time	TALK Text	Preliminary Coding
TALK 1	19:25:09	‘Are you new?’	Mild role challenge/opening interaction
TALK 2	19:25:16	‘Where is our teacher?’	Mild role challenge
TALK 3	19:25:48	‘They always shout at me.’	Bullying/vulnerability; help-seeking
TALK 4	19:26:01	‘They called me fat.’	Discrimination/identity; bullying/vulnerability; help-seeking
TALK 5	19:26:36	‘But it is true.’	Normalisation of aggression
TALK 6	19:26:52	‘Then I don’t see anything wrong with it.’	Normalisation/minimisation of harm

Note. TALK texts were translated and anonymised for publication. The original dashboard record was written in Spanish classroom interactional language. Timestamps are approximate and refer to the Mission Control session timeline.

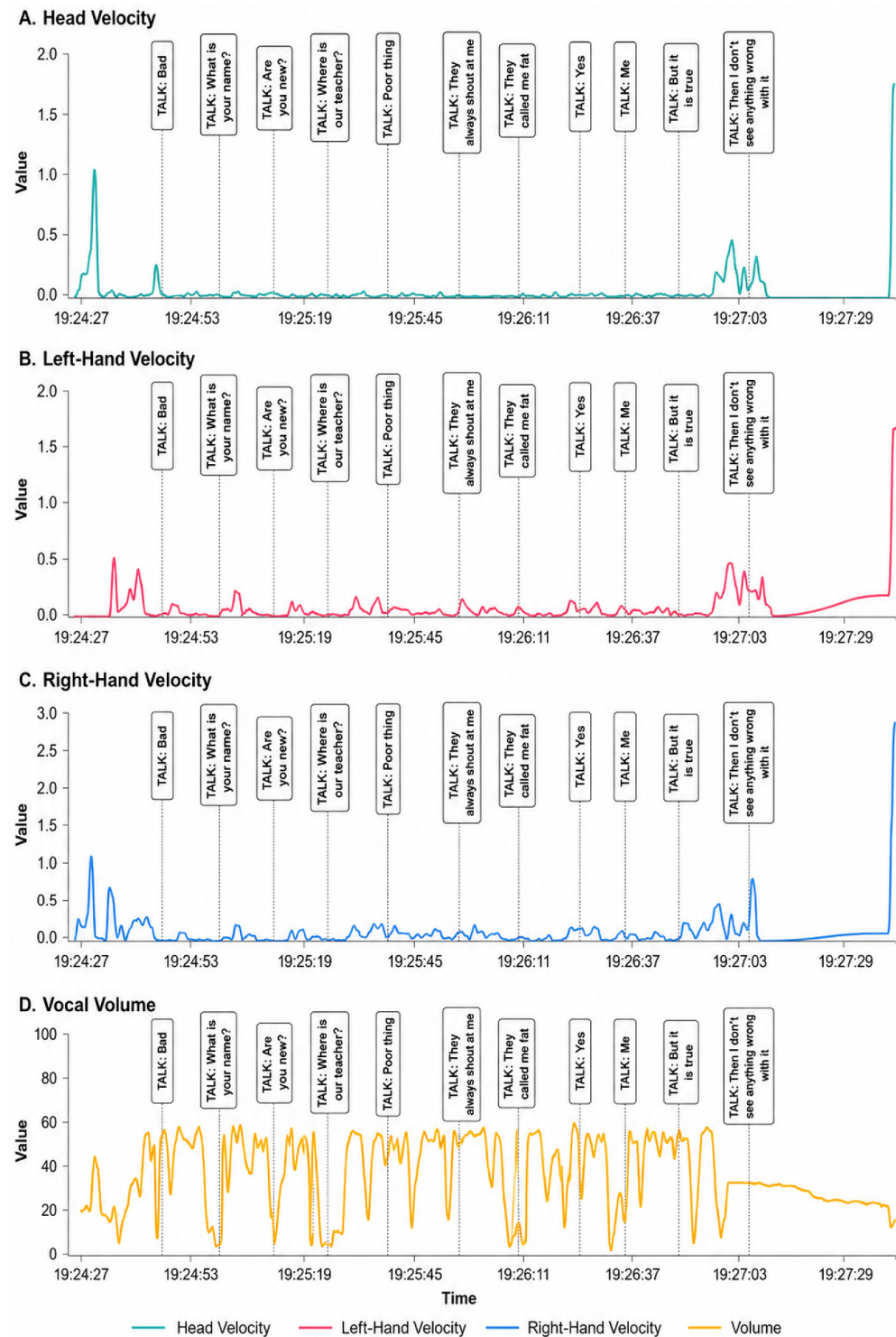


Figure A1. Exploratory multimodal dashboard evidence for Case C03. Note. The figure shows four graphs exported from the Mission Control dashboard for Case C03: head velocity, left-hand velocity, right-hand velocity and vocal volume. The x-axis displays clock time (hh:mm:ss), beginning at 19:24:27, the start of the session record. TALK events are aligned with the timeline to locate the critical TALK sequence and inspect its relationship with vocal activity, movement traces and session closure. The graph is used as a contextual visual aid, not as quantitative evidence of participant performance, emotion, stress or competence. To avoid relying on compressed in-figure annotations, the complete verbatim transcriptions of the recorded TALK events are provided in Table A5.

Appendix B.4. Mission Control Record

The Mission Control record contained several TALK events introduced during the simulated classroom interaction. These events provided the initial dashboard trace for reconstructing the sequence, but did not capture the participant's full response or the complete classroom dialogue. Table A5 summarises the relevant TALK events. No dashboard-launched programmed challenging events, such as STAND_UP, SEPARATE or INSULT, occurred in this session; therefore, the intervention window was coded as not applicable rather than absent.

Appendix B.5. From TALK Events to Interactional Sequence

The transcript excerpt illustrates how dashboard TALK events were interpreted within a broader interactional sequence. While Mission Control records identified key verbal interactions, the transcript provided additional information about participant responses and the unfolding classroom discussion.

Table A6. From TALK event to interactional sequence.

Mission Control TALK	Participant Response	Sequence Interpretation	Debriefing Focus
'They always shout at me.'	'Okay, then we need to learn to use language in a way that respects all classmates. Right? Okay, can anyone give me an example of—'	A vulnerable statement introduces a relational demand. The participant acknowledges respectful language but redirects quickly to the lesson objective.	How could the teacher recognise the pupil's vulnerability more explicitly before moving back to the planned activity?
'They called me fat.'	'Okay! That is an example of using language for something that is not positive and that is discriminatory, right? Has this happened to anyone else?'	Body-related discrimination becomes explicit. The participant identifies the comment as discriminatory and opens the issue to the group.	How can discriminatory language be addressed without exposing or re-victimising the affected pupil?
'Yes.'/'Me.'/'Me.'	'Okay. . . Okay, okay. So, for example, is calling a classmate fat using language well, and is it discriminating?'	Several avatars align with the example, turning an individual report into a group-level discussion. The participant transforms the episode into a classroom question.	How can the teacher use peer responses to build a collective reflection while protecting the affected pupil?
'But it is true.'	'Okay, but fat is ultimately an adjective and. . . it depends on the intentionality of how we use it. It can be used to describe a person or to insult them and disrespect them, right?'	A truth claim justifies aggression. The participant distinguishes between description, intention, insult and disrespect.	How can the teacher challenge the normalisation of discriminatory language while clarifying the difference between description, intention and harm?

Table A6. *Cont.*

Mission Control TALK	Participant Response	Sequence Interpretation	Debriefing Focus
'Then I don't see anything wrong with it.'	'Okay, but how did they say it to you? How did they call you fat?'	The group resists the ethical framing of the episode. The participant asks for contextual clarification and shifts attention to how the expression was used.	How can the teacher move from intentionality to impact, harm and responsibility in an inclusive classroom conversation?

Note. Transcript excerpts were translated, anonymised and abridged. Only turns required to illustrate the coding procedure are shown.

Appendix B.6. Sequence-Based Coding of Case C03

Table A7. Coding matrix.

Variable	Coding
case_id	Case C03
session_type	Sensitive sequence without programmed challenging events
data_sources	Mission Control log; dashboard timeline; CSV voice and movement signals; transcript excerpt; coding matrix
programmed_challenging_event	No
event_types_present	TALK only
critical_TALK_present	Yes
sensitive_TALK_present	Yes
main_critical_TALK	'They always shout at me'; 'They called me fat'
main_tension	Bullying/vulnerability
secondary_tensions	Discrimination/identity; relational conflict; mild role challenge
participant_response	The participant recognises the situation as discriminatory, attempts to work on intentionality and harm, and uses the example didactically.
exploratory_signal_interpretation	Vocal and movement signals were inspected as contextual traces in relation to the critical TALK sequence and closing phase, without being interpreted as direct indicators of emotion, stress, competence or performance.
intervention_window	Not applicable
reason_intervention_window	No programmed challenging event occurred in the session.
sequence_pattern	TALK-only critical sequence with pedagogical elaboration
emergent_critical_event	Yes

Table A7. Cont.

Variable	Coding
type_of_emergent_event	Sensitive verbal interaction involving vulnerability and discrimination
feedback_potential	High
feedback_focus	Recognition of vulnerability; response to body-related discrimination; distinction between description, insult, intention and harm; use of sensitive examples in classroom discussion
evidence_reuse	Suitable for debriefing and reflective teacher education; suitable as anonymised worked example of the coding framework

Appendix B.7. Possible Debriefing Prompts

The following questions illustrate how sequence-based assessment evidence can be reused during post-simulation debriefing. Derived from the coded interactional sequence, they exemplify how Mission Control records and transcript-based analysis may support reflection on classroom management, inclusion and pedagogical decision-making. The prompts are not intended to be applied exhaustively, but selected according to the focus of the debriefing and the needs of the group.

1. What pedagogical demand emerges when the avatar says that others always shout at them?
2. How does the situation change when the avatar reports being called fat?
3. What is the difference between recognising the comment as discriminatory and addressing the harm experienced by the pupil?
4. How could the teacher respond when another pupil justifies the comment by saying, 'But it is true'?
5. What would be an appropriate balance between continuing the planned lesson and protecting the vulnerable pupil?
6. How could this sequence be used to discuss inclusive language with the whole group without exposing or re-victimising the affected pupil?

Appendix B.8. Key Methodological Message

This worked example illustrates that the educational value of Mission Control does not derive from isolated dashboard events. Instead, formative evidence emerges through the multimodal reconstruction of interactional sequences, where TALK, participant responses, exploratory signals and contextual interpretation become jointly meaningful for feedback and debriefing.

References

- Alvarez, I. M., Manero, B., Romero-Hernández, A., Cárdenas, M., & Masó, I. (2024). Virtual reality platform for teacher training on classroom climate management: Evaluating user acceptance. *Virtual Reality*, 28(2), 78. [\[CrossRef\]](#)
- Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 37, 100489. [\[CrossRef\]](#)
- Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. [\[CrossRef\]](#)
- Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., Bakken, S., Kaplan, C. P., Squiers, L., Fabrizio, C., & Fernandez, M. (2009). How we design feasibility studies. *American Journal of Preventive Medicine*, 36(5), 452–457. [\[CrossRef\]](#)

- Garzón Artacho, E., Martínez, T. S., Ortega Martín, J. L., Marín Marín, J. A., & Gómez García, G. (2020). Teacher training in lifelong learning—The importance of digital competence in the encouragement of teaching innovation. *Sustainability*, 12(7), 2852. [\[CrossRef\]](#)
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. [\[CrossRef\]](#)
- Gašević, D., Kovanović, V., & Joksimović, S. (2017). Piecing the learning analytics puzzle: A consolidated model of a field of research and practice. *Learning: Research and Practice*, 3(1), 63–78. [\[CrossRef\]](#)
- Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, 3, 275–285. [\[CrossRef\]](#)
- Han, X., Luo, H., Wang, Z., & Zhang, D. (2025). Using virtual reality for teacher education: A systematic review and meta-analysis of literature from 2014 to 2024. *Frontiers in Virtual Reality*, 6, 1620905. [\[CrossRef\]](#)
- Kaliisa, R., Jivet, I., & Prinsloo, P. (2023). A checklist to guide the planning, designing, implementation, and evaluation of learning analytics dashboards. *International Journal of Educational Technology in Higher Education*, 20, 28. [\[CrossRef\]](#)
- Li, L., Hu, Y., Yang, X., Wu, M., Tao, P., Chen, M., & Yang, C. (2024). Enhancing pre-service teachers' classroom management competency in a large class context: The role of fully immersive virtual reality. *Humanities and Social Sciences Communications*, 11, 1050. [\[CrossRef\]](#)
- Liu, K., & Deris, F. D. (2025). AI feedback literacy in higher education: Understanding, measuring, and predicting student feedback uptake. *Assessment & Evaluation in Higher Education*, 1–15. [\[CrossRef\]](#)
- Malmqvist, J., Hellberg, K., Möllås, G., Rose, R., & Shevlin, M. (2019). Conducting the pilot study: A neglected part of the research process? Methodological findings supporting the importance of piloting in qualitative research studies. *International Journal of Qualitative Methods*, 18, 1609406919878341. [\[CrossRef\]](#)
- Markauskaite, L., Carvalho, L., & Fawns, T. (2023). The role of teachers in a sustainable university: From digital competencies to postdigital capabilities. *Educational Technology Research and Development*, 71, 181–198. [\[CrossRef\]](#)
- Morgan, D. L. (2023). Exploring the use of artificial intelligence for qualitative data analysis: The case of ChatGPT. *International Journal of Qualitative Methods*, 22, 16094069231211248. [\[CrossRef\]](#)
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. [\[CrossRef\]](#)
- Qazi, A. G., & Pachler, N. (2025). Conceptualising a data analytics framework to support targeted teacher professional development. *Professional Development in Education*, 51(3), 495–518. [\[CrossRef\]](#)
- Ruipérez-Valiente, J. A. (2020). El proceso de implementación de analíticas de aprendizaje [The learning analytics implementation process]. *RIED Revista Iberoamericana de Educación a Distancia*, 23(2), 85–101. [\[CrossRef\]](#)
- Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal learning analytics tell us? *British Journal of Educational Technology*, 51(5), 1450–1484. [\[CrossRef\]](#)
- Street, K. E. S., Ertesvåg, S. K., Gamlem, S. M., Rebni, L. M., Sølvi, R. M., & Vaaland, G. S. (2025). Classroom management for first-year pre-service teachers: What is difficult and why? *Teaching and Teacher Education*, 168, 105235. [\[CrossRef\]](#)
- Tao, L., Cukurova, M., & Song, Y. (2025). Learning analytics in immersive virtual learning environments: A systematic literature review. *Smart Learning Environments*, 12, 43. [\[CrossRef\]](#)
- Voordijk, H., & Vahdatikhaki, F. (2022). Virtual reality learning environments and technological mediation in construction practice. *European Journal of Engineering Education*, 47(2), 259–273. [\[CrossRef\]](#)
- Winstone, N. E., Gravett, K., & Elkington, S. (2026). Black box assessment: Rethinking integrity and learning for a time of generative AI. *Assessment & Evaluation in Higher Education*, 1–18. [\[CrossRef\]](#)
- Zhang, H., Wu, C., Xie, J., Lyu, Y., Cai, J., & Carroll, J. M. (2025). Harnessing the power of AI in qualitative research: Exploring, using and redesigning ChatGPT. *Computers in Human Behavior: Artificial Humans*, 4, 100144. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.