

Assessing Procedural Knowledge and Learner Experience in VR Safety Training Using Multimodal Learning Analytics

Timothé Bonhoure¹[0009–0003–3300–9933], Mélina Verger¹[0000–0002–5839–882X],
and Élise Lavoué^{1,2}[0000–0002–2659–6231]

¹ INSA Lyon, CNRS, Université Claude Bernard Lyon 1, LIRIS, UMR5205, 69621
Villeurbanne, France

`{timothe.bonhoure,melina.verger}@insa-lyon.fr`

² University Jean Moulin Lyon 3, iaelyon School of Management, Lyon, France
`elise.lavoue@univ-lyon3.fr`

Abstract. Virtual reality (VR) environments offer strong potential for training in high-risk domains such as medical and chemical laboratory settings. Their effect on learning is usually measured using self-reported questionnaires. However, these measures do not adequately capture how learning evolves over time in immersive environments. By combining physiological signals, eye-tracking data, and behavioral traces, multimodal learning analytics (MMLA) provide a richer and more integrated understanding of learning processes. Beyond learning assessment, MMLA is also highly relevant for risk prevention and incident management, as it may capture key indicators such as learners’ stress and attention. In this paper, we present an experimental framework that synchronizes multimodal data collection in a virtual chemistry laboratory training environment. We investigate how the analysis of these data can inform on both the development of procedural knowledge for risk prevention and incident management, and the overall learner experience (flow, cognitive load, presence). This work ultimately aims to characterize the dynamics of learning indicators to support the adaptive design of training scenarios for safer practice.

Keywords: multimodal learning analytics (MMLA) · virtual reality · procedural knowledge · learner experience · safety training

1 Introduction

Immersive virtual reality (VR) is increasingly used to support the acquisition of procedural skills in high-risk settings, including medical and chemical laboratories. In these domains, learners have to not only acquire technical gestures but also develop rapid and appropriate responses to critical incidents. Such training, therefore, requires rigorous assessment methods to ensure that learners achieve a sufficient level of competence. Yet, in immersive settings, learning outcomes are often assessed using post-hoc self-reported questionnaires. While these measures provide useful global perceptions, they capture only a partial view of how

learning unfolds over time. Behavioral traces can complement these measures by documenting actions and performance, but they still provide limited access to the cognitive and emotional processes underlying decision-making during task execution.

Multimodal Learning Analytics (MMLA) address this limitation by combining complementary sources of evidence, including self-reported questionnaires, interaction traces, physiological signals, and eye-tracking indicators. This integrated perspective is particularly relevant in VR, where learning is dynamic, situated, and tightly linked to attention allocation, stress regulation, and procedural control. However, implementing MMLA in immersive environments remains methodologically demanding. It requires intrusive instrumentation, precise temporal synchronization of heterogeneous streams, and robust preprocessing pipelines to mitigate motion-related artefacts, while also accounting for practical VR constraints such as equipment cost and setup complexity. Although interest in MMLA for education is increasing, its application to procedural knowledge acquisition in VR safety training remains limited and insufficiently consolidated [3].

To address this gap, this paper investigates the following main research question: *How can multimodal learning analytics inform our understanding of procedural knowledge acquisition in virtual reality safety training?* Our main contribution is a robust experimental framework that articulates pedagogical scenario design, synchronized multimodal data collection, and analysis strategies to jointly study procedural knowledge acquisition and learner experience in VR [7]. The paper is organized as follows. Section 2 reviews related work on physiological and eye-tracking indicators, as well as prior MMLA approaches in and beyond VR. Section 3 presents the pedagogical scenario. Section 4 details the experimental protocol and data collection procedure. Section 5 describes the multimodal analysis pipeline. Finally, Section 6 discusses contributions, limitations, and future perspectives for adaptive and explainable VR training systems.

2 Related Work

Physiological indicators are increasingly used to better characterize learning processes, each modality having specific strengths and limits. ECG-derived metrics can inform on stress regulation, cognitive load, and changes in procedural control; however, they are sensitive to movement artefacts and therefore require robust preprocessing and quality control [6]. EDA provides complementary information about arousal, engagement, and workload dynamics, yet its interpretation strongly depends on methodological rigor (sensor placement, filtering, decomposition, and normalization) [10]. Eye-tracking adds a fine-grained view of visual attention and exploration strategies through fixations, saccades, and pupillometric measures, while remaining highly dependent on calibration quality and vulnerable to tracking errors in VR environments [8].

Beyond VR, MMLA literature consistently shows that no single data stream is sufficient to comprehensively explain learning dynamics. Interaction traces capture task progression and strategy use, physiological measures approximate

effort and arousal fluctuations, and eye-tracking reveals how attention is allocated over time [4]. Recent syntheses further indicate that multimodal fusion often improves explanatory and predictive power compared with unimodal models, especially when analyses preserve temporal structure rather than relying solely on aggregated post hoc indicators [2]. These findings provide a strong methodological basis for immersive learning research.

Recent VR studies already show the feasibility of multimodal designs: for example, Abdurrahman et al. [1] used a VR-based driving system that combines eye-gaze, pupil dilation, heart rate, and performance data, and reported that multimodal fusion outperformed unimodal approaches for cognitive-load-related classification. Similarly, Sobocinski et al. [9] combined think-aloud audio, bird’s-eye-view video recordings, and physiological signals to analyse temporal relations between metacognitive monitoring, cognitive load, and motion in immersive VR learning. These contributions confirm that MMLA in immersive environments is progressing rapidly. However, most works still rely on parallel or modality-specific analyses, with limited fusion of synchronized signals, or limit their analysis to physiological indicators. In line with the broader MMLA literature [4], our objective is therefore not to propose new modalities, but to use them within a more explicitly integrated analytical framework for VR safety training.

3 Pedagogical Scenario in a VR Training Environment

We designed a pedagogical scenario as a central contribution of this study because it explicitly links instructional design choices to the experimental protocol and to the multimodal analyses that follow. Rather than treating the learning activity and data collection as separate components, we built an integrated framework in which learning situations are intentionally structured to elicit observable indicators of procedural knowledge acquisition, risk-based decision-making, and learner experience.

Concretely, we developed four simulations within the same VR training environment, each one following the same pedagogical format: a primary laboratory task (dilution or titration) is coupled with a forced critical incident (chemical spill or fire). This combination was selected to capture two complementary dimensions of procedural learning. The primary task reflects routine procedural execution (sequence compliance, precision, and timing), whereas the incident probes adaptive procedural control in a safety-critical context (situation assessment, prioritization, and appropriateness of corrective actions). By integrating both dimensions in each session, the scenario makes it possible to study not only whether learners can apply a procedure, but also how they regulate their behavior when the expected workflow is disrupted.

The sequence of activities is progressive and theory-driven. After an initial tutorial that standardizes basic interaction with the VR environment, participants complete four simulations. The first two simulations include comparable incident structures to capture short-term learning effects through repetition and consolidation. The next two simulations introduce a different incident type to

assess transfer, that is, the learner’s capacity to mobilize acquired procedural principles in a new situation. This progression supports a fine-grained analysis of learning dynamics across sessions, from initial familiarization to adaptation and transfer.

4 Methodology

The data collection protocol is structured into three main phases centred on the VR training: a pre-experimental phase, data collection during the immersive phase, and a post-experimental phase.

Before the VR training begins, participants undergo a standardized pre-experimental phase designed to ensure safety, data quality, and help us better understand baseline conditions to analyse the rest of the data. After signing the informed consent form, an initial questionnaire is administered to collect demographic information, prior experience with VR/video games, motivation in chemistry, and basic knowledge of chemistry and risk prevention. Participants are then fitted with physiological sensors, positioned comfortably, and a 5-minute rest period (baseline) is observed to stabilize signals before launching the tutorial and simulation tasks. This provides baseline reference conditions.

During the VR training, multimodal data are collected synchronously across the tutorial and the four simulations. All data streams are temporally aligned using the PLUME toolbox [5], which supports synchronized recording of behavioral and physiological data in 6DoF XR environments, including physiological signals (ECG, EDA) through LSL, eye-tracking data from OpenXR, and behavioral traces from Unity (interaction events, performance metrics). Between each simulation, participants answer two brief questions, one on their perceived level of competence and one on their sense of presence, in order to capture temporal fluctuations in learner experience and to enable a more detailed analysis of learning dynamics between simulations.

Finally, after completing the VR training simulations, participants completed a series of self-reported questionnaires to assess both their acquired knowledge of chemical hazards and their subjective experience, including cognitive load (MCLSVE), state of flow (S-FSS2), and sense of presence (IPQ). A brief interview with a few open-ended questions is then conducted to gather learners’ feedback on their experience and the challenges they encountered.

This protocol was approved by the Ethics Committee of Université de Lyon (CER-UdL).

5 Data Analysis

Drawing on the synchronised multimodal corpus collected during the VR sessions, we aim to analyse signals together to gain a more comprehensive understanding of procedural knowledge acquisition than classical unimodal analyses provide. To address the main research question, this section is organised around three complementary objectives: identifying the most informative performance

and learner experience indicators over time, quantifying the added value of multimodal fusion for prediction and characterising recurrent learner profiles during immersive training. Before performing our analyses, we perform a standardised pre-processing step to clean and structure the signals for feature extraction. We then apply a two-level temporal segmentation grounded in the pedagogical scenario: macro-level segmentation aligned with pedagogical simulations to capture longitudinal progression across simulations, and micro-level segmentation centred on critical events to capture short-term regulation processes at key moments, such as stress increases around unexpected incidents and their potential impact on procedural decisions and subsequent learning.

To address our objectives, we will first identify the unimodal and multimodal indicators that best explain procedural performance and learner experience over time. To this end, we will primarily use mixed-effects models to relate modality-specific indicators to procedural scores, declarative knowledge gains and self-reported measures. Additionally, we will explore more complex models, including neural networks, to assess whether nonlinear relationships improve explanatory power. The extracted indicators will cover all available modalities. For EDA, these will include tonic/phasic decomposition, SCR rate and amplitude, used to characterise arousal peaks and recovery phases when learners face unexpected events. For ECG, they will include HR and HRV indices, used to estimate variations in autonomic regulation and cognitive load during routine execution versus incident management. For eye tracking, they will include fixation duration, saccade metrics, pupillometry and AOI dwell time, used to capture attentional allocation, visual search efficiency and potential attentional tunnelling under pressure. For behavioural traces, they will include procedural compliance scores and incident response times, which provide direct evidence of procedural quality, timing and adaptation.

Secondly, we will compare and evaluate multimodal fusion methods to identify the most effective strategies. Early, intermediate and late fusion approaches will be contrasted and then benchmarked against unimodal reference models to determine the extent to which fusion improves the prediction of learning outcomes. We will also analyse the impact of temporal segmentation on model performance, including bottom-up approaches to automatically identifying informative segments in the data.

Finally, we will conduct a learner-profile analysis based on multimodal data. An unsupervised clustering step (GMM or hierarchical clustering) will be applied to participant-level feature vectors, integrating average indicator levels and temporal dynamics across simulations. We will then test whether these profiles are associated with differences in learning performance or learner experience while controlling for prior VR experience and initial chemistry knowledge. In parallel, we will conduct a sequence analysis to identify recurrent multimodal indicator patterns that may characterise learner profiles or specific learning moments.

6 Conclusion and Future Work

In this paper, we proposed an MMLA experimental framework for studying the acquisition of procedural knowledge in VR within the context of chemical safety. This study focuses on integrating an educational scenario, an experimental protocol, and the synchronized collection of heterogeneous data to jointly analyze procedural improvement, gains in knowledge, and the learning experience. This work thus bridges the gap identified between approaches relying solely on questionnaires or traces and multimodal VR studies, which remain largely underused.

This framework paves the way for adaptive learning environments that leverage multimodal indicators in near real time to detect difficulties, assess engagement or cognitive load, and adjust feedback or task difficulty. In this context, it seems appropriate to add a dimension of explainability: XAI approaches would enable a better understanding of the role of each modality and facilitate more precise identification of learner profiles or areas of insufficient understanding.

References

1. Abdurrahman, U.A., Yeh, S.C., Wong, Y., Wei, L.: Effects of neuro-cognitive load on learning transfer using a virtual reality-based driving system **5**(4), 54 (2021). <https://doi.org/10.3390/bdcc5040054>
2. Cukurova, M., Giannakos, M., Martinez-Maldonado, R.: The promise and challenges of multimodal learning analytics **51**(5), 1441–1449 (2020). <https://doi.org/10.1111/bjet.13015>
3. Di Mitri, D., Limbu, B., Schneider, J., İren, D., Giannakos, M., Klemke, R.: Multimodal and immersive systems for skills development and education **55** (2024). <https://doi.org/10.1111/bjet.13483>
4. Di Mitri, D., Schneider, J., Specht, M., Drachsler, H.: From signals to knowledge: A conceptual model for multimodal learning analytics **34**(4), 338–349 (2018). <https://doi.org/10.1111/jcal.12288>
5. Javerliat, C., Villenave, S., Raimbaud, P., Lavoué, G.: PLUME: Record, replay, analyze and share user behavior in 6dof XR experiences pp. 1–11 (2024). <https://doi.org/10.1109/TVCG.2024.3372107>
6. Ramasamy, S., Balan, A.: Wearable sensors for ECG measurement: a review **38**(4), 412–419 (2018). <https://doi.org/10.1108/SR-06-2017-0110>
7. Santamaría-Bonfil, G., Ibáñez, M.B., Pérez-Ramírez, M., Arroyo-Figueroa, G., Martínez-Álvarez, F.: Learning analytics for student modeling in virtual reality training systems: Lineworkers case **151**, 103871 (2020). <https://doi.org/10.1016/j.compedu.2020.103871>
8. Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N.S., Tripoliti, E., Marias, K., Fotiadis, D.I., Tsiknakis, M.: Review of eye tracking metrics involved in emotional and cognitive processes **16**, 260–277 (2023). <https://doi.org/10.1109/RBME.2021.3066072>
9. Sobocinski, M., Dever, D., Wiedbusch, M., Mubarak, F., Azevedo, R., Järvelä, S.: Capturing self-regulated learning processes in virtual reality: Causal sequencing of multimodal data **55**(4), 1486–1506 (2024). <https://doi.org/10.1111/bjet.13393>
10. Viana-Matesanz, M., Sánchez-Ávila, C.: Adaptive normalization and feature extraction for electrodermal activity analysis **12**(2), 202 (2024). <https://doi.org/10.3390/math12020202>