Electronic, didactic and innovative platform for learning based on multimedia assets

e-DIPLOMA

Funded by the European Union

D.5.4 Specifications Report

Version No. V1.2

19.09.2025

Disclaimer:

"Funded by the European Union. However, the views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the European Research Executive Agency (REA)can be held responsible for them."

HISTORY OF CHANGES			
Version*	**Publication date**	**Beneficiaries**	**Changes**
V1.0	30.07.2025	UPV	▪ Initial version of Deliverable Owner
V1.1	01.08.2025	BRA, TU Delft	▪ Version including suggestions of WP Contributors
V1.2	15.09.2025	UPV	▪ Pre-final version reviewed by Internal Reviewers
V1.2	19.09.2025	UPV	▪ Final version approved by Project Coordinator

(*) According to the section “Review and Submission of Deliverables” of the Project Handbook

1. Technical References

Project Number	101061424
Project Acronym	e-DIPLOMA
Project Title	Electronic, Didactic, and Innovative Platform for Learning Based on Multimedia Assets
Granting Authority	European Research Executive Agency (REA)
Call	HORIZON-CL2-2021-TRANSFORMATIONS-01
Topic	HORIZON-CL2-2021-TRANSFORMATIONS-01-05
Type of the Action	HORIZON Research and Innovation Actions
Duration	1 September 2022 – 31 October 2025 (38 months)
Entry into force of the Grant	1 September 2022
Project Coordinator	Inmaculada Remolar Quintana

Deliverable No.	D5.4: Specifications Report
Work Package	WP5: Assessment of the User Quality of Experience and Social and Educational Impact
Task	T5.2: Selection of appropriate metric for evaluation of learning competencies
Dissemination level*	PU- Public
Type of license:	CC-BY
Lead beneficiary	Universitat Politécnica de Valencia (UPV)
PIC of the Lead beneficiary	▪ 999421653

Contributing beneficiary/ies	▪ Universitat Jaume I (UJI) ▪ Technische Universiteit Delft (TU Delft) ▪ Budapesti Muszaki es Gazdasagtudomanyi Egyetem (BME) ▪ Universitat Politécnica de Valencia (UPV) ▪ Innogrowth European Association for Innovation and Growth (INN) ▪ Centre for Social Innovation LTD (CSI) ▪ Aris Formazione e Ricerca Societa Cooperativa (ARIS FR) ▪ Brainstorm Multimedia SL (BRAINSTORM)
PIC of the Contributing beneficiary/ies	▪ UJI: 999882985 ▪ TU Delft: 999977366 ▪ BME: 999904228 ▪ UPV: 999864846 ▪ INN: 900529668 ▪ CSI: 913552403 ▪ ARIS FR: 911643734 ▪ BRAINSTORM: 999441732
Due date of deliverable	31 July 2025
Actual submission date	19 September .2025

Table of contents

Executive Summary

This deliverable presents the technical and methodological framework developed in e-DIPLOMA to support the collection, synchronization, and interpretation of multimodal data in XR-based learning environments. The document focuses on the specification and validation of physiological, behavioural, and interaction metrics across the three educational prototypes developed in the project.

The main contribution of the deliverable is the definition of a multimodal logging system that integrates cognitive, affective, and psychomotor data streams from wearable sensors (e.g., EmotiBit, HP Omnicept) and system-level interactions. These metrics are aligned with the assessment model defined in WP5 and were validated in a large-scale study involving over 400 participants from five countries.

The results demonstrate that the integration of XR technologies enhances psychomotor learning (Prototype 3) and that physiological variables such as heart rate (HR) and cognitive load (CL) can be effectively used to model learning processes. Additionally, the document explores the relationships between self-reported psychological traits (e.g., self-efficacy, curiosity) and physiological profiles through multivariate analyses (e.g., PCA), providing early evidence for personalized adaptation strategies in immersive learning.

The findings will feed directly into the upcoming deliverables D5.5 (Impact Assessment) and D5.6 (Teacher Training), as well as guide future deployment and scaling-up activities in WP6.

1. Introduction

This deliverable provides a comprehensive overview of the technical and methodological infrastructure developed in e-DIPLOMA for the capture, organization, and interpretation of multimodal data across the three educational prototypes. The first section defines a unified recording schema and describes the event types and metrics to be collected per module, including behavioural actions, system interactions, and psychophysiological signals. These definitions are critical to ensuring data consistency across prototypes, analytical replicability, and future integration into adaptive learning systems.

Building on this foundation, the deliverable details the signal processing process, specifying how raw physiological data (such as heart rate, electrodermal activity, and gaze coordinates) were cleaned, synchronized, and transformed into high-level features. These features are used to derive cognitive, affective, and performance indicators, aligned with the assessment objectives of WP5.

Below, we briefly summarize the data collection methodology, previously detailed in deliverable D5.2 and implemented in the large-scale validation study presented in the same deliverable. This methodology is based on a multimodal, evidence-centred design approach, linking learning outcomes with behavioural and biometric evidence. The data acquisition processes were closely aligned with tasks T5.1, T5.2, and T5.3, and integrated with the technological and pedagogical frameworks defined in Work Packages 3 and 4.

Finally, the deliverable presents the main results of the validation study, illustrating how the metrics and recording strategies presented here allowed for the evaluation of learning effectiveness in the prototypes. These results not only support the methodological soundness of our registration and synchronization strategy but also lay the groundwork for future work in Work Packages 5 (impact assessment, D5.5) and 6 (scaling up and policy recommendations).

2. Methodological Framework for Learning Assessment and Multimodal Data Collection

The e-DIPLOMA evaluation methodology is grounded in a rigorous and multidimensional approach, aimed at assessing how disruptive technologies influence learning performance, user experience, and engagement. The strategy integrates **pre-post testing, psychometric profiling, behavioural logging, and psychophysiological sensing**, forming a comprehensive framework aligned with the Evidence-Centred Design (ECD) and the Cognitive Affective Model of Immersive Learning (CAMIL).

At the core of this methodology is a **comparative design** that contrasts the learning outcomes and cognitive-affective dynamics of users exposed **to disruptive technologies** (VR, AR, AI) against those using **non-disruptive (traditional) digital methods**. Participants are grouped into **experimental** (disruptive technology) and **control** (non-disruptive) conditions, ensuring that both experience the **same instructional content**, differing only in **how** that content is delivered. This design controls for content bias and allows for the isolated analysis of **technological mediation as the key variable**.

Each prototype developed in the project corresponds to a different type of learning content:

- **Prototype 1**: Logical and computational thinking (programming and electronics).
- **Prototype 2**: Social and entrepreneurial reasoning.
- **Prototype 3**: Psychomotor and technical competence in immersive environments.

Each prototype is broken down into multiple **lessons or modules**, and for every lesson, users complete a **pre- and post-assessment** of domain knowledge, emotional state, and a post-experience cognitive load questionnaire. This is complemented by baseline **cognitive profiling** using validated instruments that assess attention, working memory, self-efficacy, curiosity, emotional regulation, and locus of control.

2.1 Triangulated measures for learning evaluation

To address the research questions of WP5, three levels of analysis are integrated:

1. **Subjective and cognitive-affective measures:**
   - Knowledge tests (pre/post).
   - NASA-TLX for perceived workload.
   - SAM scale for emotional valence and arousal.
   - Usability and engagement scales (e.g., presence, agency).

- Baseline cognitive/affective questionnaires (D2, ERQ, CEI-II, GSE, etc.).

2. **Behavioural and interaction metrics:**

- Detailed event logging during interaction (actions, timings, decisions, chat history).
- Gaze data (fixation duration and object mapping).
- Navigation patterns and task completion strategies.

These behavioural markers are used to **infer engagement, strategy use,** and **learning flow**, and are fully synchronized with both system timestamps and physiological signals.

3. **Psychophysiological measures:**

- **EDA (Electrodermal Activity)**: used to infer emotional arousal and sustained attention
- **HR and HRV (Heart Rate Variability)**: used to assess autonomic nervous system responses and stress regulation

All of these measures are captured in real time via **EmotiBit sensors** and the **HP Omnicept HMD**, offering high-resolution data on the **user's cognitive and emotional state throughout the learning experience**.

2.2 Importance of synchronization and multimodal metric extraction

A key aspect of the methodology lies in the **temporal synchronization** of all recorded data: behavioural events, system interactions, gaze data, and physiological signals are **time-aligned** using unified timestamping mechanisms. This enables **fine-grained correlation analyses**—for example, relating peaks in EDA to moments of confusion, stress, or insight during task performance, or assessing whether gaze fixation correlates with learning success in VR object manipulation.

The collected multimodal data serve not only to validate usability and effectiveness across prototypes, but also as a **foundation for future adaptive systems**. The rich logs allow researchers to extract **linked metrics** such as:

- Engagement duration per task
- Response latency in interactive modules
- Physiological reactivity to specific instructional challenges
- Visual attention patterns per content type
- Affective trajectories across sessions

These insights will be key inputs to the upcoming **impact assessment (D5.5)** and **teacher training guidelines (D5.6)**, as well as to the long-term goal of building **evidence-based pedagogical recommendations** for XR-enabled education.

In summary, this methodological framework provides both **robust internal consistency** for comparing learning effectiveness across technologies and the **technical foundation** required to interpret data-driven user experiences in rich, immersive environments. The next sections of this deliverable detail the specific events, interactions and data streams recorded in each prototype, forming the empirical core of this evaluation effort.

3. Event types and logging by Prototype

The e-DIPLOMA project has developed three innovative educational prototypes, each leveraging extended reality (XR), artificial intelligence, and multimodal interaction to enhance learning outcomes in different domains. **Prototype 1** focuses on introductory programming and electronics, combining theoretical audiovisual content with hands-on block-based coding and sensor-based activities in virtual and augmented reality. **Prototype 2** targets the development of social entrepreneurship competencies through immersive simulations, conversational AI, and decision-making games that simulate the complexity of managing social enterprises. **Prototype 3**, designed for educators, explores the use of virtual reality in education, offering experiential modules on interaction, navigation, visualization, collaboration, and advanced educational applications.

To evaluate the impact and effectiveness of each prototype, a detailed **event logging system** has been implemented across modules. These systems capture behavioural, physiological, and interaction data—including eye-tracking, EmotiBit biosignals, chat interactions, and gameplay metrics—providing a rich foundation for the analysis of user engagement, learning processes, and system usability. This section outlines the key events recorded per prototype and module, highlighting their alignment with the educational objectives, evaluation framework, and data collection needs defined in WP5.

The objective of this event logging strategy is twofold: (1) to support **quantitative assessment of user experience and learning progression**, and (2) to enable **multimodal correlation between behavioural and psychophysiological data**.

The event collection mechanisms have been implemented in line with the **project's evaluation framework (linked to WP5)** and the **educational and technological requirements of each prototype (WP3 and WP4)**. By synchronizing data across modalities—such as physiological signals, user decisions, gaze behaviour, and interaction patterns—these logs allow us to **evaluate user engagement, cognitive load, and usability**, while also informing the design of adaptive learning strategies.

Each module logs data that is specific to its interaction design and technical capabilities, ensuring coverage of the evaluation dimensions defined in **Deliverable D5.2**, and providing the raw material for the upcoming **impact (D5.5) and teacher training reports (D5.6)**. These data points also directly support the educational validation activities described in **T5.2 and T5.3**, enabling fine-grained analysis of prototype effectiveness.

4. Event logging across Prototypes

4.1 Prototype 1 – Block Programming

This prototype introduces fundamental concepts of programming and electronics using immersive and interactive tools. It alternates between **theoretical instruction** and **practice-based learning** using VR and AR technologies.

4.1.1 P1M1 – Theoretical video module

This module delivers foundational knowledge through a series of short instructional videos on basic programming concepts.

- **Technology**: Edison-based audiovisual content accessed through Moodle.
- **Logged events**: Timestamps for video start and completion.
- **Evaluation focus**: Measures passive engagement, attention span, and completion rate, which are later cross-referenced with learning assessments and retention scores

4.1.2 P1M2 – Explore BlockCoding in VR

A VR-based interactive module where students solve programming challenges using block-based logic to control a robot in a virtual environment.

- **Technology:** Unity VR, HMD (HP Reverb)
- **Events captured:**
    - Teleportation.
    - Object manipulation (grab, move, release).
    - Object removal/reset.
    - UI interactions.
    - Zoom and rotation (via joysticks).
- **Data collected**: Time-stamped logs of each interaction.

- **Evaluation focus: Tracks problem-solving processes, interaction fluency, and understanding of programming constructs through behavioural analytics.**

4.1.3 P1M3 – Basis of Electronics

Similar to Module 1, this component introduces microcontroller concepts (Arduino, sensors, actuators) through video explanations.

- **Technology**: Audiovisual content with 3D object integration (Edison).
- **Logged events**: Passive viewing timestamps.
- **Evaluation focus**: Completeness of video consumption and correlation with performance in the following hands-on AR module.

4.1.4 P1M4 – Arduino Learn

This AR application enables students to recognize and assemble real electronic components over a printed template using their smartphone/tablet.

- **Technology**: Android AR app; physical components + printed interaction board.
- **Logged events**:
    - Component recognition.
    - Interaction attempts.
    - Placement errors.
    - Duration of interaction.
    - Final accuracy.
- **Evaluation focus**: Assesses user autonomy in applying theoretical knowledge to real-world component assembly; supports cognitive and motor learning evaluation

A two-player VR scenario where participants must jointly assemble and program a robot to solve a global challenge using sensors and code.

- **Technology**: Multiplayer VR (Unity + Photon Fusion), Arduino logic simulation, HMDs.
- **Logged events**:
    - Teleportation.
    - Object manipulation (grab, move, release).
    - Object removal/reset.
    - UI interactions.
    - Zoom and rotation (via joysticks).
- **Data collected:** Time-stamped logs of each interaction.
- **Evaluation focus**: Integration of learned concepts and collaboration skills, enabling cross-validation with group-based engagement metrics.

4.2 Prototype 2 – Social entrepreneurship

This prototype provides a diverse and immersive training program on how to conceptualize, plan, and manage social enterprises, using simulations, games, AI-based roleplay, and strategic decision-making tools. Each module presents different types of technology.

4.2.1 P2M1 – Module 1 Lodestars: The Social Enterprise and the Social Entrepreneur

A single-user VR experience where students interact with AI-driven NPCs representing social actors. Conversations reveal traits and motivations of characters, which students must classify through group discussion.

- **Technology**: Unity VR, Whisper/AWS Transcribe (STT), LLMs (Llama 3, GPT), TTS (LMNT), EmotiBit, raycasting-based eye tracking.
- **Logged events**:
    - EmotiBit physiological data (heart rate, temp, GSR).
    - Eye-tracking via raycasting (object focus + duration).
    - Chat logs with AI NPCs (content, time, discovered values/traits).
    - UI interaction events.
- **Evaluation focus:** Deep learning, critical analysis, empathy development and attention mapping within immersive narratives.

4.2.2 P2M2 – Module 2 Heroes: Urban management for social outcomes

A multiplayer VR simulation where students act as policy makers managing urban resources to reduce inequality.

- **Technology**: Unity VR, HP Omnicept for biometric capture.
- **Logged events**:
    - Heart rate, HRV, eye movement, cognitive load.
    - IMU data (acceleration/rotation).
    - Timestamped synchronizations.
- **Evaluation focus**: Real-time cognitive and affective state tracking, foundation for adaptive feedback mechanisms.

4.2.3 P2M3 – Module 3 Painters: Immersive BMC for Energetic Communities

A Teleconferencing session employing Teams where students participate in a Business model canvas exercise supported by a moderator. Users can log in a remote desktop session into a Virtual machine hosted by AWS for the e-Diploma Platform. The exercise is focused on developing a business model to improve local energetic resources sustainability.

- **Technology**: MS Teams, Edison PRO, Emotibit.
- **Logged events**:
  - GSR, Heart Rate
  - Business Model Canvas diagram.
  - Timestamped synchronizations.

Evaluation focus: Real-time cognitive and affective state tracking, computer interaction abilities, Business modelling abilities.

4.2.4 P2M4 – Module 4: Allies: Human resources and team management

A management simulation game where students hire, interview (via AI), and manage employees in a virtual social enterprise.

- **Technology**: Unity + LLM (Llama 3), EmotiBit, JSON-based configuration for personas.
- **Logged events**:
- Game state: time per scene, hires, company stats
- Chat: full message history with categorization (Skills, Wage, etc.)
- Traits and values discovered
- Game events:
  - jobEnabled, jobDisabled
  - hireCharacter, fireCharacter
  - selectChat, sendMessage
  - generateTask, startTask, finishTask
  - sceneChange
- **Evaluation focus**: Provides a detailed trace of decision patterns, interpersonal judg ment, and multitasking behaviour.

4.2.5 P2M5 – Module 5: Angels: Social product management and marketing

An online card-based business simulation game where students must fund, develop, market, and scale socially impactful products.

- **Technology**: Web-based multiplayer (Kotlin + JS), Unity-based EmotiBit logging companion app.
- **Logged events**:
  - **Two-level event system**:
    - **Game events**: HoverCard, ClickCard, PlayBrand, PlayFinancing, PlayStoreOnZone, etc.
    - **Score events**: PlayerFinished, PhaseChange, final score.
  - **Companion app**: Handles EmotiBit stream logging + real-time WebSocket communication.

- **Evaluation focus**: Enables full synchronization between browser interactions and physiological responses, supporting end-user feedback loops and adaptive UX analysis.

4.3 Prototype 3 – Virtual Reality in education

Prototype 3 is structured into a sequence of modular learning activities, implemented in XR environments designed to foster interaction, exploration, collaboration, and content creation in STEM education. Each module defines a set of tracked events that enable detailed analysis of user behaviour, interaction quality, and learning progression. All modules feature the same technology, so the events and logs collected are similar but adapted to the specific content of each lesson/module.

4.3.1 P3M1 – Interaction

- **Phase**: Object manipulation with cubes.
- **Logged events**:
    - start_phase: start of the interaction sequence.
    - selectEntered: when the user grabs a cube (CubeX-Y).
    - selectExited: when the user releases the cube.
- **Evaluation dimension**:
    - Measures the time required for understanding and executing manipulation.
    - Indicators of task comprehension and motor interaction (WP5, T5.2).

4.3.2 P3M2 – Navigation

- **Phase**: Free navigation and objective completion.
- **Logged events**:
    - teleport [start|stop]: player teleportation movement.
    - goal: enclosing sheep within green areas (SheepX-Y).
    - controller: joystick/button input events (type, hand, value).
- **Evaluation dimension**:
    - Spatial reasoning and environmental awareness.
    - Precision and timing of navigational decisions (WP5, T5.3).

4.3.3 P3M3 – Visualization

- **Phase**: Visual object stylization and manipulation.
- **Logged events**:
    - button_pressed: interaction with buttons.
    - move_buttom: grip activation.

- object_move: repositioning of visual elements.

- **Evaluation dimension**:
  - Active engagement in visual content transformation.
  - Competence in visual representation and decision-making.

4.3.4 P3M4 – Demonstration

- **Phase**: Guided collaborative experience.
- **Logged events**:
  - times_joystick_pressed: frequency of joystick actions.
  - achieve_goal: reaching designated collaborative targets.
- **Evaluation dimension**:
  - Assessment of group coordination and communication.
  - Behavioural markers of collaborative learning (WP5, D5.5).

4.3.5 P3M5 – Builder

- **Phase**: Virtual museum construction.
- **Logged events**:
  - object_graved: selected objects for construction.
  - button_pressed: activation of construction mechanics.
  - Level_achieve: successful completion of the building activity.
- **Evaluation dimension**:
  - Indicators of creativity, agency, and task planning.
  - Final learning outcome in applied content creation (WP5, T5.2).

5. Linking measured constructs to data sources and events.

To ensure validity and coherence in the evaluation process, each key educational or psychological metric has been explicitly linked to its corresponding **data capture modality** (sensor, event, or questionnaire) and the **cognitive/affective construct** it informs. This traceability is crucial for interpreting the logs recorded during the XR lessons and ensures that data extraction aligns with the evaluation framework defined in WP5.

Each extracted metric is triangulated from at least one **objective source** (e.g., sensor or event log) and one **subjective source** (e.g., validated questionnaire), allowing robust cross-validation. This mapping supports the data interpretation pipeline and ultimately informs the guidelines for adaptation and evaluation to be delivered in D5.6.

The following table presents a consolidated view of this mapping, covering metrics extracted across all lessons and technological conditions.

Extracted Metric	Data Source	Associated Constructs
Cognitive Load	HP system (task time + effectiveness)	Cognitive Effort
Visual Attention	Eye Tracking (AOI, scan path, ray cast)	Attention & Focus
Emotional Arousal	EDA + HR (GSR, peaks under stress)	Stress/Activation
Emotional Regulation	HR + GSR + time-limited decision triggers	Emotional Control
Self-Efficacy	System behaviour + distractors + GSE Questionnaire	Confidence in Action
Autonomy	System logs + response to help triggers	Decision-Making Independence
Curiosity	Exploration behaviour + Eye Tracking	Exploratory Drive
Critical Thinking	Event logs + decision paths	Reasoning & Evaluation
Cognitive Flexibility	Task switching patterns + sensor reactivity	Adaptability under Pressure
Knowledge Acquisition	Pre/Post Questionnaires	Learning Outcome
Motivation	Locus of Control Questionnaire	Internal Motivation
Collaboration	Number of interactions per session	Teamwork & Social Dynamics
Empathy	Behavioural activity + reflection questions	Perspective Taking
Presence & Usability	Post-session UX and Presence Questionnaires	User Experience

Table 1. Metrics extracted across all lessons and technological conditions.

6. Multimodal data architecture and synchronization

The meaningful interpretation of psychophysiological and behavioural data in immersive learning environments requires not only robust experimental design, but also a **systematic data architecture** capable of aligning and organizing the different signal types recorded across prototypes. This section presents the **technical structure and synchronization strategies** used to process and harmonize the data collected in e-DIPLOMA, with particular emphasis on the heterogeneity of devices, formats, and storage conventions employed in each prototype.

This section serves as a **bridge between raw data acquisition and high-level analysis**, detailing how heart rate (HR), cognitive load (CL), electrodermal activity (EDA), and task event logs were extracted, cleaned, and temporally aligned. Given the project's multimodal nature and the distributed implementation across countries and partners, the reliability of the analytical outputs hinges on the integrity of this foundational step.

Specifically, the section describes the **file structures and custom parsing functions** developed for each prototype—Prototype 1 (logical content), Prototype 2 (social content), and Prototype 3 (psychomotor content)—as well as the differences in data sources (e.g., OMNICEPT JSONs, EmotiBit streams, .log files). It also explains the modular data extraction scripts implemented in MATLAB and Python, which allow for scalable, reusable pipelines.

In short, this section outlines how diverse multimodal data streams were **transformed into a coherent analytical dataset**, ready for modelling and hypothesis testing. It lays the technical foundation for the psychophysiological analyses presented in the Results section.

6.1 Prototype 1

CSV files (MATLAB implementation)

[dataMatrix] = extract_data(fileMatrix, rootPath, subfolder)

This function is designed to systematically extract physiological data from multiple subjects and experimental modules. It receives as inputs a matrix containing the subject identifiers and the corresponding modules, the root directory where the data is stored, and the name of the subfolder containing the physiological data files (e.g., SENSORS).

For each subject and module combination, the function searches within the corresponding folder for CSV files related to two specific physiological signals: Heart Rate and Cognitive Load. These files are expected to have a standardized structure, from which the function extracts two key columns: one containing time information (e.g., timestamp) and another containing the recorded physiological values.

The extracted data are cleaned and converted into numeric format, ensuring the validity of each entry. The result is organized into a cell array (a structured matrix), where each element contains the cleaned data for a specific subject and module. Each entry in this matrix consists of a structured format with two fields: one for Heart Rate and one for Cognitive Load, each of which may contain one or more arrays with time-value pairs.

This modular data extraction procedure facilitates the organization and subsequent analysis of physiological recordings across multiple experimental conditions and participant profiles, and it can be adapted to other formats or programming languages in future implementations.

EMOTIBIT files

[dataMatrix] = extract_dataEMOTIBIT(fileMatrix, rootPath, subfolder)

The function extract_dataEMOTIBIT is designed to extract physiological data from JSON files associated with the EMOTIBIT system. It operates on a file matrix (fileMatrix) that contains subject identifiers and experimental module names and uses a given rootPath and subfolder (subfolder) to locate the corresponding JSON files.

The function begins by parsing the header and content of fileMatrix, separating subject IDs and module labels. It initializes a cell matrix dataMatrix where each row corresponds to a subject and each column to a module. The function iterates through each subject and each module, building the file path to the expected EMOTIBIT subfolder. It then searches for files that contain the substring emotibit_biometric and have a .json extension.

Each JSON file is parsed to extract entries from the values field, which are assumed to contain a timestamp (time) and a comma-separated string of measurements (value). The function converts the timestamp into Unix format and splits the value string into individual components. It focuses specifically on three types of signals: Heart Rate (HR), Electrodermal Activity (EA), and Skin Temperature (TH), corresponding to positions within the string format. If the identified signal type matches one of these three, the data sample (timestamp and values) is appended to the respective field in a temporary structure named moduleData.

An important difference from the OMNICEPT version is that in this case, the data arrays are not initially converted into numeric matrices. Instead, they are collected as cell arrays, and only after confirming that the structure is not empty, the arrays are homogenized. To do this, the function ensures that all data rows within each field (HR, EA, or TH) have the same number of columns by padding with NaN if necessary. Then, the data is concatenated vertically. However, only the first two columns of each sample are preserved, under the assumption that they correspond to the Unix timestamp and a primary measurement value.

If the module subfolder does not exist or no valid data is found, an empty structure is inserted in dataMatrix for that subject and module combination. The function returns a complete dataMatrix containing the structured physiological data for each subject and module.

Compared to the extract_data function designed for OMNICEPT, this version is tailored for the specific formatting and structure of EMOTIBIT files, which may differ in how data is embedded within the value field and the expected signal types. Additionally, this function explicitly enforces column alignment and truncation to two columns, a step not presented in the OMNICEPT version.

6.2 Prototype 2

OMNICEPT files (read and load of JSON data)

[dataMatrix] = extract_data(fileMatrix, rootPath, subfolder)

The extract_data function is designed to read, process, and organize physiological data stored in .json files generated by the OMNICEPT system (Prototype 2). Its overall structure and traversal logic are very similar to the previously described extract_dataEMOTIBIT function but specifically adapted to the file format and naming conventions used in OMNICEPT.

This function takes as input a file matrix (fileMatrix) containing subject identifiers and module names, along with the root directory path (rootPath) and the name of the subfolder where the data files are located. The output is a cell array (dataMatrix) organized by subject (rows) and module (columns), where each cell contains a struct with fields corresponding to the available physiological variables: HeartRate and CognitiveLoad.

Unlike the previous function, this one explicitly searches for .json files whose names start with HeartRate_ or CognitiveLoad, in line with OMNICEPT's naming conventions. Each file is read using the decode_safe_json function, and the extracted data includes both the timestamp (Timestamp) and the corresponding physiological value (HeartRate or CognitiveLoadValue). Timestamps are converted to Unix format (posixtime) to facilitate downstream processing.

If no files are found or read errors occur, warnings are printed to the console, and the corresponding dataMatrix cell is filled with an empty struct containing the expected fields.

This function enables systematic organization of physiological data extracted from Prototype 2, and its modular structure allows for easy integration with subsequent analysis functions, while maintaining design consistency with similar functions used for other prototypes.

EMOTIBIT files

[dataMatrix] = extract_dataEMOTIBIT(fileMatrix, rootPath, subfolder)

Same as above, corresponding to Prototype 1

6.3 Prototype 3

LOG files (MATLAB implementation)

For the processing of data corresponding to Prototype 3, a MATLAB script was developed to automate the reading and organization of .log files stored within the project's hierarchical folder structure. The root path contains individual folders for each experimental subject, identified by unique codes (e.g., SP_EXP_026, SP_EXP_027, etc.), and within each subject folder, there are five subfolders corresponding to the different experimental modules (M1 through M5).

Initially, a cell array (fileMatrix) with headers was created to organize the names of available .log files for each subject and module. Subsequently, module-specific matrices (e.g., eventMatrix_M1) were generated, where only files starting with the prefix "Interaction" were loaded, as these contain the relevant information about task execution and physiological variables recorded during each session.

Due to the large size and detailed nature of the original data, a subsequent filtering step was implemented to reduce the amount of stored information. A reduced version of each matrix (e.g., eventMatrix_M1_light) was created, retaining only the fields of interest: sessiontimestamp, complete, hr, cl, and hrv. This selection preserves key physiological variables (heart rate, cognitive load, and heart rate variability) as well as a basic task performance indicator (complete), without compromising further analysis or overloading memory. The same procedure was applied across all five modules, ensuring consistency in processing all available records.

7. Validation study

Following the detailed specification of logging strategies, event structures, and multimodal data synchronization protocols, this deliverable shifts focus to the empirical validation of the e-DIPLOMA prototypes.

The second part of this document presents the outcomes of the validation study, demonstrating how the structured metrics and logging systems previously described enable rigorous analysis of learning performance across content domains and technological conditions.

8. Research Objectives and Hypotheses

The validation study presented in this deliverable is grounded in the theoretical and methodological framework detailed in *Deliverable D5.2*, where the scientific rationale, experimental design, and learning hypotheses of the e-DIPLOMA project were established. That document outlines the integration of evidence-centred design (ECD) principles with the Cognitive Affective Model of Immersive Learning (CAMIL), providing the conceptual backbone for the assessment activities conducted in WP5.

Building on this foundation, the current study addresses three central research objectives, each linked to a core hypothesis that guides the evaluation of the platform's educational impact:

1. **Objective 1: Effectiveness of Immersive Technologies**: The first objective is to assess whether disruptive technologies—such as Virtual Reality (VR), Augmented Reality (AR), and AI-driven tools—enhance learning outcomes compared to traditional, non-immersive approaches. As hypothesised in D5.2, immersive technologies are expected to foster greater engagement, attention, and cognitive integration, resulting in superior performance in post-intervention knowledge assessments.

2. **Objective 2: Role of Content Type:** The second objective is to investigate how the nature of the learning content (logical, social, or psychomotor) modulates the effectiveness of immersive learning. The three prototypes developed within the project target distinct content domains: logical (Prototype 1), social (Prototype 2), and psychomotor (Prototype 3). In line with the theoretical framework proposed in D5.2, it is hypothesised that immersive technologies yield stronger benefits for logical and psychomotor content—due to their high visual-spatial and procedural demands—whereas social content may require alternative strategies or modalities to fully exploit the potential of immersion.
3. **Objective 3: Interaction with Learner Characteristics**: The third objective focuses on the influence of individual cognitive and emotional traits on learning performance. Specifically, it examines whether baseline attributes such as attentional control, self-regulation, and emotional resilience moderate the efficacy of immersive environments. As outlined in D5.2, the working hypothesis is that learners with stronger self-regulatory and cognitive-affective skills will derive greater benefit from immersive and complex learning settings.

These hypotheses structure the experimental logic of the validation study and are operationalised across the three prototypes and learning modalities. The present validation study complements D5.2 by providing empirical results that test these hypotheses through a combination of behavioural data, psychophysiological measures, and self-reported metrics.

9. Experimental Methodology

9.1 Participants and sampling strategy

The validation study involved a total of 408 participants distributed across five EU countries: Spain, Hungary, Cyprus, Italy, and Estonia. Participants were recruited through partner institutions and stratified according to the three educational prototypes developed within the project. Each prototype included approximately 150 individuals, divided into three primary profiles: students (n≈90), trainers (n≈30), and professionals (developers, content creators, or social entrepreneurs, n≈30), depending on the prototype. Eligibility criteria required participants to be over 16 years old, possess reading and comprehension skills in the national language or English, and have no history of seizures or photosensitive conditions. The experimental design ensured gender and cultural diversity to assess inclusiveness and generalisability of the findings.

9.2 Experimental protocol and design

Each participant was randomly assigned to either an experimental group, which engaged with immersive learning technologies (VR, AR, AI), or a control group, which received the same content in a traditional, non-immersive format. Both groups completed a multi-phase protocol structured around three main stages: baseline, learning experience, and post-evaluation.

The baseline phase began with informed consent and collection of sociodemographic information. Participants then completed a battery of cognitive and affective assessments (detailed below), followed by the placement of biometric monitoring devices (Emotibit). A two-minute resting state recording was taken to establish a psychophysiological baseline (HR and EDA). Before each learning session, participants completed the SAM (Self-Assessment Manikin) and a pre-test of content-specific knowledge.

During the learning experience, participants engaged with a sequence of lessons designed according to their assigned prototype and condition. In the experimental group, lessons used HP Reverb G2 Omnicept (for VR) integrated with biometric recording. In the control group, the same learning content was delivered via presentations, videos or printable materials without immersive features. Actions, task performance and gaze behaviour were also logged in the experimental group for behavioural analysis.

After each lesson, participants completed a post-test version of the knowledge questionnaire, a second SAM scale, and the NASA Task Load Index to measure perceived cognitive load. Upon finishing all lessons of a prototype, they filled out a comprehensive usability questionnaire covering presence, system ease-of-use, motivation and emotional engagement.

9.3 Questionnaires and measures

A wide range of standardised instruments was employed to assess cognitive capacities, affective states, and learning outcomes. The overview of these instruments, including their purpose, administration format and timing, is presented in the table titled "**c**". These include:

- **D2 Test** and **Digit Span** (administered on paper) to measure attentional control and working memory.
- **ERQ**, **Locus of Control**, **CEI-II**, and **GSE** scales (digital) to assess motivation, self-regulation, and curiosity.
- **Self-Assessment Manikin (SAM)** was administered before and after each lesson to track emotional valence, arousal, and dominance.

- **NASA-TLX** and a custom-designed **Usability Questionnaire** were used to assess cognitive workload and platform usability.
- **Knowledge Questionnaires**, aligned to each prototype, were delivered before and after every lesson to quantify learning gains.

Questionnaire	Purpose	Format	Administration
1. D2 Test	Selective attention and processing speed	Paper	Before XR
2. Digit Span	Working memory capacity	Paper	Before XR
3. Self-Assessment Manikin (SAM)	Emotional state (valence, arousal, dominance)	Digital	Before & after each lesson
4. Locus of Control	Motivation (internal vs external locus)	Digital	Before XR
5. ERQ	Emotion regulation strategies (reappraisal, suppression)	Digital	Before XR
6. CEI-II	Curiosity (exploratory behaviour and uncertainty acceptance)	Digital	Before XR
7. GSE	Self-efficacy (belief in one's capacity to perform tasks)	Digital	Before XR
8. Knowledge Questionnaire	Domain-specific learning assessment (pre/post)	Digital	Before & after each lesson
9. NASA-TLX	Cognitive load after each lesson	Digital	After each lesson
10. Usability Questionnaire	Ease of use, interaction and engagement with the system	Digital	After full prototype

Table 2: e-DIPLOMA Questionnaires Overview

These assessments allowed a multidimensional analysis of user experience, learning effectiveness, and the modulation of outcomes by individual characteristics.

9.4 Materials and devices

Three main hardware elements were used in the immersive learning experiences:

- **HP Reverb G2 Omnicept Edition**, a VR headset equipped with real-time physiological sensors for eye tracking, HR, and Cognitive Load.

- **Emotibit**, a wearable sensor for high-resolution measurement of electrodermal activity and heart rate variability, connected via Bluetooth.

These tools enabled simultaneous collection of behavioural and physiological data throughout the learning sessions, ensuring a fine-grained analysis of user engagement and stress/arousal patterns.

9.5 Learning content and prototype overview

Each prototype was composed of multiple lessons with increasing complexity. Prototype 1 (logical content) included five lessons on block-based programming and sensor manipulation. Prototype 2 (social content) contained five lessons oriented toward social entrepreneurship and ethical business simulation. Prototype 3 (psychomotor content) comprised five VR navigation and interaction lessons aimed at building fluency with immersive systems. All prototypes integrated gamification elements, collaborative scenarios, and autonomous problem-solving tasks in the experimental group. These prototypes are explained in greater detail in the prototype description sections in each of the usability studies in this same deliverable, as well as in deliverable D4.1.

9.6 Timeline and session flow

Each experimental session was structured to last no more than 90 minutes. Participants progressed through informed consent, baseline data collection, one or more learning lessons, and a post-evaluation phase. For multi-day prototypes, new psychophysiological baselines were recorded at the start of each session. A detailed visual timeline of the protocol phases and device setup is available in the original methodological deliverable (D5.2).

10. Analysis of objectives 1 and 2: Results and conclusions

10.1 Data structure

The dataset analysed in this study originates from a series of educational experiments designed to evaluate the impact of interventions mediated by digital and immersive technologies, in comparison with traditional teaching methods. The interventions were organized into three prototypes, differentiated by the type of content addressed: Prototype 1: logical-mathematical content, Prototype 2: social content, and Prototype 3: psychomotor content.

Each prototype consists of five modules (M1 to M5), resulting in a total of fifteen modules implemented. For each module, a knowledge assessment questionnaire was administered at two time points: before (pre-test) and after (post-test) the intervention. Although the questionnaires differ in structure (e.g., number of items), all share a common maximum score of 10 points, allowing for comparative analysis across modules.

The interventions were conducted with participants from three different countries. The experimental design included two groups: an experimental group, exposed to technology-supported interventions (with varying degrees of immersion), and a control group, which received equivalent instruction without technological mediation. Both groups were evaluated using the same pre- and post-module questionnaires.

Table 3 provides a descriptive summary of the number of available observations, categorized by prototype, group (experimental or control), country of origin, and corresponding module.

Prototype	**Group**	**Country**	**Module 1**		**Module 2**		**Module 3**		**Module 4**		**Module 5**	
			Pre	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**
1 Logical	Control n=76	CY, n=24	23	24	24	24	22	22	22	22	21	20
		HU, n=28	28	28	28	28	28	28	28	28	23	23
		SP, n=24	23	23	23	23	23	23	22	22	23	23
	Experimental n=72	CY, n=20	20	20	20	20	20	20	20	20	19	19
		HU, n=28	27	27	27	27	26	26	26	26	26	26
		SP, n=24	24	24	24	24	24	24	24	24	24	24
2 Social	Control n=51	EE, n=7	7	7	7	7	7	7	7	7	7	7
		HU, n=19	18	18	18	18	18	18	18	17	18	18
		IT, n=25	25	25	25	23	25	25	24	24	24	25
	Experimental n=72	EE, n=24	24	24	22	24	24	24	24	23	24	24
		HU, n=23	23	23	23	23	23	23	23	23	23	23
		IT, n=25	25	9	25	24	25	25	25	25	25	25
3 Psychomotor	Control n=65	CY, n=22	18	18	17	18	18	18	18	18	18	17
		HU, n=24	24	24	24	24	24	24	24	24	24	24
		SP, n=19	18	18	17	17	17	17	17	17	17	16
	Experimental n=72	CY, n=22	22	22	0	22	19	18	14	14	10	9
		HU, n=26	18	18	0	18	18	18	18	18	18	18
		SP, n=25	23	23	0	23	23	23	23	23	23	23

Table 3: Prototype data organization.

10.2 Preprocessing

A. Preliminary organization and filtering

The dataset was categorized based on two main factors: the prototype (1, 2, or 3) and the group assignment (experimental or control). Observations were organized according to each participant's performance in each of the applied modules. To ensure the validity of comparisons, only those observations with complete data in both the pre-test and post-test knowledge assessments were selected. These were processed independently for each module. Table 4 presents the total number of valid observations obtained after the filtering process.

Filtering was strictly limited to the availability of both knowledge scores per module. No additional exclusions were made due to missing data in other variables, such as sociodemographic characteristics. After filtering, 89% of the original observations were retained. The retention rate by prototype was as follows: 96.1% in Prototype 1, 95.3% in Prototype 2, and 75.8% in Prototype 3.

Prototype	Group	Module 1	Module 2	Module 3	Module 4	Module 5	Total	Possible
1 Logical	Control n=76	74	75	73	72	66	**360**	**380**
	Experimental n=72	71	71	70	70	69	**351**	**360**
2 Social	Control n=51	50	48	50	48	49	**245**	**255**
	Experimental n=72	56	70	72	71	72	**341**	**360**
3 Psychomotor	Control n=65	60	58	59	59	57	**293**	**325**
	Experimental n=72	63	0	59	55	50	**227**	**365**

Table 4: Frequency and distribution of filtered data.

B. Learning metrics

The knowledge questionnaire scores are numerical and normalized on a scale from 0 to 10 points, regardless of the module or prototype in which they were administered. However, since the questionnaires differ in their structure—particularly in the number and type of items—the value assigned to each item may vary between modules. To quantify each participant's learning, two classical metrics were used: Absolute Difference (AD), calculated as the difference between the post-test and pre-test scores for each module. This metric allows a direct identification of whether the participant's performance improved, declined, or remained stable between the two evaluation points; and Normalized

Gain (NG), which quantifies the relative percentage of improvement with respect to the margin available to reach the maximum score. This measure is defined within the interval [0, 1], and considers the distance between the pretest score and the ideal value (10 points). Equations 1 and 2 formally describe both metrics.

$AD = Post - Pre$	**Equation 1**
$NG = (Post - Pre)/\,(10 - Pre)$	Equation 2

The AD metric allows the evaluation of performance changes across all available observations, regardless of the magnitude or direction of the change. On the other hand, NG is undefined for cases where the pretest score is equal to 10, as its calculation involves division by zero. Furthermore, due to its formulation, the NG metric produces negative values in observations where the post-test score is lower than the pretest score, which introduces interpretative distortion if the focus is solely on improvement. As shown in Figure 1A, NG tends to amplify cases of performance decline. For this reason, subsequent analyses based on NG included only observations with calculated values greater than or equal to zero. This restriction resulted in retention of 68% of valid observations in the control group and 60% in the experimental group (Figure 1B).

A

Pre\Post	0	2	4	6	8	10
0	0.00	0.20	0.40	0.60	0.80	1.00
2	-0.25	0.00	0.25	0.50	0.75	1.00
4	-0.67	-0.33	0.00	0.33	0.67	1.00
6	-1.50	-1.00	-0.50	0.00	0.50	1.00
8	-4.00	-3.00	-2.00	-1.00	0.00	1.00
10	—	—	—	—	—	—

B

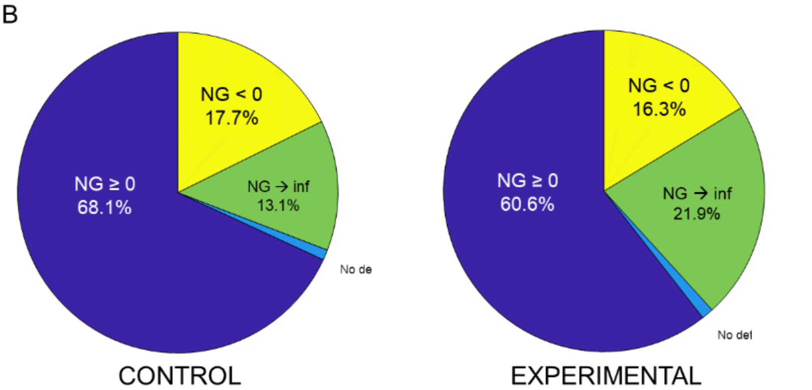


Figure 1: Distribution of Normalized Gain (NG. A) Illustrative matrix of possible NG values based on combinations of pre and post-test scores on a 0 to 10 scale. B) Distribution of observations according to NG values in the control group. C) Equivalent distribution in the experimental group.

10.3 Descriptive statistics

Data normality was assessed using the Kolmogorov-Smirnov test, which is recommended for samples larger than 50 observations. The test was performed on the distributions of the learning metrics, grouped independently by prototype and by group (experimental and control). The normality test results indicated that the distributions did not fit the normal model in any of the cases evaluated. Consequently, non-parametric statistics were chosen for subsequent stages of the analysis.

Table 5 summarizes the descriptive statistics obtained for each group within the three prototypes. For each combination, the total number of valid observations, the medians of pretest and post-test scores, the absolute difference (AD), and the normalized gain score (NG) are presented. In all cases, the full range of observed values is also reported.

Prototype	Group	Observations	Pre	Post	Absolute Difference	Normalized Gain
			Median [Min Max]	Median [Min Max]	Median [Min Max]	Median [Min Max]
1 Logical	Control	360	7.14 [0 10]	8 [0 10]	0.36 [-6 7.14]	0.45 [0 1]
	Experimental	351	7.14 [0.5 10]	8.33 [1.1 10]	0.37 [-4.3 7.5]	0.43 [0 1]
2 Social	Control	245	5.56 [0 10]	6.67 [0 10]	0.55 [-6.6 8.3]	0.37 [0 1]
	Experimental	340	6.19 [0 10]	6.67 [0 10]	0 [-6.6 6.6]	0.20 [0 1]
3 Psychomotor	Control	292	5 [0 10]	6.67 [0 10]	1 [-4 9.2]	0.33 [0 1]
	Experimental	227	5 [0 10]	7 [0 10]	1.66 [-4.6 8.6]	0.40 [0 1]

Table 5: Descriptive statistics of learning.

10.4 Comparative statistics using absolute difference (AD)

A. Objective 1: Results and conclusions

The first general hypothesis of the study was formulated as follows: Learning is more effective using technologies vs. non-technologies. To evaluate this, observations corresponding to each prototype were grouped considering only the type of group: experimental (technology use) and control (no technology use). Normality tests applied to both groups revealed that the distributions did not conform to a normal distribution; therefore, non-parametric statistical tests were employed. The significance level adopted for the comparison was $p < 0.05$. Table 6 presents the descriptive statistics corresponding to this grouping, including median, range of values, and sample size per group.

Group	Observations	Pre	Post	Absolute Difference

		Median [Min Max]	Median [Min Max]	Median [Min Max]
Control	897	6.00 [0 10]	7.14 [0 10]	0.83 [-6.7 9.2]
Experimental	918	6.11 [0 10]	7.14 [0 10]	0.50 [-6.7 8.6]

Table 6: Descriptive statistics in large groups with Absolute Difference.

To compare the initial knowledge level between the experimental and control groups, the Mann-Whitney-Wilcoxon test was applied. To evaluate learning within each group, that is, the change between pretest and post-test scores, the Wilcoxon signed-rank test was used. Finally, to compare learning between groups, measured by the AD metric, the Mann-Whitney test was employed again.

The results indicate that both groups started with equivalent knowledge levels (Figure 2A). Likewise, significant learning was observed within each group, with p values < 0.001 for both the experimental and control groups (Figure 2B). However, no significant differences were found in the magnitude of learning between the two groups (Figure 2C). Together, these findings suggest that the use of educational technologies was not associated with superior learning improvement compared to traditional teaching, at least under the evaluated conditions.

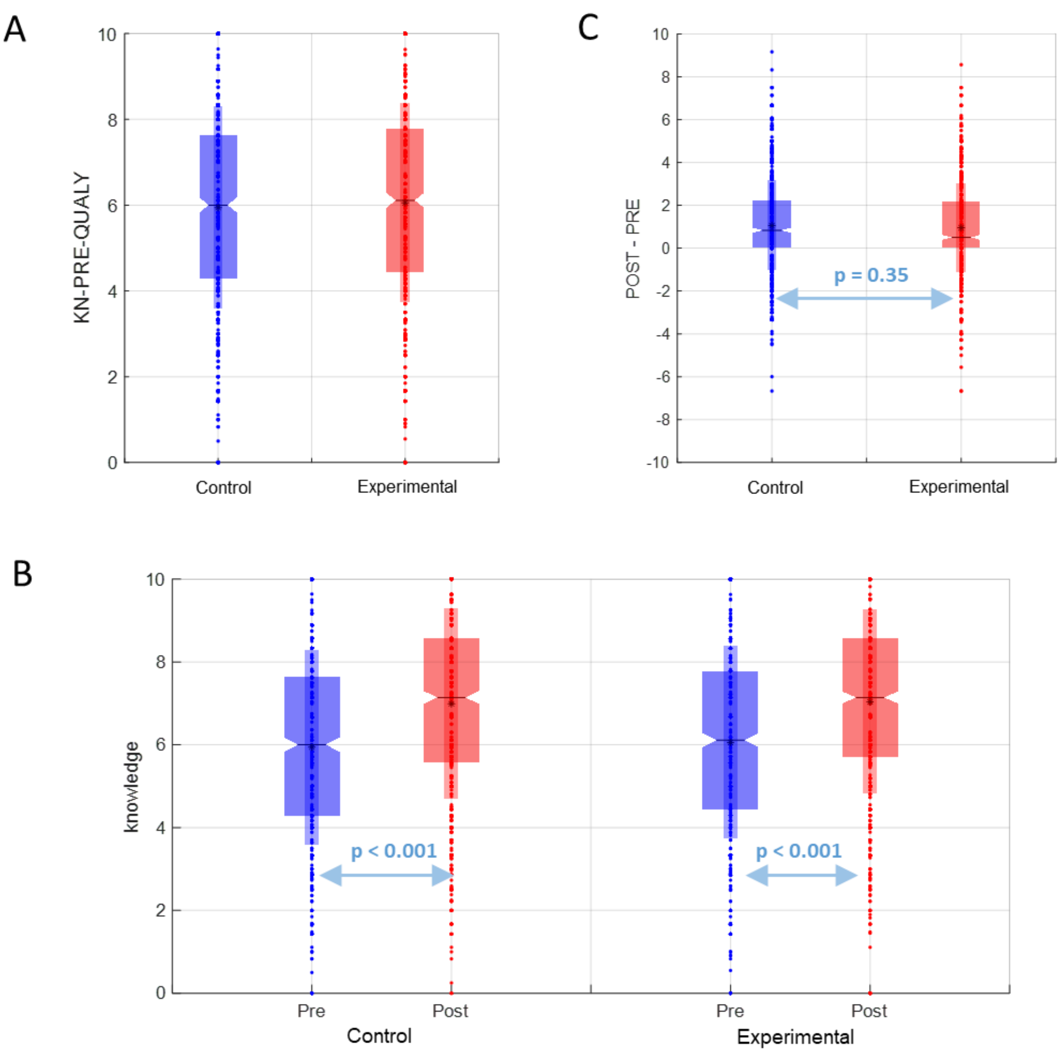


Figure 2: Distribution of AD values and results of the statistical analysis comparing initial knowledge levels, learning within each group, and between control vs. experimental methodologies. A) Initial knowledge levels in both groups. B) Comparison of the acquired data using control vs. experimental methodologies. C) Pre-post differences within each group.

B. Objective 2: Results and conclusions

The second hypothesis of the study was stated as follows: The type of content influences the acquisition of new knowledge mediated by technologies. To evaluate this, two levels of comparison were applied: 1) an intra-prototype analysis aimed at determining whether each type of content (logical, social, or psychomotor) produced a significant effect on learning within its own group; and 2) an inter-prototype analysis intended to identify whether any type of content generated a differential impact on the magnitude of learning, considering the control and experimental groups separately. Since the observed distributions in each comparison did not conform to a normal model, non-parametric statistical tests were applied, with a significance level set at $p < 0.05$. The descriptive statistics corresponding to these comparisons are summarized in Table 5.

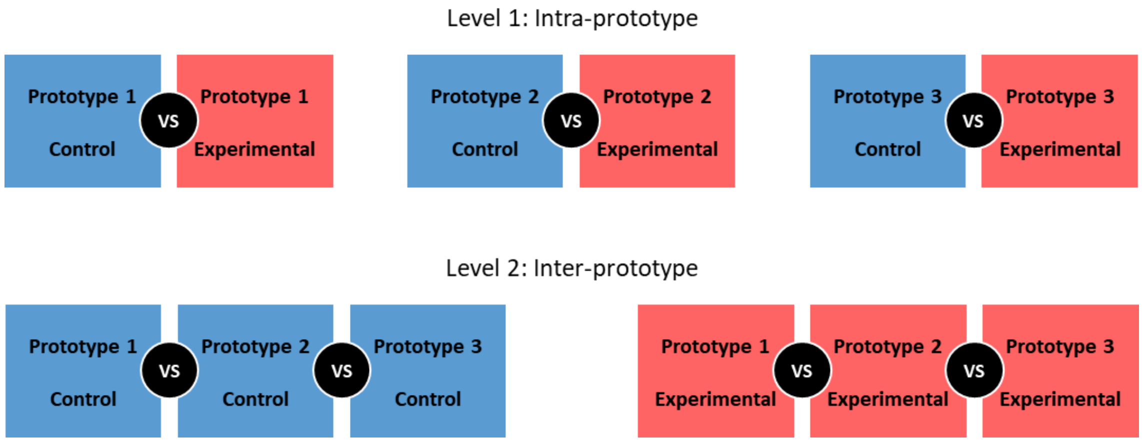


Figure 3: Strategies employed for comparisons between different conditions.

Various non-parametric statistical tests were performed to evaluate differences between and within groups, as well as between prototypes. To compare the initial knowledge level between the experimental and control groups, the Mann-Whitney test was used. Learning within each group was assessed using the Wilcoxon signed-rank test. Differences in learning between groups were also analysed with the Mann-Whitney test. Finally, to compare learning across prototypes, the Kruskal-Wallis test was applied, accompanied by a Bonferroni correction for multiple comparisons, establishing an adjusted significance threshold of $p = 0.0167$.

At the first level of analysis (intra-prototype comparisons), results indicate that the experimental and control groups started from equivalent knowledge levels in Prototype 1 (logical content) and Prototype 3 (psychomotor content). In contrast, in Prototype 2 (social content), the experimental group presented a significantly higher initial knowledge level compared to the control group (Figure 4, left column). Both groups demonstrated significant improvement within their respective intervention contexts ($p < 0.001$), as shown in the central column of Figure 4. However, when comparing the magnitude of learning between groups, no significant differences were found between the experimental and control groups in Prototype 1 (Figure 4, right column, upper row). In Prototype 2, the control group exhibited significantly greater improvement than the experimental group ($p < 0.001$), although this result should be interpreted with caution since the experimental group started with a higher pre-test level (Figure 4, right column, middle row). In Prototype 3, the experimental group showed significantly superior learning compared to the control group ($p < 0.05$), as observed in Figure 4, right column, lower row.

In summary, logical content (Prototype 1) showed no differences between groups in terms of learning, social content (Prototype 2) favoured the control group, and psychomotor content (Prototype 3) favoured the experimental group.

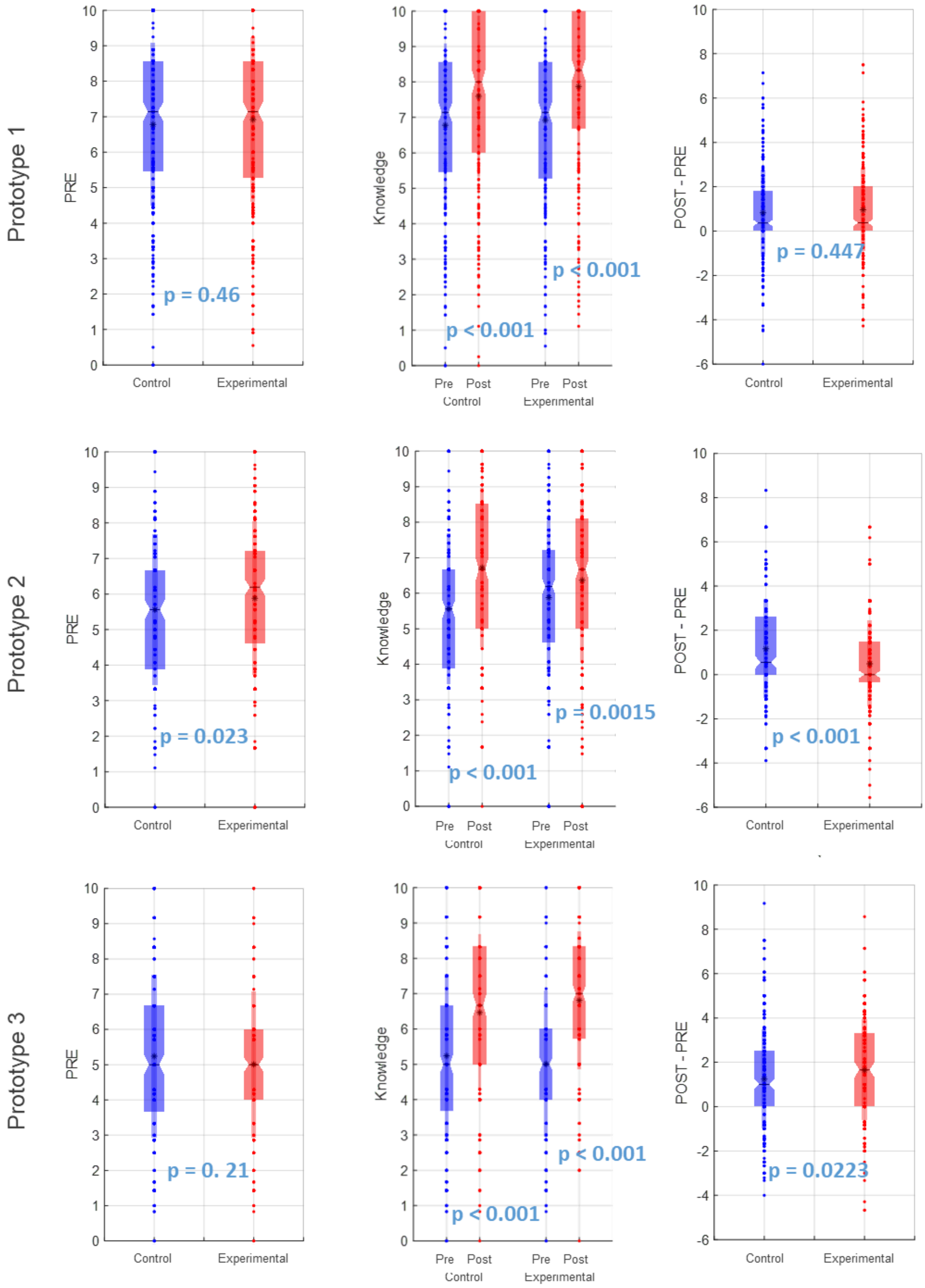


Figure 4: Results of the proposed statistical analysis in intra-prototype comparisons.

At the second level of analysis (inter-prototype comparisons), differences in prior knowledge and learning were evaluated among the three types of content, separately for the control and experimental groups. The results indicate that participants began with significantly different knowledge levels depending on the type of content. In particular, logical content (Prototype 1) showed higher initial levels than the other two content types in both groups, with significant differences ($p < 0.001$) (Figure 5, top row).

Regarding learning, the comparison between prototypes in the control group did not reveal significant differences in the magnitude of learning among the three content types, suggesting that all functioned equivalently (Figure 5, left column, bottom row). In contrast, in the experimental group, significant differences between content types were identified. Both logical content (Prototype 1) and psychomotor content (Prototype 3) produced significantly greater improvements in learning compared to social content (Prototype 2). Additionally, psychomotor content showed a superior learning effect compared to logical content within the experimental group (Figure 5, right column, bottom row).

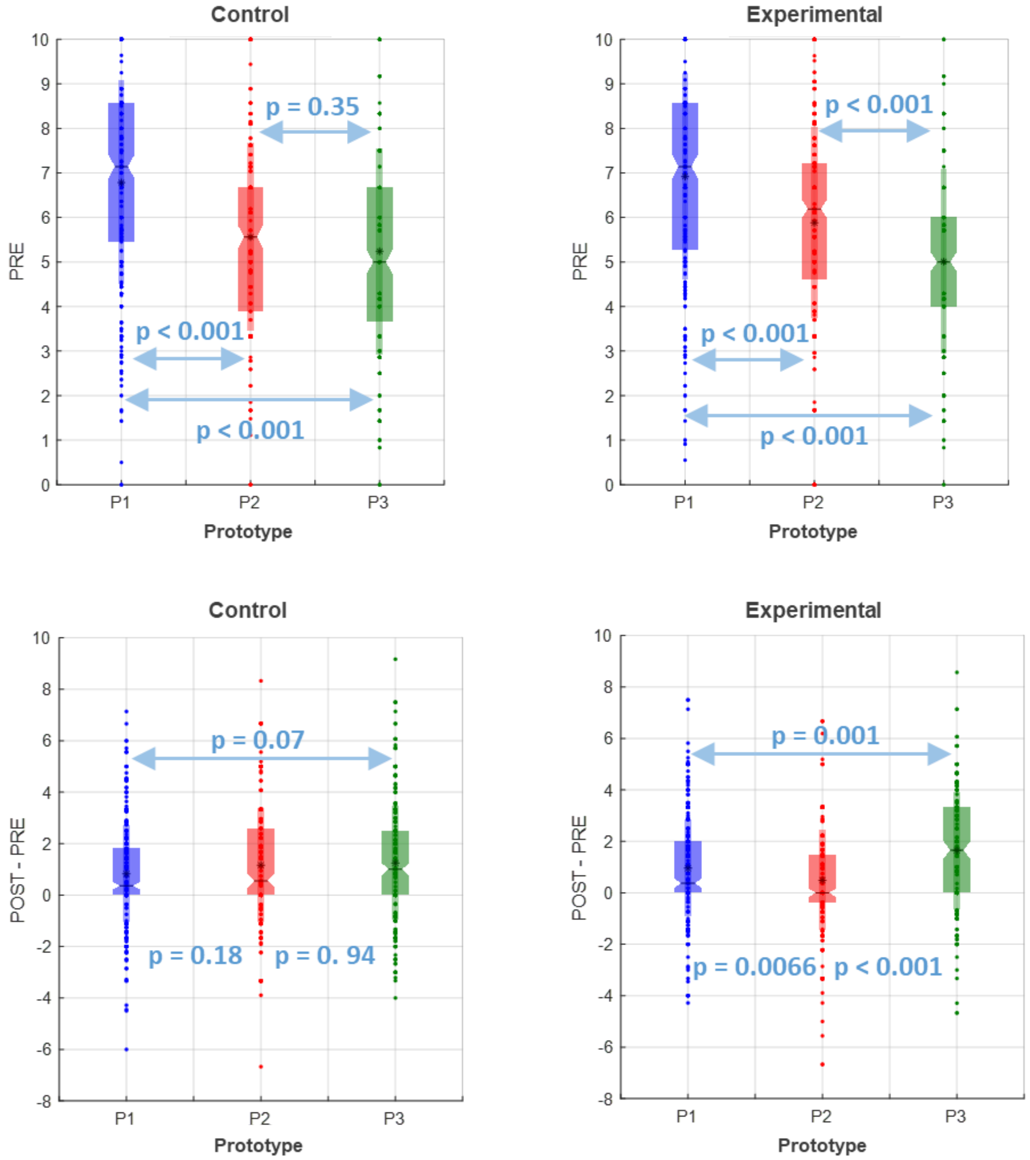


Figure 5: Results of the proposed statistical analysis in inter-prototype comparisons.

10.5 Comparative statistics using normalised gain (NG)

To complement the analysis conducted in Section 4, the same statistical procedure was repeated, substituting the Absolute Difference (AD) metric with the Normalized Gain (NG) metric. This metric quantifies each participant's relative improvement based on their potential margin for progress, considering a maximum scale of 10 points. The results obtained using this variable are presented below.

A. Objective 1: Results and conclusions

The hypothesis evaluated was "learning is more effective using technologies vs. non-technologies". For this purpose, observations from the three prototypes were grouped, classified only according to the group type: control (traditional teaching) and experimental (technology-mediated teaching). The distribution of the NG variable in both groups did not conform to the normal model; therefore, non-parametric tests were employed for the comparison, with a significance level set at $p < 0.05$. Table 5 summarizes the descriptive statistics corresponding to the proposed grouping, including median, value range, and the number of observations per group.

Group	Observations	Pre	Post	NG
		Median [Min Max]	Median [Min Max]	Median [Min Max]
Control	664	6.00 [0 10]	7.14 [0 10]	0.36 [0 1]
Experimental	670	6.11 [0 1]	7.14 [0 1]	0.33 [0 1]

Table 7: Descriptive statistics in large groups using NG. The number of observations in both groups is lower than those in Table 6, due to the exclusion of subjects with NG values < 0 and Pre = 10 from the analysis.

To compare the difference in learning between the experimental and control groups, the Mann-Whitney test was used, as the NG variable did not follow a normal distribution. The results indicated no significant differences in normalized gain between the two groups (Figure 6), suggesting that the use of technologies was not associated with a superior relative improvement in learning compared to traditional teaching. These findings replicate and confirm previous observations using the AD metric, reinforcing the conclusion that, overall, the incorporation of technologies did not demonstrate a differential effect on knowledge improvement.

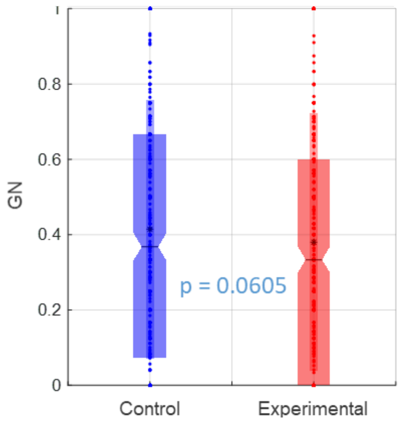


Figure 6: Distribution of NG values and statistical results for the difference in learning between the experimental and control groups.

B. Objective 2: Results and conclusions

A second hypothesis was formulated: The type of content influences the acquisition of new knowledge mediated by technologies. To analyse this, two levels of comparison were applied: 1) an intra-prototype level to determine whether, within each type of content, there were differences in learning between the experimental and control groups; and 2) an inter-prototype level intended to evaluate if any of the content types generated a superior learning effect, analysed separately for each group. Since the data groupings did not meet normality assumptions, non-parametric tests were used with a significance level of $p < 0.05$. The descriptive statistics corresponding to these comparisons are presented in Table 5, previously mentioned.

For comparisons between groups within each prototype (intra-prototype level), the Mann-Whitney test was applied. For comparisons between content types (inter-prototype level), the Kruskal-Wallis test was used, with a Bonferroni correction applied to the significance threshold for multiple comparisons ($p = 0.0167$).

Regarding the intra-prototype results (level 1), in Prototype 1 (logical content), no significant differences in normalized gain were observed between the control and experimental groups (Figure 7, left column). In Prototype 2 (social content), the control group showed a significantly higher gain compared to the experimental group ($p < 0.001$), as shown in the central column. It is noteworthy that the experimental group started from a higher initial knowledge level, which may have influenced this result. In Prototype 3 (psychomotor content), no significant differences were found between groups either (Figure 7, right column).

In summary, social content showed a greater effect in the control group, while logical and psychomotor contents did not show significant differences in learning between conditions.

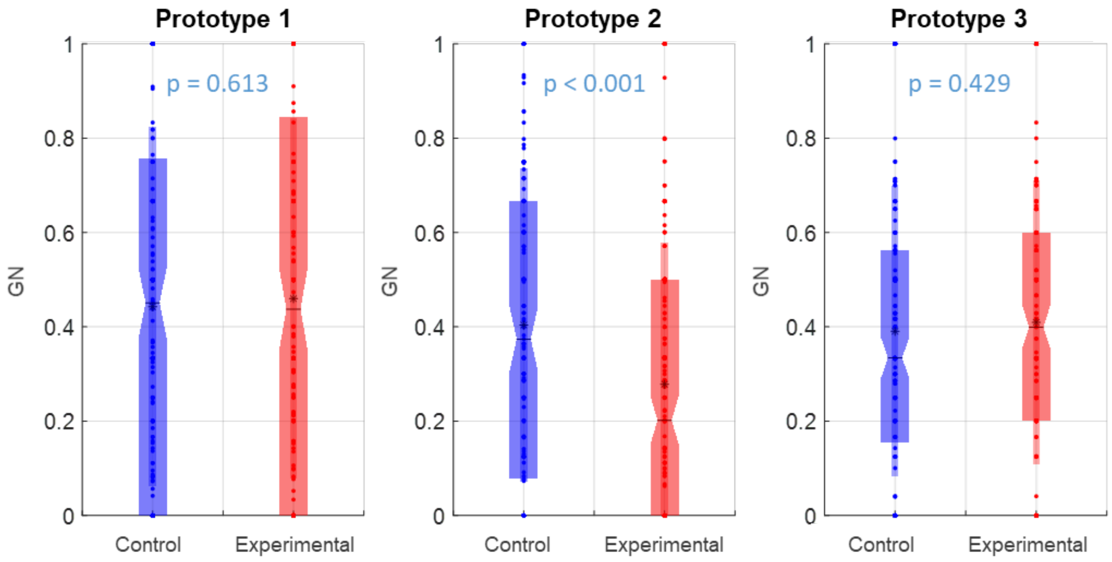


Figure 7: Results of the proposed statistical analysis for intra-prototype comparisons using NG.

In the inter-prototype analysis (level 2), the results showed that no significant differences in NG were observed among the three types of content within the control group. In other words, the prototypes performed equivalently in terms of learning (Figure 8, left column). This finding is consistent with the results previously obtained using the AD metric. Conversely, in the experimental group, significant differences between prototypes were observed. Both logical content (Prototype 1) and psychomotor content (Prototype 3) exhibited significantly higher normalized gains compared to social content (Prototype 2), replicating the pattern previously identified with the AD metric (Figure 8, right column).

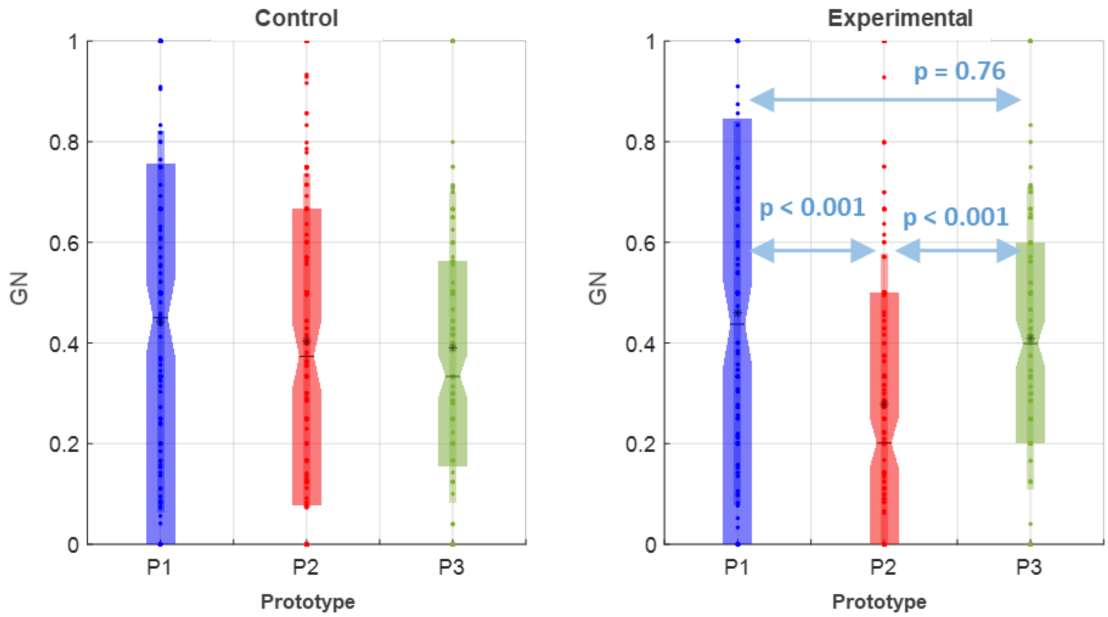


Figure 8: Results of the proposed statistical analysis for inter-prototype comparisons using NG.

10.6 Summary of results

<table>
<tr><th></th><th>Absolute Difference (AD)</th><th></th><th>Normalized Gain (NG)

(only NG ≥ 0)</th></tr>
<tr><td>Objective 1</td><td>Using technology is NOT more effective compared to control, both work equally.</td><td>=</td><td>Using technology is NOT more effective compared to control, both work equally. But almost, Control > Experimental (p=0.0605)</td></tr>
<tr><td rowspan="3">Objective 2

Intra-proto types</td><td>P1 (logical): Works equally well in both groups.</td><td>=</td><td>P1 (logical): Works equally well in both groups.</td></tr>
<tr><td>P2 (social): Works better in the control group. Caution: technology group started with a higher prior level.</td><td>=</td><td>P2 (social): Works better in the control group. Caution: technology group started with a higher prior level.</td></tr>
<tr><td>P3 (psychomotor): Works better in the technology group.</td><td>≠</td><td>P3 (psychomotor): Works equally well in both groups.</td></tr>
<tr><td rowspan="2">Objective 2

Inter-proto types</td><td>In the control group:

All contents work; none better than others.</td><td>=</td><td>In the control group:

All contents work; none better than others.</td></tr>
<tr><td>In the technology group:

P1 (logical) and P3 (psychomotor) work better than P2 (social).

P3 (psychomotor) works better than P1 (logical).</td><td>≠</td><td>In the technology group:

P1 (logical) and P3 (psychomotor) work better than P2 (social).</td></tr>
</table>

Table 8: Summary of results for each objective.

10.7 Association with subjective tests

To explore potential associations between emotional regulation strategies and changes in learning outcomes, the Chi-square test (χ^2) was applied, given that the variables involved are categorical and did not meet normality assumptions. Learning changes were categorized into three groups: 1) Positive or neutral change—when the post-test score was equal to or higher than the pre-test score; 2) Negative change—when the post-test score was lower than the pre-test score; and 3) Expert subject—when the pre-test score was equal to the maximum possible score on the scale (10 points). Emotional regulation strategies were assessed using the Emotion Regulation Questionnaire (ERQ), a validated instrument that identifies individual differences in the habitual use of two strategies: cognitive reappraisal and expressive suppression. Both strategies were classified into three levels: low, medium, and high.

Statistical analyses were conducted separately for each group (control and experimental) and for both ERQ strategies. For each module within each prototype, contingency tables were constructed relating the learning change categories and the corresponding levels of emotional regulation strategy use. Cases lacking information in either variable were excluded. The resulting data were then grouped by participant group (control or experimental), and the findings are summarized in Table 9, which presents the contingency tables alongside the χ^2 test results. No statistically significant associations were found between emotional regulation strategies and learning changes in either the control or experimental groups.

CONTROL n=882

Cognitive reappraisal

	Pos/neutral	Negative	Experts
High	233	53	24
Med	276	60	27
Low	148	41	20

χ^2 = 2.0795, p = 0.7211

Expressive suppression

	Pos/neutral	Negative	Experts
High	277	65	29
Med	185	38	17
Low	195	51	25

χ^2 = 1.8716, p= 0.7594

EXPERIMENTAL n=882

Cognitive reappraisal

	Pos/neutral	Negative	Experts
High	276	58	26
Med	275	57	37
Low	107	32	14

χ^2 = 4.3851, p = 0.3564

Expressive suppression

	Pos/neutral	Negative	Experts
High	261	60	42
Med	212	42	22
Low	185	45	13

χ^2 = 8.1805, p = 0.852

Table 9: Contingency tables and Chi-square test results.

With the aim of identifying possible relationships between changes in subjective states during activities and changes in learning, Spearman's rank correlation coefficient (ρ) was calculated. This non-parametric measure assesses dependence between ordinal or non-normally distributed variables. Subjective states were evaluated using the Self-Assessment Manikin (SAM), a non-verbal pictorial questionnaire that measures an individual's affective responses to an experience. The SAM distinguishes three dimensions of affective state: emotional valence, physiological arousal, and sense of control (dominance). In this study, scores were recorded before and after the activity, and changes were computed as the difference between these two instances, with a possible range from −4 to +4 per dimension. Learning was quantified using the Absolute Difference (AD) metric between pretest and post-test scores. Cases lacking complete data for both variables were excluded from the analysis. Correlations were calculated separately for the control and experimental groups, pooling observations across all modules. Results are presented in Figure 9 (control group) and Figure 10 (experimental group), showing the correlation coefficients obtained for each SAM dimension in relation to learning changes. No significant correlations or consistent trends were observed between changes in subjective states and performance improvements, in any of the prototypes or groups analysed.

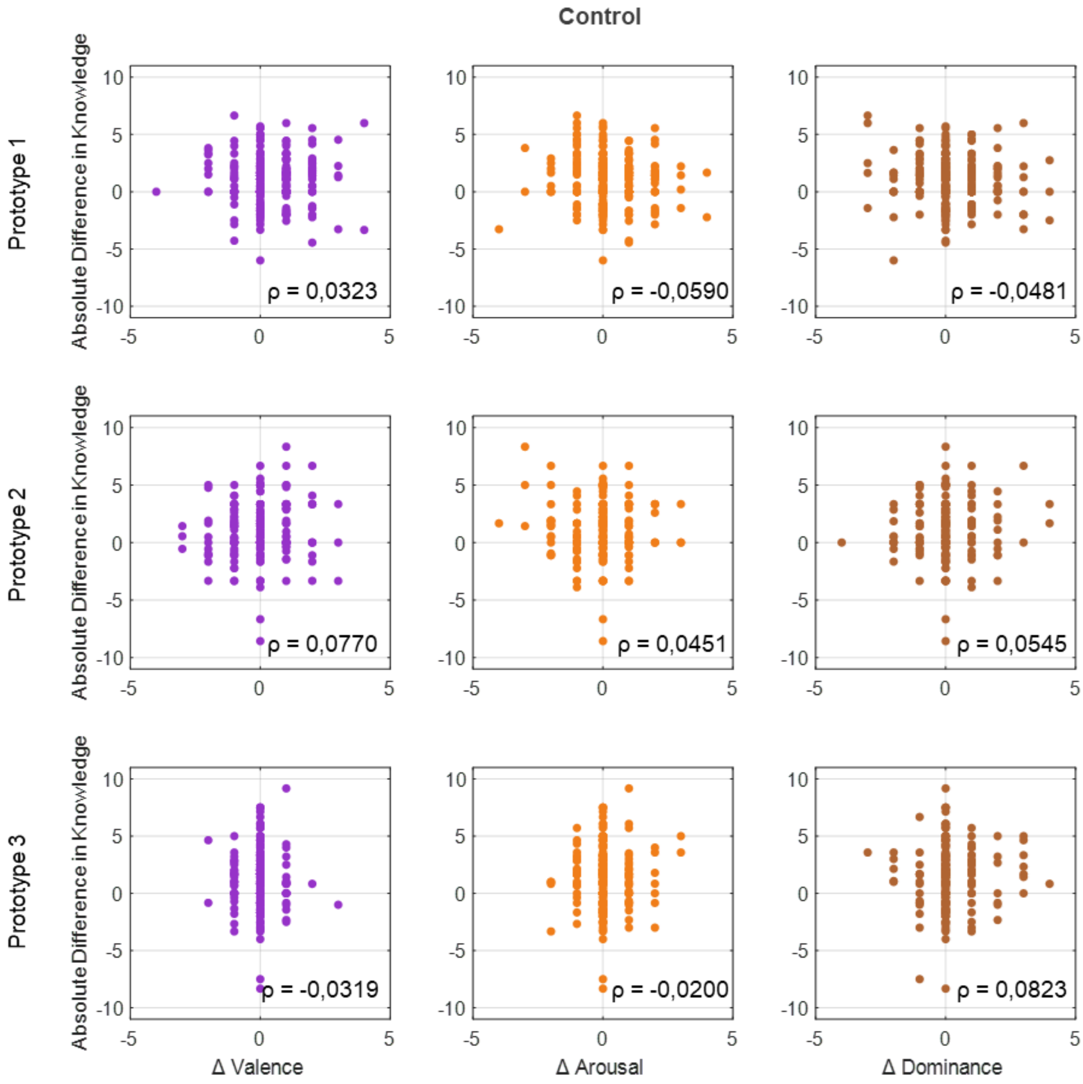


Figure 9: Correlations between changes in subjective states and learning in the control group. Spearman correlation coefficients (ρ) are shown between changes in each of the three SAM dimensions (valence, arousal, and dominance) and the absolute difference in knowledge. Each point represents an individual observation, and correlations are displayed separately for each prototype.

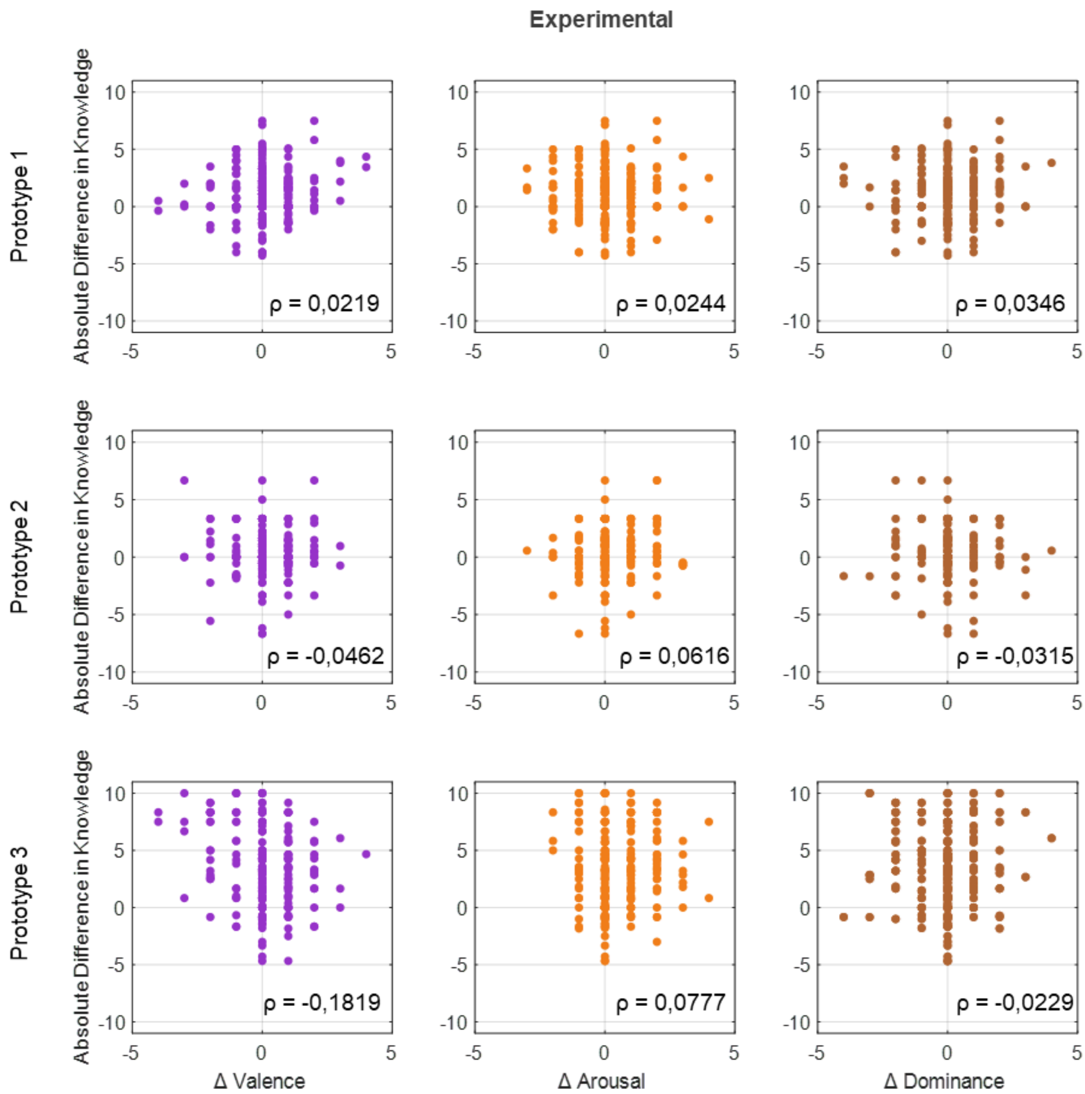


Figure 10: Correlations between changes in subjective states and learning in the experimental group. Spearman correlation coefficients (ρ) are shown between changes in each of the three SAM dimensions (valence, arousal, and dominance) and the absolute difference in knowledge. Each point represents an individual observation, and correlations are displayed separately for each prototype.

10.8 Final conclusions related to objectives 1 and 2

This comparative analysis allowed for the evaluation of the impact of technology-mediated educational interventions on different types of content (logical, social, and psychomotor), using two learning metrics: Absolute Difference (AD) and Normalized Gain (NG). The main considerations and conclusions drawn from the study are detailed below:

A. Methodological Considerations:

- The AD metric includes all observations and reflects both improvements and deteriorations in performance. However, its interpretation must consider the possibility of negative differences associated with decreases in post-test performance.
- The NG metric, while useful for estimating relative improvement with respect to the potential learning margin, requires the exclusion of approximately 65% of the data due to undefined cases (e.g., pretest score = 10) or negative results. The use of this metric should be justified and limited to contexts where its calculation is valid.
- No significant associations were found between emotional regulation strategies (evaluated by the ERQ) or changes in affective states during the activity (evaluated by the SAM) and learning performance. These results suggest that the affective and personality factors assessed do not predict performance in the contexts analysed.
- Since both metrics (AD and NG) yielded generally similar patterns in the results, and AD allowed for maintaining a larger data volume, its preferential use in this analysis is recommended.

B. General Conclusions

Both groups (experimental and control) showed significant learning improvements regardless of technology use. For logical content, no differences were observed between traditional and technology-mediated methodologies. For social content, the control group (without technology) showed superior performance, suggesting that technology use might represent a relative disadvantage in this type of content. Conversely, for psychomotor content, the technology-mediated intervention showed a significant advantage over the classical methodology.

11. Analysis of objective 3: Results and conclusions

Understanding the role of **psychophysiological signals** is crucial for unpacking the cognitive and emotional processes that underlie immersive learning. Within the **Cognitive Affective Model of Immersive Learning (CAMIL)**, these signals provide a direct, continuous, and objective window into how learners react to immersive environments—beyond what they can report in questionnaires or demonstrate in task performance. By tracking bodily responses such as **heart rate (HR)**, **electrodermal activity (EDA)**, and **cognitive load (CL)** in real time, we can infer dynamic states like emotional arousal, sustained attention, or mental fatigue, and relate them to specific moments within the learning experience.

The integration of psychophysiological data into the CAMIL-based evaluation framework allows for **multilevel triangulation**—linking **subjective**, **behavioural**, and **biometric** indicators of engagement and learning. This multimodal approach is particularly relevant in XR-based educational contexts, where the user's embodied interaction with the content often triggers high cognitive and affective variability.

The analysis strategy for these signals follows a **model-driven approach**. Instead of relying solely on raw signal averages, each participant's time series is processed using **linear modelling techniques**, extracting interpretable parameters such as trend, intercept, and magnitude of change (ΔHR/ΔCL). These features allow us to quantify not just the presence of activation, but also its **evolution over time**—for example, identifying whether a user adapts, disengages, or accumulates cognitive load across a session.

Furthermore, by combining these physiological markers with **psychometric profiles** (e.g., self-efficacy, emotional regulation) and **task performance metrics**, we can explore **inter-individual differences** in learning strategies and outcomes. This enables a nuanced understanding of how immersive educational systems may differentially benefit users depending on their baseline characteristics—informing future personalization and adaptive feedback mechanisms.

The following sections present the results of this analysis for **Prototypes 1, 2 and 3**, illustrating how psychophysiological signals were processed, modelled, and interpreted in alignment with the methodological principles outlined above.

11.1 Prototype 1

A. **Methods.**

- ***CSV data recordings.***

The physiological data analysed in this study originate from recordings of variables such as heart rate (HR), eye tracking (ET), and cognitive load (CL), acquired through sensing devices during the execution of various experimental tasks. The data are stored in comma-separated text files (CSV), with one file per physiological variable and experimental session. These files are organized within a hierarchical folder structure, following the scheme shown in Figure 11.

The countries involved in data collection were Cyprus, Hungary, and Spain. Despite the number of participants included in each group (N = 20, 28, and 24, respectively), in some cases, records contained no stored data, while in others, the records were entirely missing. The EMOTIBIT subfolder contains data collected during the execution of specific tasks using an EMOTIBIT acquisition system. Module 4 does not include CSV-format data; therefore, its analysis was based exclusively on the data obtained through EMOTIBIT.

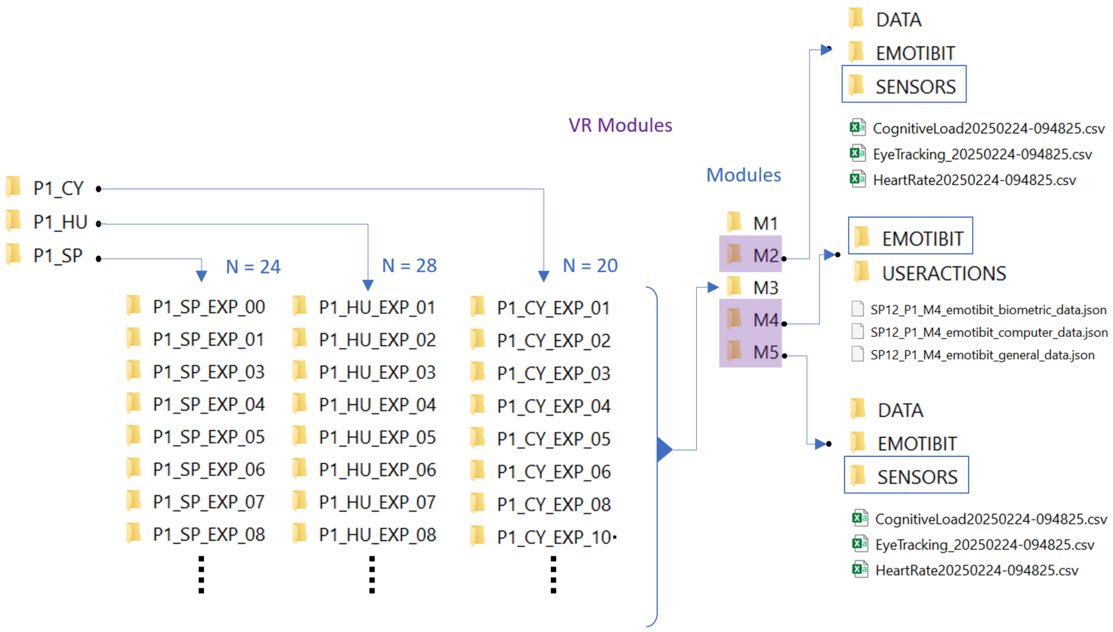


Figure 11: Organizational structure of the data acquired during the sessions of Protocol 1. The countries involved in data collection are Cyprus, Hungary, and Spain. Only the modules involving virtual reality were analysed (M2, M4, and M5).

The physiological data contained in the CSV files are organized in columns. One of the main columns is unixTimestamp, which represents the date and time in numerical format (Unix timestamp), derived from the original format yyyy-MM-dd HH:mm:ss.SSSSSS. This field precisely indicates the moment at which each measurement was recorded.

In the case of Heart Rate data, the corresponding values are located in column 4. Similarly, for Cognitive Load data, the primary values are also stored in column 4. Additionally, in these files, an estimate of the measurement error or deviation of the cognitive load is specified in column 5.

- ***Availability of physiological data.***

In the MATLAB environment, the physiological data were initially organized into a cell array of dimensions N×3. In this array, the first column contained the subject identifiers, while the second and third columns stored structures (struct data) corresponding to sessions M2 and M5, respectively (M4 did not have data stored in CSV format). Each of these structures has two main fields: 1) HeartRate and 2) CognitiveLoad. Both fields are cell arrays containing matrices of size N×2, where the first column corresponds to the timestamp (expressed as Unix timestamp) and the second column contains the value of the corresponding physiological measure (heart rate or cognitive load).

Subsequently, to facilitate analyses, these data were concatenated by subject and physiological variable. As a result, two new cell arrays were generated: HeartRate and CognitiveLoad, which contained the heart rate and cognitive load information, respectively. Both matrices have the following structure: column 1 contains the subject identifier; column 2 contains a numeric matrix of dimensions N×2 with all concatenated HR or CL records from session M2; and finally, column 3 contains the information from session M5. Figure 12 schematically illustrates this structure.

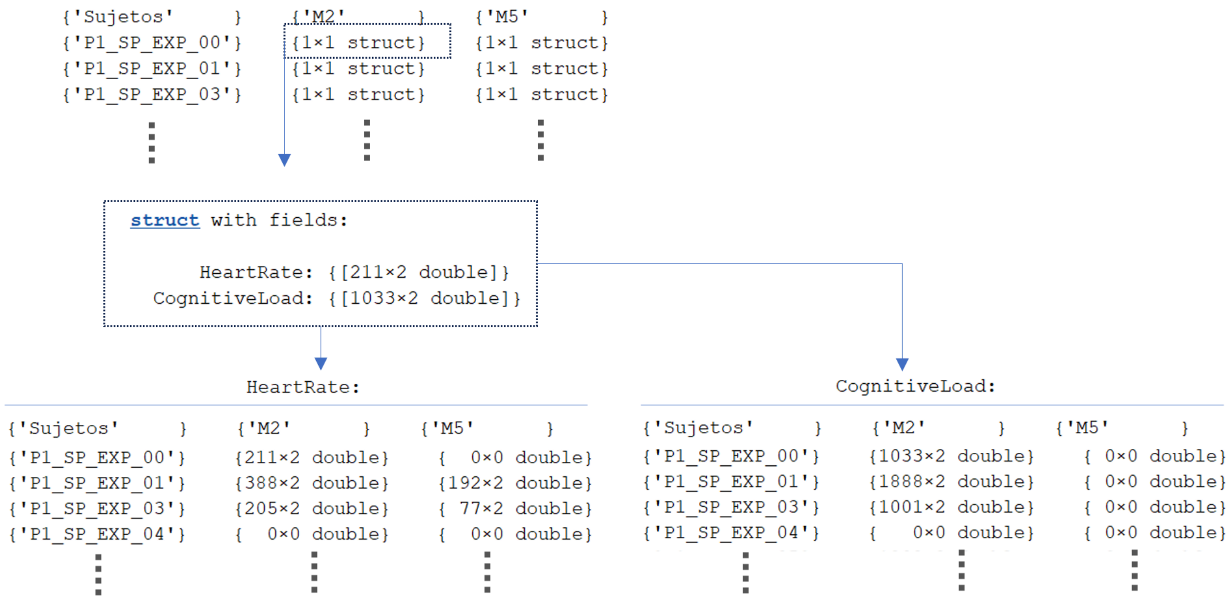


Figure 12: Structural data organization. This scheme was implemented in the MATLAB environment using specific data reading and loading functions.

In Figure 12, it can be observed that some elements of the matrices comprising the HR and CL data contain empty matrices. These cases correspond to missing data and, therefore, were not considered in subsequent analyses. Table 10 shows the available data for the analysis of physiological profiles in Prototype 1.

Spain	**Hungary**	**Cyprus**
Heart Rate	*Heart Rate*	*Heart Rate*
N = 17 (M2) N = 8 (M5)	N = 16 (M2) N = 9 (M5)	N = 7 (M2) N = 8 (M5)

Cognitive Load		Cognitive Load		Cognitive Load	
N = 15 (M2)	N = 1 (M5)	N = 25 (M2)	N = 14 (M5)	N = 0 (M2)	N = 0 (M5)

Table 10: Available data for analysis in Prototype 1.

- ***Linear modelling of heart rate dynamics during task execution.***

To characterize the physiological profile of each subject based on heart rate (HR) recordings obtained during the execution of the experimental tasks, a simple linear regression model was applied to the temporal evolution of the HR signal. This approach allowed the extraction of a set of relevant features describing the physiological dynamics of each individual over time. The main parameters extracted include:

- **HR Trend (slope)**: Represents the slope coefficient of the fitted line and indicates whether the HR increased, decreased, or remained stable throughout the task. This trend is interpreted as an indicator of sustained physiological activation or autonomous adaptation processes in response to task demands.
- **Initial HR (intercept)**: Corresponds to the estimated HR value at the start of the task. This parameter serves as an approximation of the subject's basal physiological state at task onset.
- **Coefficient of determination (R^2):** Evaluates the quality of the linear fit. A high R^2 value indicates that HR variability can be consistently explained by a linear trend, whereas low values reflect greater momentary variability or more complex patterns not captured by the linear model.
- **p-value associated with the slope**: Assesses the statistical significance of the HR change over time. Although a low p-value suggests that the observed trend is statistically significant, its interpretation should be approached cautiously given the inherent complexity of physiological signals.
- **Final HR**: The HR value at the end of the task, representing the subject's final physiological state.
- **ΔHR (Final HR – Initial HR):** This metric summarizes the total change in heart rate during task execution, providing a measure of the overall degree of physiological activation or modulation induced by the activity.

Figure 13 graphically presents the parameters obtained from the linear regression analysis.

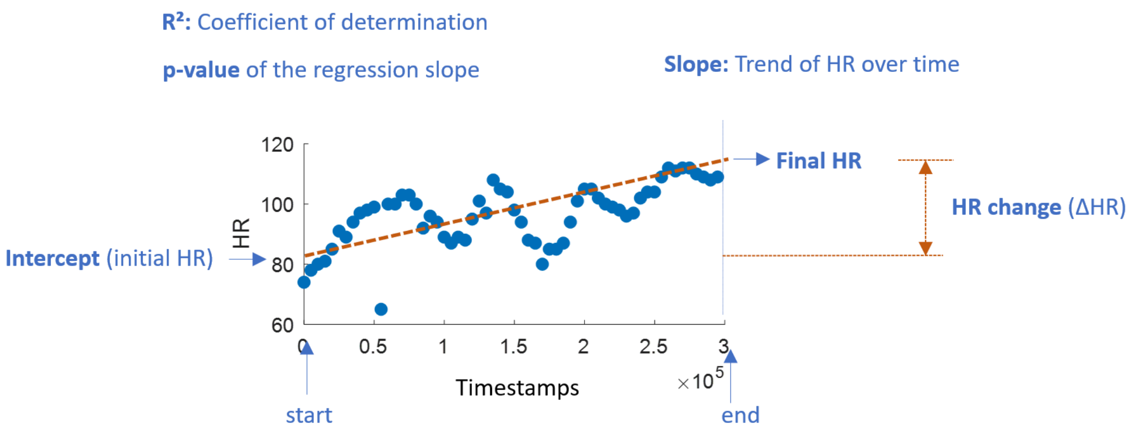


Figure 13: Time series of HR obtained from an experimental subject during task execution (M2). The orange dotted line represents the result of a linear trend fit.

- ***Linear modelling of cognitive load dynamics during task execution.***

Analogous to the heart rate analysis, the physiological profile of each subject was characterized using the cognitive load (CL) recordings obtained during task execution. A simple linear regression model was applied to the temporal evolution of the CL series, allowing extraction of metrics that describe the dynamics of sustained cognitive effort. The parameters extracted from the model included:

- **CL trend (slope)**: Indicates the direction of change in cognitive load over time (increase, decrease, or stability). This value is interpreted as an estimate of the progressive change in cognitive demand or the subject's adaptation to the task.
- **Initial CL value (intercept)**: Represents the estimated cognitive load at the start of the task, serving as an approximation of the basal state or starting point of the subject's cognitive effort.
- **Coefficient of determination (R^2)**: Evaluates how well the linear model fits the temporal evolution of CL. High R^2 values indicate a more consistent physiological response, while low values suggest greater variability or nonlinear responses throughout the task.
- **p-value associated with the slope**: Allows determination of whether the observed trend is statistically significant. As in the HR analysis, this value should be interpreted with caution since cognitive responses may present complex and not necessarily linear fluctuations.
- **Final CL value**: Reflects the estimated cognitive state of the subject at the end of the task, which may relate to the degree of fatigue or residual load following task completion.
- **ΔCL (final CL – initial CL)**: This difference quantifies the total change in cognitive load during the task, providing a measure of cumulative cognitive effort or adaptation to the execution context.

- ***Associations between HR, CL, and subjective measures.***

This analysis aimed to explore whether the physiological parameters recorded during task execution are related to different psychological profiles and levels of perceived cognitive load, assessed through standardized instruments. The psychological profiles were established based on the following questionnaires:

- ERQ (Emotion Regulation Questionnaire): separately considering the subscales of cognitive reappraisal (ERQcog) and emotional suppression (ERQemo).
- CEI (Curiosity and Exploration Inventory): measures curiosity and openness to experience.
- LOC (Locus of Control): assesses the perception of internal or external control over events.
- GSE (General Self-Efficacy Scale): estimates the perceived level of general self-efficacy.
- NASA-TLX: measures the perceived cognitive load during the task across multiple dimensions.

The physiological parameters derived from linear regression models applied to the HR and CL time series were compared between subgroups defined by questionnaire scores (e.g., high vs. low ERQcog). This procedure allowed exploration of the validity of HR and CL as potential physiological markers sensitive to psychological variations in experimental interaction contexts.

- ***Multivariate pattern analysis of HR and CL via PCA.***

In addition to the aforementioned univariate analyses, a principal component analysis (PCA) was applied to the physiological variables extracted from the heart rate and cognitive load time series, with the aim of exploring multivariate patterns potentially associated with individual differences in psychological profiles or self-reported cognitive load levels.

For this purpose, a data matrix was constructed where each row represented a participant, and each column corresponded to one of the physiological features derived from the linear models. The physiological variables were analysed separately for HR and CL. Since the variables within each domain shared a common scale (e.g., beats per minute for HR), normalization of the data prior to PCA was not required.

The projection of participants into the reduced space defined by the first principal components (e.g., PC1 and PC2) allowed visual examination of potential clustering or separation among individuals based on their psychological profiles or perceived load levels. This multivariate approach provided a complementary perspective, capable of capturing complex relationships and patterns that may not be evident in univariate analyses.

Funded by
the European Union

To assess group differences without assuming data normality, a permutation-based multivariate analysis of variance (PERMANOVA) was conducted on the principal components derived from the HR PCA (PC1 and PC2). This technique enables comparison of multivariate structures between groups defined by psychological profiles or perceived load levels, offering a robust alternative to traditional MANOVA in contexts with moderate sample sizes and potential violations of parametric assumptions. This analysis was applied independently to the principal components of both HR and CL.

B. Results

- ***Univariate tests.***

The results of the univariate comparisons using the Mann-Whitney U test, employed to evaluate whether the psychological profiles and self-reported cognitive loads can be described by the recorded physiological variables, are presented in the following tables. Table 11 summarizes the univariate comparisons corresponding to the heart rate physiological variables during the execution of tasks M2 and M5. Meanwhile, Table 12 shows the equivalent results for the cognitive load variables assessed in the same tasks. Statistically significant differences are highlighted in blue.

HR parameters vs self-reports					
ERQ_reev_cog	***ERQ_sup_emo***	***CEI***	***LOC***	***GSE***	***NASA***
M2					
Slope: p = 0.85 Intercept: p = 0.24 R2: p = 0.62 pValue: p = 0.39 HR_final: p = 0.22 HR_variation: p = 0.97	Slope: p = 0.95 Intercept: p = 0.26 R2: p = 0.50 pValue: p = 0.55 HR_final: p = 0.26 HR_variation: p = 0.89	Slope: p = 0.91 Intercept: p = 0.30 R2: p = 0.54 pValue: p = 0.51 HR_final: p = 0.28 HR_variation: p = 0.57	Slope: p = 0.69 Intercept: p = 0.37 R2: p = 0.55 pValue: p = 0.83 HR_final: p = 0.39 HR_variation: p = 0.80	Slope: p = 0.06 Intercept: p = 0.93 R2: p = 0.89 pValue: p = 0.89 HR_final: p = 0.40 HR_variation: p = 0.10	Slope: p = 0.19 Intercept: p = 0.44 R2: p = 1.00 pValue: p = 0.74 HR_final: p = 0.32 HR_variation: p = 0.12
M5					
Slope: p = 0.2268 Intercept: p = 0.39 R2: p = 0.01 pValue: p = 0.01 HR_final: p = 0.77 HR_variation: p = 0.20	Slope: p = 0.26 Intercept: p = 0.37 R2: p = 0.90 pValue: p = 0.37 HR_final: p = 0.81 HR_variation: p = 0.14	Slope: p = 0.88 Intercept: p = 0.31 R2: p = 0.56 pValue: p = 0.66 HR_final: p = 0.66 HR_variation: p = 0.47	Slope: p = 0.9769 Intercept: p = 0.43 R2: p = 0.12 pValue: p = 0.21 HR_final: p = 0.74 HR_variation: p = 0.79	Slope: p = 0.2860 Intercept: p = 0.73 R2: p = 0.94 pValue: p = 0.64 HR_final: p = 0.31 HR_variation: p = 0.14	Slope: p = 0.9168 Intercept: p = 0.91 R2: p = 0.40 pValue: p = 0.67 HR_final: p = 0.67 HR_variation: p = 1.00

Table 11: Univariate comparisons (Mann-Whitney U) of physiological variables (Heart Rate) between psychological groups during tasks M2 and M5.

CL parameters vs self-reports					
ERQ_reev_cog	***ERQ_sup_emo***	***CEI***	***LOC***	***GSE***	***NASA***
M2					
Slope: p = 0.98 Intercept: p = 0.42 R2: p = 0.79 pValue: p = 0.31 CL_final: p = 0.28 CL_variation: p = 0.57	Slope: p = 0.36 Intercept: p = 0.94 R2: p = 0.58 pValue: p = 1.00 CL_final: p = 0.12 CL_variation: p = 0.18	Slope: p = 0.95 Intercept: p = 0.71 R2: p = 0.62 pValue: p = 0.57 CL_final: p = 0.65 CL_variation: p = 0.95	Slope: p = 0.47 Intercept: p = 0.54 R2: p = 0.61 pValue: p = 0.76 CL_final: p = 0.14 CL_variation: p = 0.41	Slope: p = 0.49 Intercept: p = 0.83 R2: p = 0.51 pValue: p = 0.65 CL_final: p = 0.32 CL_variation: p = 0.39	Slope: p = - Intercept: p = - R2: p = - pValue: p = - CL_final: p = - CL_variation: p = -
M5					
Slope: p = 0.69 Intercept: p = 1.00 R2: p = 0.28	Slope: p = 0.07 Intercept: p = 0.23 R2: p = 0.44	Slope: p = - Intercept: p = - R2: p = -	Slope: p = 0.83 Intercept: p = 0.62 R2: p = 0.73	Slope: p = 0.93 Intercept: p = 0.46 R2: p = 0.37	Slope: p = - Intercept: p = - R2: p = -

pValue: p = 0.81 CL_final: p = 0.93 CL_variation: p = 0.81	pValue: p = 0.36 CL_final: p = 0.18 CL_variation: p = 0.05	pValue: p = - CL_final: p = - CL_variation: p = -	pValue: p = 0.62 CL_final: p = 0.44 CL_variation: p = 0.94	pValue: p = 0.69 CL_final: p = 0.46 CL_variation: p = 0.81	pValue: p = - CL_final: p = - CL_variation: p = -

Table 12: Univariate comparisons (Mann-Whitney U) of physiological variables (Cognitive Load) between psychological groups during tasks M2 and M5.

- ***Multivariate pattern analysis of HR and CL via PCA.***

For tasks M2 and M5, an independent principal component analysis (PCA) was performed on physiological features derived from heart rate (HR) and cognitive load (CL). In none of the cases did the first two principal components (PC1 and PC2) clearly differentiate the psychological profiles evaluated in this prototype. Specifically, the projection of subjects in the PC1-PC2 space showed no consistent clustering or significant separations between profiles, suggesting that when analysed separately, the physiological dynamics of HR and CL do not capture distinctive patterns associated with psychological group differences in these tasks.

C. Methods with EMOTIBIT

- ***EMOTIBIT data recordings.***

As an alternative for monitoring and obtaining physiological variables during the execution of tasks corresponding to Prototype 1, the EMOTIBIT system was used. This system recorded both movement variables, such as motion signals (IMU: accelerometer, gyroscope, and magnetometer), as well as physiological signals (heart rate, electrodermal activity, electrodermal level, and skin temperature, in addition to PPG signals at different wavelengths: infrared, green, and red). All these data were stored in JSON format files. Specifically, for this study, the physiological variables considered were heart rate (HR), electrodermal activity (EA), and skin temperature (TH).

Each JSON file contains records with timestamps in ISO 8601 format (yyyy-MM-ddTHH:mm:ss.SSSZ), which were converted to numeric Unix timestamp format to facilitate processing. The physiological values corresponding to HR, EA, and TH were extracted from specific fields within the JSON structure, ensuring accurate temporal alignment among them.

For organization and analysis in MATLAB, the data were initially structured into cell arrays with dimensions N×3, where the first column contained unique subject identifiers, and the second and third columns stored numerical matrices corresponding to each physiological variable (HR, EA, or TH) for different sessions or experimental blocks.

Each of these matrices contained records organized in two columns: the first with timestamps in Unix format and the second with the numerical values of the corresponding physiological variable. Subsequently, these data were concatenated by subject and physiological variable, generating final matrices where the first column contained the subject identifier, and the following columns grouped the temporal records and physiological values corresponding to each session. This structuring allowed for integrated and homogeneous statistical and multivariate analyses of physiological data.

Table 13 shows the data available for the analysis of the physiological profiles of Prototype 1 based on EMOTIBIT.

Spain			**Hungary**			**Cyprus**		
Heart Rate			*Heart Rate*			*Heart Rate*		
8 (M2)	17 (M4)	14 (M5)	4 (M2)	0 (M4)	0 (M5)	18 (M2)	18 (M4)	19 (M5)
Epidermal activity			*Epidermal activity*			*Epidermal activity*		
8 (M2)	17 (M4)	14 (M5)	4 (M2)	0 (M4)	0 (M5)	18 (M2)	18 (M4)	19 (M5)
Skin temperature			*Skin temperature*			*Skin temperature*		
8 (M2)	16 (M4)	13 (M5)	4 (M2)	0 (M4)	0 (M5)	18 (M2)	18 (M4)	19 (M5)

Table 13: Data available for analysis.

- ***Multivariate pattern analysis of HR, EA and TH via PCA.***

Based on the data distribution presented in Table 2, it was possible to organize the physiological variables corresponding to heart rate (HR), electrodermal activity (EA), and skin temperature (TH) into a common set of features. This allowed for an integrated multivariate analysis, simultaneously considering all physiological measures recorded during the execution of the experimental tasks.

For each of the three physiological variables, four descriptive features reflecting the signal dynamics were defined: the initial value, the final value, the slope of change throughout the task, and the delta (difference between final and initial values). For example, for the heart rate signal, HR-initial, HR-final, HR-slope, and HR-delta were calculated. The same procedure was applied to EA and TH, resulting in a total of 12 features per subject.

Since these features are expressed in different units and absolute scales, a z-score normalization was applied prior to conducting the principal component analysis (PCA). This normalization helped reduce bias due to scale differences and facilitated comparison between subjects. An illustrative diagram of this multivariate physiological analysis strategy is presented in Figure 14.

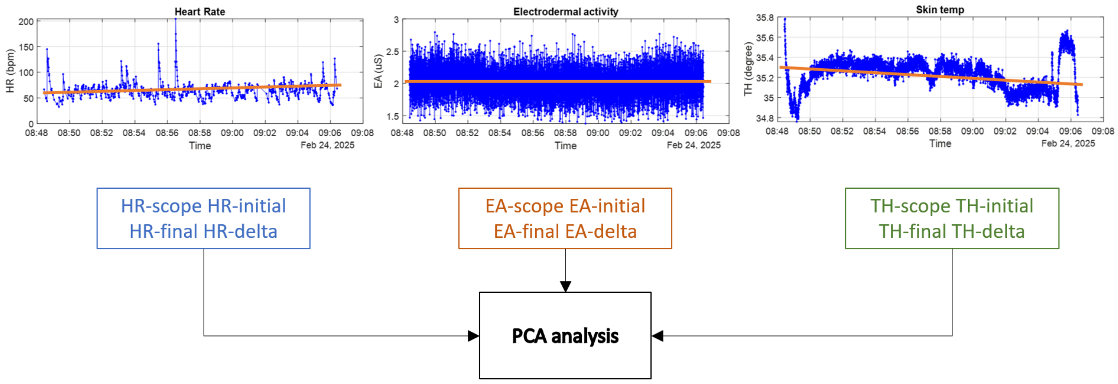
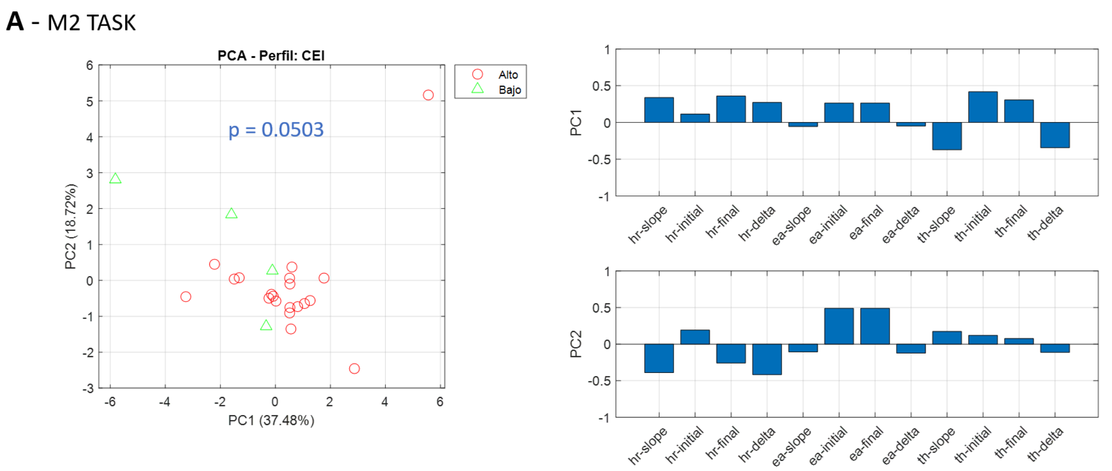


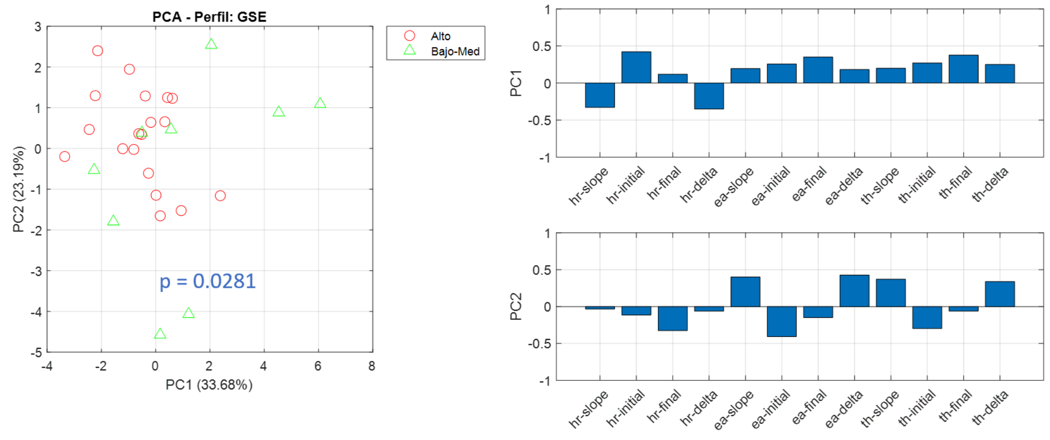
Figure 14: Diagram of the multivariate physiological pattern analysis.

D. Results

Figure 15 presents the PCA results, showing that the reported psychological profiles and cognitive load are statistically described by the physiological profiles recorded during tasks involving virtual reality technologies. The contribution of each physiological feature to the principal components (PC1 and PC2) is displayed in the bar charts on the right, allowing identification of which variables most influence the group separation observed in the PC1-PC2 space. This representation also highlights the clustering of experimental subjects based on their physiological responses.



B – M4 TASK



C – M5 TASK

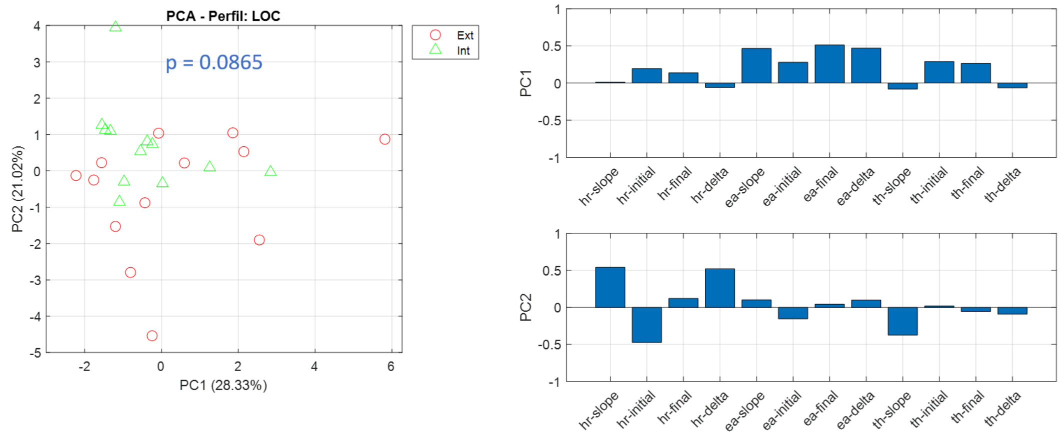


Figure 15: Results of the PCA, showing that the reported psychological profiles and cognitive load are statistically described by the physiological profiles recorded during task execution. A) Task M2 – CEI profile. B) Task M4 – GSE profile. C) Task M5 – LOC profile.

11.2 Prototype 2

A. Methods

- ***JSON data recordings***

The physiological data analysed in this prototype are stored in separate files located in the folder named OMNICEPT. These files contain recordings of physiological variables such as heart rate (HR), eye tracking (ET), and cognitive load (CL), acquired using a specific device during the execution of various experimental tasks. Each physiological variable and experimental session is recorded in a separate JSON file. These files are organized within a hierarchical folder structure, as shown in Figure 16.

Data were collected in three countries: Italy, Hungary, and Estonia. Although N = 25, 23, and 24 experimental subjects were included in each country, respectively, cases were identified where files contained no valid data, and others where the OMNICEPT subfolder was missing. Additionally, the

EMOTIBIT subfolder contains records obtained during certain tasks using the EMOTIBIT physiological acquisition system. It is noteworthy that experimental subjects from Hungary did not have recordings in either OMNICEPT or EMOTIBIT. For this analysis, only modules M1 and M2 were considered, as these were the only ones implementing virtual reality technologies.

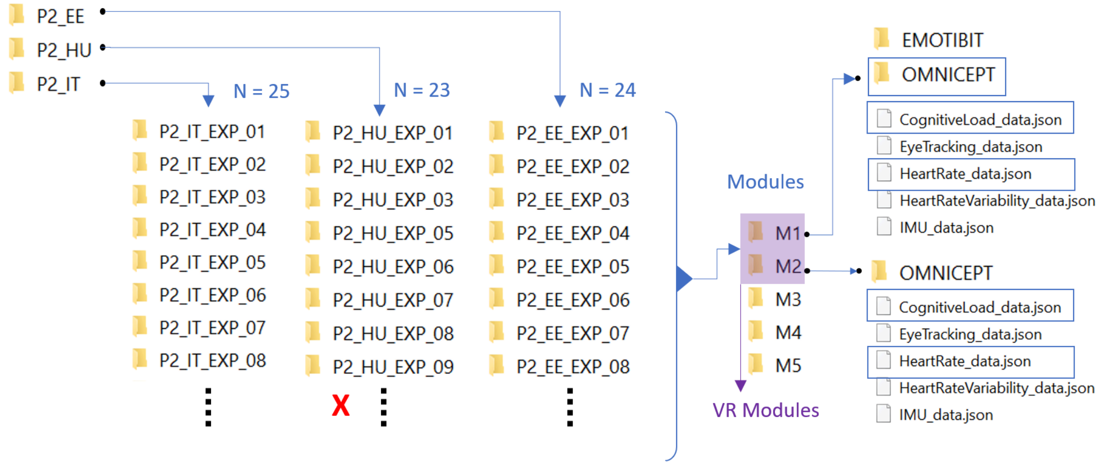


Figure 16: Organizational structure of the data acquired during the sessions of Protocol 2. The countries involved in data collection are Italy, Hungary, and Estonia. Only the modules involving virtual reality were analysed (M1 and M2).

The physiological data analysed are contained in JSON format files, one per physiological variable and experimental session. Each JSON file consists of a series of objects representing individual records, organized as a list of dictionaries. One of the main keys is unixTimestamp, which indicates the exact moment each measurement was recorded, expressed as a number in Unix timestamp format, derived from the original format yyyy-MM-dd HH:mm:ss.SSSSSS.

For Heart Rate data, the heart rate value is typically found under the key "HeartRate," while for Cognitive Load data, the estimated cognitive load value is stored under the key "CognitiveLoadValue." Additionally, the cognitive load files include an estimate of the measurement uncertainty, recorded as "CognitiveLoadStd," representing the standard deviation of the estimated value.

During the data loading and analysis process, various issues related to the JSON file format were detected. Several files presented incorrect structures, such as leading commas ([{,...), missing closing brackets (]), or improperly placed brackets within the content. These inconsistencies prevented automatic decoding and caused read errors. To resolve this, each malformed JSON file was manually inspected and corrected to ensure compliance with the standard JSON format before processing.

- ***Availability of physiological data***

In the MATLAB environment, the physiological data were initially organized in a cell array with dimensions N×3. In this array, the first column contained the subject identifiers, while the second and third columns stored structures (data structs) corresponding to sessions M1 and M2, respectively.

The data organization in MATLAB followed the same structural scheme used in Prototype 1. Therefore, the data layout corresponds to that illustrated in Figure 12, applied here specifically to tasks M1 and M2. Table 14 shows the data available for the analysis of physiological profiles in Prototype 2.

Italy		**Hungary**		**Estonia**	
Heart Rate		*Heart Rate*		*Heart Rate*	
N = 5 (M1)	N = 6 (M2)	N = 0 (M1)	N = 0 (M2)	N = 19 (M1)	N = 20 (M2)
Cognitive Load		*Cognitive Load*		*Cognitive Load*	
N = 25 (M1)	N = 25 (M2)	N = 0 (M1)	N = 0 (M2)	N = 24 (M1)	N = 24 (M2)

Table 14: Data available for analysis in Prototype 2.

- ***Analyses Applied in Prototype 2***

The analyses conducted on the Prototype 2 data replicate exactly the procedures described for Prototype 1. Therefore, to avoid unnecessary repetition, the methods will not be detailed again in this section. Specifically, the following analyses were performed, following the same methodological specifications previously described:

- *Linear modelling of heart rate (HR) dynamics during task execution.*
- *Linear modelling of cognitive load (CL) dynamics during task execution.*
- *Analysis of associations between HR, CL, and subjective measures (self-reported psychological profiles).*
- *Multivariate Pattern Analysis (MPA) of HR and CL via Principal Component Analysis (PCA).*

Each of these analyses was applied identically to those established for Prototype 1, both in terms of task segmentation and statistical procedures used.

B. **Results**

- ***Univariate Tests***

Univariate comparison results using the Mann-Whitney U test, employed to evaluate whether self-reported psychological profiles and cognitive loads can be described by the recorded physiological variables, are presented in the following tables. Table 15 summarizes the univariate comparisons corresponding to the heart rate physiological variables during the execution of tasks M1 and M2. Table 16 shows the equivalent results for the cognitive load variables evaluated in the same tasks. Statistically significant differences are highlighted in blue.

HR parameters vs self reports					
ERQ_reev_cog	***ERQ_sup_emo***	***CEI***	***LOC***	***GSE***	***NASA***
M1					
Slope: p = 0.75 Intercept: p = 0.39 R2: p = 0.13 pValue: p = 0.39 HR_final: p = 1.00 HR_variation: p = 0.39	Slope: p = 0.96 Intercept: p = 0.29 R2: p = 0.84 pValue: p = 1.00 HR_final: p = 0.46 HR_variation: p = 0.84	Slope: p = 0.75 Intercept: p = 0.75 R2: p = 0.57 pValue: p = 0.94 HR_final: p = 0.34 HR_variation: p = 0.57	Slope: p = 0.47 Intercept: p = 0.34 R2: p = 0.05 pValue: p = 0.03 HR_final: p = 0.96 HR_variation: p = 0.57	Slope: p = 0.94 Intercept: p = 0.94 R2: p = 0.48 pValue: p = 1.00 HR_final: p = 0.94 HR_variation: p = 1.00	Slope: p = 0.34 Intercept: p = 0.03 R2: p = 0.08 pValue: p = 0.18 HR_final: p = 0.47 HR_variation: p = 0.24
M2					
Slope: p = 0.57 Intercept: p = 0.52 R2: p = 0.73 pValue: p = 0.43 HR_final: p = 0.97 HR_variation: p = 0.68	Slope: p = 0.23 Intercept: p = 0.59 R2: p = 0.42 pValue: p = 0.38 HR_final: p = 0.97 HR_variation: p = 0.21	Slope: p = 0.01 Intercept: p = 0.43 R2: p = 0.16 pValue: p = 0.47 HR_final: p = 0.43 HR_variation: p = 0.01	Slope: p = 0.76 Intercept: p = 0.64 R2: p = 0.19 pValue: p = 0.31 HR_final: p = 0.60 HR_variation: p = 0.80	Slope: p = 0.12 Intercept: p = 0.73 R2: p = 0.31 pValue: p = 0.79 HR_final: p = 0.73 HR_variation: p = 0.12	Slope: p = 0.10 Intercept: p = 0.16 R2: p = 0.68 pValue: p = 0.39 HR_final: p = 0.39 HR_variation: p = 0.13

Table 15: Univariate comparisons (Mann-Whitney U) of physiological variables (Heart Rate) between psychological groups during tasks M1 and M2.

CL parameters vs self reports					
ERQ_reev_cog	***ERQ_sup_emo***	***CEI***	***LOC***	***GSE***	***NASA***
M1					
Slope: p = 0.96 Intercept: p = 0.73 R2: p = 0.90 pValue: p = 0.75 CL_final: p = 0.39 CL_variation: p = 0.90	Slope: p = 0.14 Intercept: p = 0.04 R2: p = 0.20 pValue: p = 0.10 CL_final: p = 0.46 CL_variation: p = 0.12	Slope: p = 0.76 Intercept: p = 0.67 R2: p = 0.43 pValue: p = 0.78 CL_final: p = 0.51 CL_variation: p = 0.74	Slope: p = 0.29 Intercept: p = 0.91 R2: p = 0.99 pValue: p = 0.91 CL_final: p = 0.21 CL_variation: p = 0.35	Slope: p = 0.90 Intercept: p = 0.75 R2: p = 0.01 pValue: p = 0.01 CL_final: p = 0.28 CL_variation: p = 0.70	Slope: p = 0.40 Intercept: p = 1.00 R2: p = 0.17 pValue: p = 0.22 CL_final: p = 0.05 CL_variation: p = 0.39
M2					
Slope: p = 0.67 Intercept: p = 0.69 R2: p = 0.22 pValue: p = 0.35 CL_final: p = 0.86 CL_variation: p = 0.79	Slope: p = 0.59 Intercept: p = 0.09 R2: p = 0.80 pValue: p = 0.75 CL_final: p = 0.69 CL_variation: p = 0.49	Slope: p = 0.24 Intercept: p = 0.26 R2: p = 0.47 pValue: p = 0.54 CL_final: p = 0.67 CL_variation: p = 0.29	Slope: p = 0.07 Intercept: p = 0.02 R2: p = 0.97 pValue: p = 0.86 CL_final: p = 0.94 CL_variation: p = 0.10	Slope: p = 0.25 Intercept: p = 0.88 R2: p = 0.38 pValue: p = 0.53 CL_final: p = 0.20 CL_variation: p = 0.29	Slope: p = 0.52 Intercept: p = 0.40 R2: p = 0.53 pValue: p = 0.81 CL_final: p = 0.67 CL_variation: p = 0.62

Table 16: Univariate comparisons (Mann-Whitney U) of physiological variables (Cognitive Load) between psychological groups during tasks M1 and M2.

- ***Multivariate pattern analysis of HR and CL via PCA***

A Principal Component Analysis (PCA) was independently applied to the physiological features derived from heart rate (HR) and cognitive load (CL) for tasks M1 and M2. Overall, the projection of subjects in the space defined by the first two principal components (PC1 and PC2) did not show a clear separation or consistent grouping according to psychological profiles, indicating a weak discriminative ability of these physiological variables to characterize profiles in most analysed situations. However, a notable exception was observed in task M1. When applying PCA to the physiological variables associated with cognitive load, subjects within the GSE profile showed significant differentiation ($p < 0.05$) compared to other profiles, revealing a distinctive latent structure in the CL physiological data for this particular group (Figure 17).

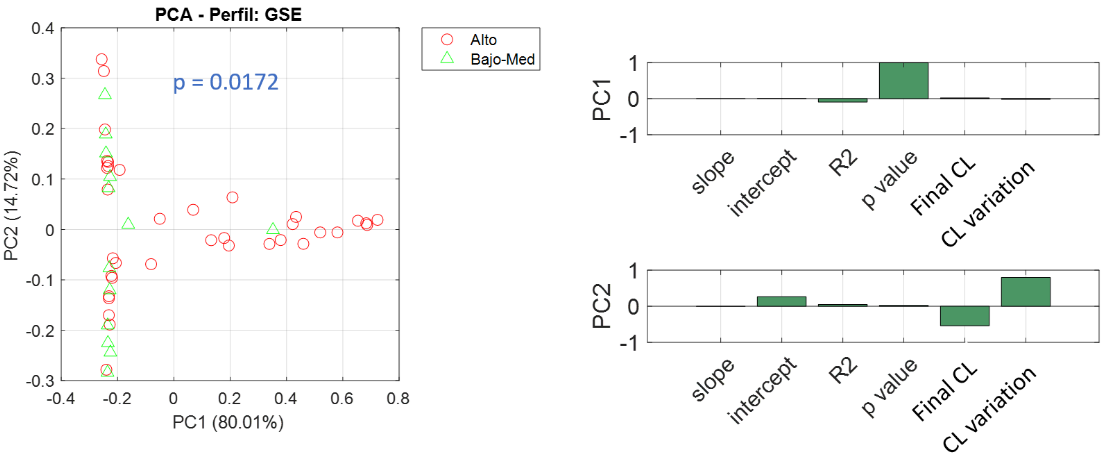


Figure 17: Results of the Principal Component Analysis (PCA), showing that the GSE profile is statistically characterized by the physiological profiles of cognitive load recorded during the execution of task M1.

C. Analysis of Physiological Data Obtained with EMOTIBIT

For the analysis of physiological data in Prototype 2, the EMOTIBIT system was used again under the same conditions and procedures previously described for Prototype 1. Both the JSON file format and the process of extraction, organization, and analysis of physiological variables (HR, EA, and TH) were exactly the same as those applied earlier. Therefore, the reader is referred to the corresponding section of Prototype 1 for further methodological details. Table 17 provides a summary of the data available in this instance for the EMOTIBIT-based analysis during the execution of tasks M1 and M2.

Italy		**Hungary**		**Estonia**	
Heart Rate		*Heart Rate*		*Heart Rate*	
18 (M1)	5 (M2)	0 (M1)	0 (M2)	20 (M1)	23 (M2)

Epidermal activity		*Epidermal activity*		*Epidermal activity*	
19 (M1)	5 (M2)	0 (M1)	0 (M2)	20 (M1)	23 (M2)
Skin temperature		*Skin temperature*		*Skin temperature*	
19 (M1)	5 (M2)	0 (M1)	0 (M2)	20 (M1)	23 (M2)

Table 17: Data available for analysis in Prototype 2 using EMOTIBIT.

- ***Multivariate pattern analysis of HR, EA and TH via PCA***

The multivariate analysis of the physiological signals recorded during Prototype 2 was conducted following the same procedure previously described for the EMOTIBIT data from Prototype 1. Based on the data organization shown in Table 17, a common set of features was extracted for the HR, EA, and TH variables, allowing evaluation of their evolution throughout the experimental tasks using an integrated approach. As detailed earlier, four features were calculated for each physiological variable (initial value, final value, slope, and delta), resulting in a total of 12 descriptors per subject. These variables were normalized using z-score and analysed through principal component analysis (PCA) to identify differentiating physiological patterns between conditions or groups.

D. Results

Figure 18 presents the results of the PCA applied to HR, EA, and TH data obtained during task M2, which corresponds to the only case where significant differences between profiles were observed. Specifically, a statistically significant separation was found among subjects according to their GSE profile, indicating that this profile is differentially represented in the space derived from the physiological measures.

In this figure, the projection of subjects onto the plane defined by the first two principal components (PC1 and PC2) allows visualization of this differential grouping, while the bar charts show the relative contribution of each of the 12 physiological features to the explained variance. This analysis highlights that certain combinations of physiological responses are specifically associated with the GSE profile during task execution in virtual reality environments. It is noteworthy that for the other profiles and in task M1, no consistent separations or significant differences were observed in the PCA space.

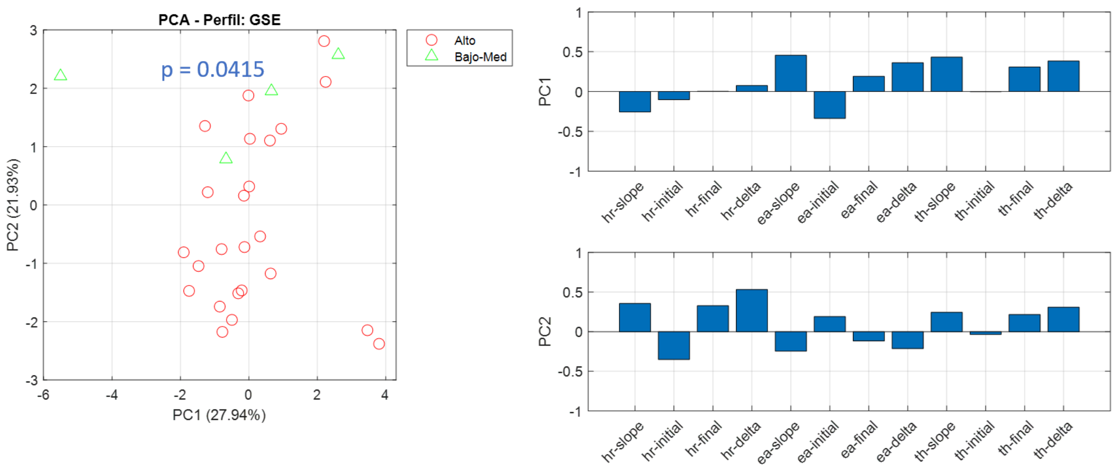


Figure 18: Results of the Principal Component Analysis (PCA), showing that the reported GSE profile is statistically described by the physiological profiles recorded during the execution of task M2.

11.3 Prototype 3

A. Methods

- ***LOG data recordings***

Physiological data analysed in Prototype 3 are contained in single .log files for each experimental session, located in subfolders corresponding to each module (M1, M2, M3, M4, and M5). These files jointly include recordings of various physiological variables such as heart rate (HR), eye tracking (ET), and cognitive load (CL), acquired via a system during the execution of different experimental tasks. Since the physiological signals are stored combined in a single file per session, it was necessary to structure and separate them by variable type for subsequent analysis. The file organization follows a hierarchical folder structure as illustrated in Figure 19.

Data collection was conducted in three countries: Spain (N = 25), Hungary (N = 26), and Cyprus (N = 22). However, cases were identified in which files lacked valid information or were missing. Unlike previous prototypes, no EMOTIBIT devices were used for data acquisition in this case. For this prototype, all modules from M1 to M5 implemented virtual reality technologies.

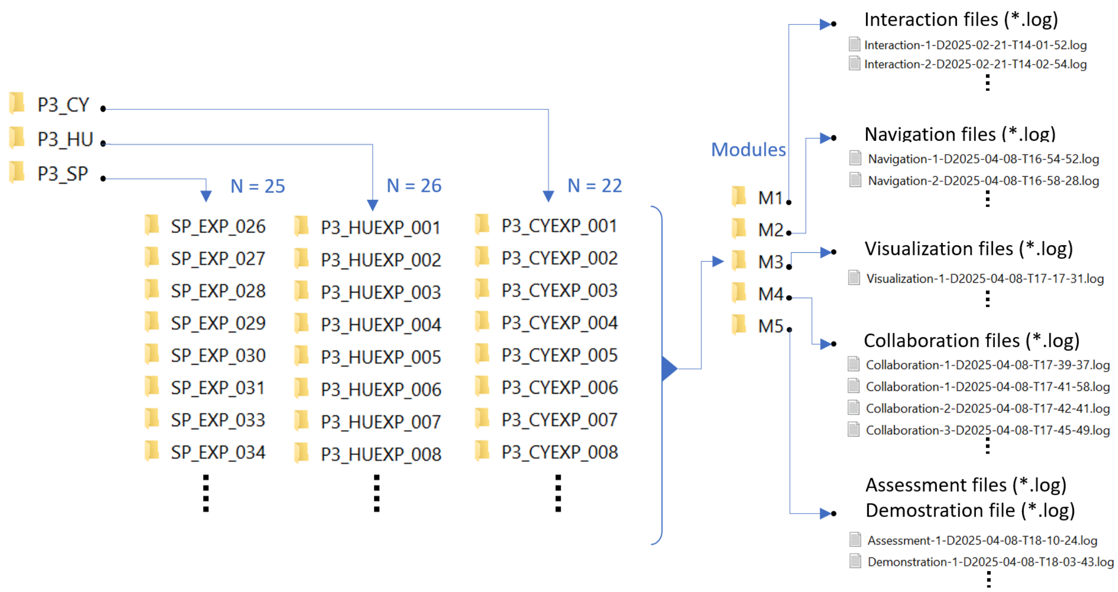


Figure 19: Organizational structure of the data acquired during the sessions of Protocol 3. The countries involved in data collection are Spain, Hungary, and Cyprus. In this prototype, all modules involved virtual reality.

Each of the LOG files generated during the experimental sessions of the different modules (M1 to M5) contains multiple event records and physiological information stored in a structured format. For analysis, these files were processed in MATLAB using a specific script designed to interpret and organize their contents into a hierarchical data structure.

The LOG files include various fields corresponding to user interaction events with the virtual environment, system parameters, and physiological variables. Specifically, the signals related to Heart Rate (HR) and Cognitive Load (CL), along with their respective timestamps, were selectively extracted for subsequent analyses.

The structuring procedure consisted of reading each file as a nested struct-type structure, from which the relevant subfields were accessed. This approach enabled the automation of valid session identification and efficient extraction of the physiological measures of interest. Additionally, the temporal information integrity was preserved, ensuring the proper alignment of signals for comparative analysis across subjects, tasks, and countries.

- ***Availability of physiological data***

In the MATLAB environment, the physiological data were initially organized into a cell array with dimensions N×6. In this array, the first column contained the subject identifiers, while columns two through six stored data structures corresponding to sessions M1 through M5, respectively.

The data were organized in MATLAB following the same structural scheme used in Prototypes 1 and 2. Therefore, the data layout corresponds to that illustrated in Figure 12, applied here specifically to tasks M1 through M5. Table 18 shows the data available for the analysis of physiological profiles in Prototype 3.

Spain	**Hungary**	**Cyprus**
Heart Rate	*Heart Rate*	*Heart Rate*
N = 4 *Cognitive Load*	N = 6 *Cognitive Load*	N = 15 *Cognitive Load*
N = 1	N = 0	N = 25

Table 18: Data available for analysis in Prototype 3. The numbers presented in this table are approximately the same for all modules.

- ***Analyses applied in Prototype 3***

The analyses applied to the data from Prototype 3 exactly replicate the procedures described for Prototypes 1 and 2. Specifically, the following analyses were conducted, following the same methodological specifications previously detailed:

- *Linear modelling of the heart rate (HR) dynamics during task execution.*
- *Linear modelling of the cognitive load (CL) dynamics during task execution.*
- *Analysis of associations between HR, CL, and subjective measures (self-reported psychological profiles).*
- *Multivariate pattern analysis (Multivariate Pattern Analysis) of HR and CL using Principal Component Analysis (PCA).*

Each of these analyses was applied identically to the approach established for Prototype 1, both in terms of task segmentation and statistical procedures used.

B. Results

- ***Univariate Tests***

Results of univariate comparisons using the Mann-Whitney U test, employed to evaluate whether self-reported psychological profiles and cognitive loads can be described by the recorded physiological variables, are presented in the following tables. Table 19 summarizes the univariate comparisons corresponding to heart rate physiological variables during the execution of tasks M1 to M5. Meanwhile, Table 20 shows the equivalent results for cognitive load variables evaluated in the same tasks. Statistically significant differences are highlighted in blue.

HR parameters vs self reports					
ERQ_reev_cog	*ERQ_sup_emo*	*CEI*	*LOC*	*GSE*	*NASA*
		M1			
Slope: p = 0.87 Intercept: p = 0.07 R2: p = 0.94 pValue: p = 0.87 HR_final: p = 0.15 HR_variation: p = 0.87	Slope: p = 0.55 Intercept: p = 0.61 R2: p = 0.24 pValue: p = 0.18 HR_final: p = 0.85 HR_variation: p = 0.50	Slope: p = 0.94 Intercept: p = 0.29 R2: p = 0.40 pValue: p = 0.47 HR_final: p = 0.35 HR_variation: p = 0.94	Slope: p = 0.12 Intercept: p = 0.58 R2: p = 0.38 pValue: p = 0.26 HR_final: p = 0.12 HR_variation: p = 0.10	Slope: p = 0.12 Intercept: p = 0.45 R2: p = 0.65 pValue: p = 0.58 HR_final: p = 0.30 HR_variation: p = 0.15	Slope: p = 0.83 Intercept: p = 0.73 R2: p = 0.07 pValue: p = 0.04 HR_final: p = 0.93 HR_variation: p = 0.93
		M2			
Slope: p = 1.00 Intercept: p = 0.10 R2: p = 0.38 pValue: p = 0.14 HR_final: p = 0.20 HR_variation: p = 0.64	Slope: p = 0.01 Intercept: p = 0.44 R2: p = 0.05 pValue: p = 0.06 HR_final: p = 0.02 HR_variation: p = 0.01	Slope: p = 0.01 Intercept: p = 0.44 R2: p = 0.05 pValue: p = 0.06 HR_final: p = 0.02 HR_variation: p = 0.01	Slope: p = 0.91 Intercept: p = 0.30 R2: p = 0.19 pValue: p = 0.06 HR_final: p = 0.45 HR_variation: p = 0.64	Slope: p = 0.89 Intercept: p = 0.68 R2: p = 0.34 pValue: p = 0.39 HR_final: p = 0.75 HR_variation: p = 0.96	Slope: p = 0.22 Intercept: p = 0.61 R2: p = 0.55 pValue: p = 0.96 HR_final: p = 0.45 HR_variation: p = 0.66
		M3			
Slope: p = 0.09 Intercept: p = 0.62 R2: p = 0.02 pValue: p = 0.05 HR_final: p = 0.09 HR_variation: p = 0.16	Slope: p = 0.36 Intercept: p = 0.58 R2: p = 0.91 pValue: p = 0.32 HR_final: p = 0.40 HR_variation: p = 0.91	Slope: p = 0.29 Intercept: p = 0.80 R2: p = 0.72 pValue: p = 0.88 HR_final: p = 0.80 HR_variation: p = 0.80	Slope: p = 0.64 Intercept: p = 0.69 R2: p = 0.34 pValue: p = 0.41 HR_final: p = 1.00 HR_variation: p = 0.41	Slope: p = 0.20 Intercept: p = 0.45 R2: p = 0.14 pValue: p = 0.65 HR_final: p = 0.72 HR_variation: p = 0.20	Slope: p = 0.78 Intercept: p = 0.41 R2: p = 0.78 pValue: p = 0.61 HR_final: p = 0.41 HR_variation: p = 0.78
		M4			
Slope: p = 0.21 Intercept: p = 0.30 R2: p = 0.63 pValue: p = 0.63 HR_final: p = 0.08 HR_variation: p = 0.21	Slope: p = 0.75 Intercept: p = 0.47 R2: p = 0.47 pValue: p = 0.20 HR_final: p = 0.39 HR_variation: p = 0.65	Slope: p = 0.70 Intercept: p = 0.58 R2: p = 0.58 pValue: p = 0.52 HR_final: p = 0.52 HR_variation: p = 0.96	Slope: p = 0.27 Intercept: p = 0.81 R2: p = 1.00 pValue: p = 0.48 HR_final: p = 0.27 HR_variation: p = 0.12	Slope: p = 0.63 Intercept: p = 0.96 R2: p = 0.37 pValue: p = 0.52 HR_final: p = 0.89 HR_variation: p = 0.63	Slope: p = 0.25 Intercept: p = 0.07 R2: p = 0.85 pValue: p = 0.48 HR_final: p = 0.08 HR_variation: p = 0.28
		M5			
Slope: p = 0.44 Intercept: p = 0.77 R2: p = 0.50 pValue: p = 0.92 HR_final: p = 0.24 HR_variation: p = 0.50	Slope: p = 0.06 Intercept: p = 0.75 R2: p = 1.00 pValue: p = 1.00 HR_final: p = 0.08 HR_variation: p = 0.15	Slope: p = 0.95 Intercept: p = 0.44 R2: p = 0.72 pValue: p = 0.95 HR_final: p = 0.64 HR_variation: p = 0.79	Slope: p = 0.31 Intercept: p = 0.89 R2: p = 0.20 pValue: p = 0.45 HR_final: p = 0.31 HR_variation: p = 0.57	Slope: p = 1.00 Intercept: p = 0.65 R2: p = 0.36 pValue: p = 0.82 HR_final: p = 0.65 HR_variation: p = 1.00	Slope: p = 0.11 Intercept: p = 0.11 R2: p = 0.88 pValue: p = 1.00 HR_final: p = 0.66 HR_variation: p = 0.11

Table 19: Univariate comparisons (Mann-Whitney U) of physiological variables (Heart Rate) between psychological groups during tasks M1 to M2.

CL parameters vs self-reports					
ERQ_reev_cog	*ERQ_sup_emo*	*CEI*	*LOC*	*GSE*	*NASA*
		M1			
Slope: p = 0.94 Intercept: p = 0.17 R2: p = 0.57 pValue: p = 0.13 CL_final: p = 0.41 CL_variation: p = 1.00	Slope: p = 0.77 Intercept: p = 0.45 R2: p = 0.77 pValue: p = 1.00 CL_final: p = 0.68	Slope: p = 1.00 Intercept: p = 0.11 R2: p = 0.17 pValue: p = 0.03 CL_final: p = 0.68	Slope: p = 0.14 Intercept: p = 0.27 R2: p = 0.86 pValue: p = 0.77 CL_final: p = 0.86	Slope: p = 0.38 Intercept: p = 0.22 R2: p = 0.68 pValue: p = 0.68 CL_final: p = 0.30	Slope: p = 0.22 Intercept: p = 0.22 R2: p = 0.57 pValue: p = 0.68 CL_final: p = 0.17 CL_variation: p = 0.22

	CL_variation: p = 1.00	CL_variation: p = 0.93	CL_variation: p = 0.06	CL_variation: p = 0.30	
M2					
Slope: p = 0.31 Intercept: p = 0.21 R2: p = 0.44 pValue: p = 0.37 CL_final: p = 0.95 CL_variation: p = 0.31	Slope: p = 0.18 Intercept: p = 0.71 R2: p = 0.71 pValue: p = 0.79 CL_final: p = 0.26 CL_variation: p = 0.49	Slope: p = 1.00 Intercept: p = 0.81 R2: p = 0.70 pValue: p = 0.60 CL_final: p = 1.00 CL_variation: p = 0.70	Slope: p = 0.83 Intercept: p = 0.35 R2: p = 0.91 pValue: p = 0.75 CL_final: p = 0.91 CL_variation: p = 0.46	Slope: p = 0.93 Intercept: p = 0.26 R2: p = 0.70 pValue: p = 0.70 CL_final: p = 0.26 CL_variation: p = 0.93	Slope: p = 0.52 Intercept: p = 1.00 R2: p = 0.10 pValue: p = 0.05 CL_final: p = 0.10 CL_variation: p = 0.31
M3					
Slope: p = 0.57 Intercept: p = 0.13 R2: p = 0.57 pValue: p = 0.57 CL_final: p = 0.34 CL_variation: p = 0.34	Slope: p = 0.68 Intercept: p = 0.22 R2: p = 0.77 pValue: p = 0.95 CL_final: p = 0.52 CL_variation: p = 0.86	Slope: p = 0.36 Intercept: p = 0.01 R2: p = 0.18 pValue: p = 0.23 CL_final: p = 0.63 CL_variation: p = 0.36	Slope: p = 0.38 Intercept: p = 0.86 R2: p = 0.86 pValue: p = 0.95 CL_final: p = 0.77 CL_variation: p = 0.68	Slope: p = 0.57 Intercept: p = 0.47 R2: p = 0.03 pValue: p = 0.07 CL_final: p = 0.93 CL_variation: p = 0.93	Slope: p = 0.73 Intercept: p = 0.44 R2: p = 0.53 pValue: p = 0.63 CL_final: p = 1.00 CL_variation: p = 1.00
M4					
Slope: p = 0.85 Intercept: p = 0.85 R2: p = 0.57 pValue: p = 0.57 CL_final: p = 0.85 CL_variation: p = 0.94	Slope: p = 0.77 Intercept: p = 0.86 R2: p = 0.68 pValue: p = 0.95 CL_final: p = 0.68 CL_variation: p = 0.68	Slope: p = 0.53 Intercept: p = 0.44 R2: p = 0.83 pValue: p = 1.00 CL_final: p = 0.73 CL_variation: p = 0.63	Slope: p = 0.69 Intercept: p = 1.00 R2: p = 0.18 pValue: p = 0.23 CL_final: p = 0.61 CL_variation: p = 0.69	Slope: p = 0.80 Intercept: p = 0.57 R2: p = 0.93 pValue: p = 0.93 CL_final: p = 0.30 CL_variation: p = 0.93	Slope: p = 0.53 Intercept: p = 0.83 R2: p = 0.83 pValue: p = 0.94 CL_final: p = 0.53 CL_variation: p = 0.53
M5					
Slope: p = 0.80 Intercept: p = 0.21 R2: p = 0.68 pValue: p = 0.68 CL_final: p = 0.28 CL_variation: p = 0.56	Slope: p = 0.39 Intercept: p = 0.24 R2: p = 1.00 pValue: p = 0.81 CL_final: p = 0.69 CL_variation: p = 0.48	Slope: p = 0.20 Intercept: p = 0.72 R2: p = 0.48 pValue: p = 0.48 CL_final: p = 0.14 CL_variation: p = 0.14	Slope: p = 0.43 Intercept: p = 0.20 R2: p = 0.63 pValue: p = 0.43 CL_final: p = 0.34 CL_variation: p = 0.87	Slope: p = 1.00 Intercept: p = 1.00 R2: p = 1.00 pValue: p = 0.83 CL_final: p = 0.66 CL_variation: p = 0.83	Slope: p = 0.80 Intercept: p = 0.21 R2: p = 0.68 pValue: p = 0.68 CL_final: p = 0.28 CL_variation: p = 0.56

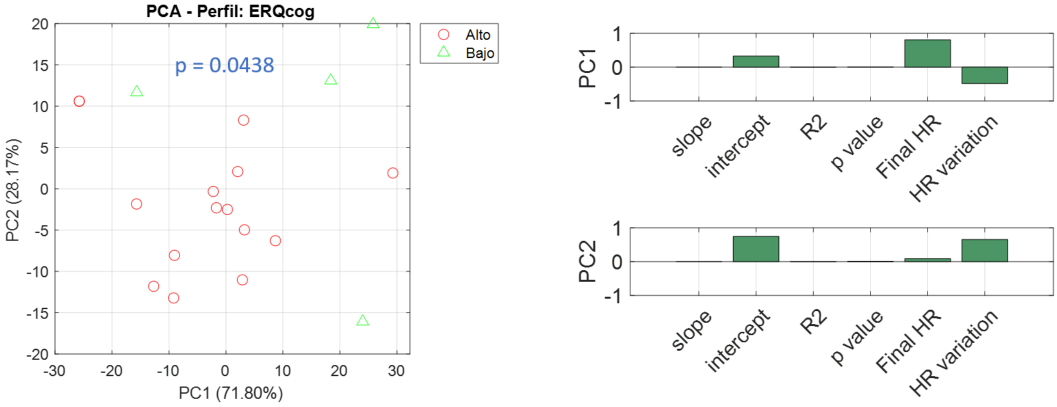
Table 20: Univariate comparisons (Mann-Whitney U) of physiological variables (Cognitive Load) between psychological groups during tasks M1 to M2.

- ***Multivariate pattern analysis of HR and CL via PCA.***

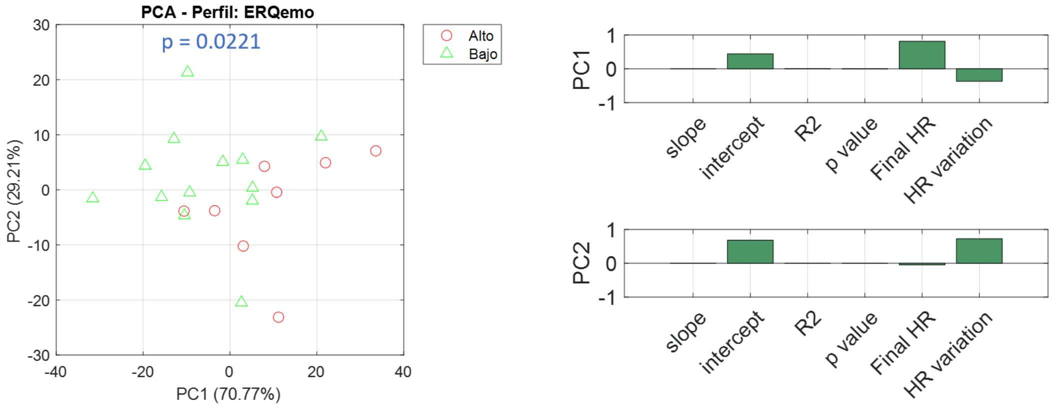
Only statistically significant results are shown in Figure 20. Principal Component Analysis (PCA) was used to explore possible group differences in the space defined by the first two components (PC1 and PC2), based on physiological variables recorded during the experimental tasks. The results indicated that significant differences between groups were observed only in specific tasks and for certain psychological profiles, and exclusively when heart rate–related physiological variables were used. In contrast, no significant differences were detected in any of the cases analysed when cognitive load variables were employed.

e-DIPLOMA
Funded by the European Union

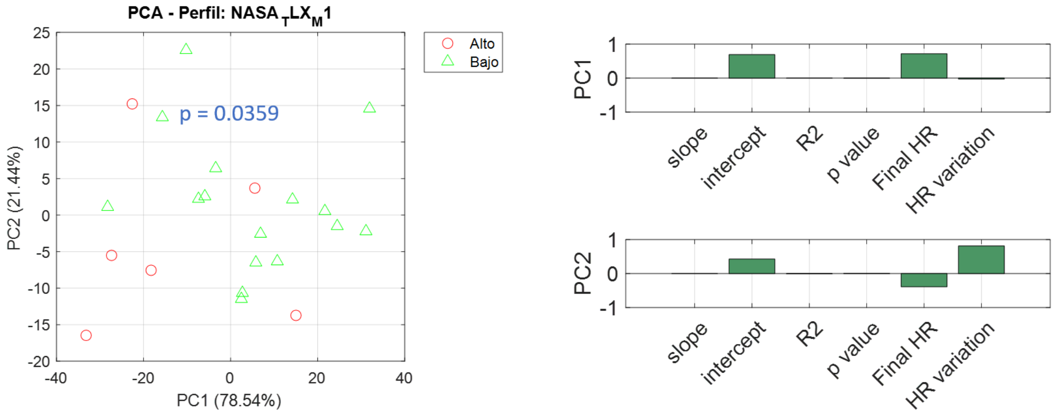
A- M1 TASK, HR variables



B- M2 TASK, HR variables



C- M4 TASK, HR variables



D- M5 TASK, HR variables

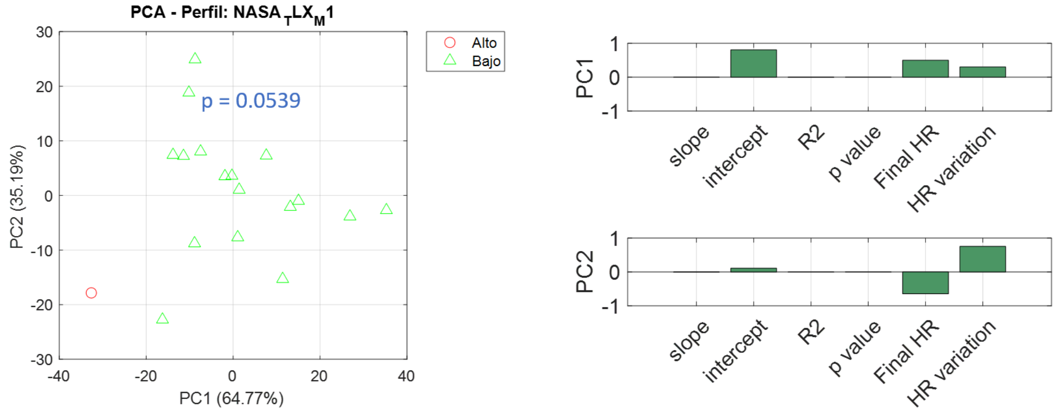


Figure 20: Results of the Principal Component Analysis (PCA), showing all situations where the reported psychological profiles and cognitive load are statistically described by the heart rate physiological profiles recorded during the execution of the different tasks.

11. Conclusions

The results of Objective 3, which focuses on the interaction between learner characteristics and the effectiveness of immersive learning technologies, reinforce the hypothesis that emotional and cognitive traits significantly influence how learners engage with and benefit from disruptive technologies. In particular, we observed that learners' cognitive and emotional profiles—including locus of control, emotional self-regulation, and curiosity—have a measurable impact on learning outcomes in virtual reality (VR) environments.

Participants with higher emotional self-regulation, a stronger internal locus of control, and greater curiosity consistently showed more favourable learning outcomes across tasks, as measured by both Absolute Difference (AD) and Normalized Gain (NG). These results highlight that students with better cognitive-emotional adaptability tend to process information more effectively and acquire learning content more efficiently in immersive VR settings. The learning gains were more pronounced in the psychomotor content of Prototype 3, where task complexity and interactive requirements create a higher cognitive load.

Interpretation of Results: Influence of Cognitive-Emotional Skills

Curiosity: Learners with higher curiosity levels displayed a greater willingness to explore the VR environment and engage in tasks autonomously. This was reflected in their longer interaction times and higher engagement levels in modules requiring decision-making and exploration. These students not only performed better in terms of knowledge acquisition but also demonstrated higher retention rates. Curiosity-driven exploration leads to deeper engagement with content and an active search for challenges, which improves sustained attention and learning retention.

Emotional Self-Regulation: Students with stronger emotional self-regulation abilities showed lower fluctuations in cognitive load during the VR tasks, allowing them to sustain attention and effort without being overwhelmed. These learners exhibited greater consistency in task completion and faster adaptation to task demands, particularly in complex modules that required rapid decision-making and spatial awareness. Emotional regulation helps reduce the impact of stress or frustration, enabling a smoother learning experience and better overall performance, especially in tasks that present higher cognitive complexity.

Internal Locus of Control: Learners with a more internal locus of control demonstrated greater initiative in taking responsibility for their learning and showed higher problem-solving efficiency. These students managed to navigate the challenges of the VR environment with more autonomy and confidence, as they were less likely to rely on external help or guidance. They tended to approach tasks more strategically, which was reflected in their improved learning outcomes. This result emphasizes the importance of self-direction in immersive learning, where the learner's sense of control directly affects engagement and performance.

Adaptation of Disruptive Technologies Based on Learner Profiles

The findings from Objective 3 suggest that disruptive technologies, such as VR, need to be adapted to the cognitive and emotional profiles of students to maximize their learning potential. Given that locus of control, emotional self-regulation, and curiosity significantly influence learning outcomes, adaptive systems can be developed to dynamically adjust the learning experience based on real-time data from physiological sensors (such as heart rate (HR) and cognitive load (CL)).

Adaptation to Curiosity: The VR systems could provide enhanced exploratory opportunities for students with higher curiosity, such as open-ended tasks, unexpected challenges, and higher levels of interactivity. These adjustments would align with their intrinsic motivation to explore and experiment, promoting autonomous learning. For students with lower curiosity, the system could offer guided experiences, gently encouraging exploration and curiosity-driven tasks to foster greater engagement.

Adaptation to Emotional Self-Regulation: For learners with lower emotional self-regulation, VR environments could be equipped with built-in emotional support mechanisms such as real-time feedback about their emotional state, stress-relief exercises, or task pacing adjustments. This would help prevent emotional overload and ensure that learners are not overwhelmed by the VR content, thereby maintaining cognitive load at an optimal level for learning. For learners with strong emotional regulation, the system could present more complex challenges that test their ability to maintain composure and manage high-stakes tasks.

Adaptation to Internal Locus of Control: Learners with an internal locus of control could benefit from more autonomous, self-directed VR learning environments that allow them to choose tasks, set goals, and track progress. These learners are typically motivated by control over their learning process, and offering them opportunities to make decisions about the flow of the learning experience would align with their self-determined nature. In contrast, students with a more external locus of control might benefit from structured guidance and scaffolded tasks that provide clear instructions and step-by-step support.

Improving Information Processing and Content Acquisition

The ability to tailor the VR learning experience to students' cognitive-emotional profiles can significantly enhance their information processing and content acquisition. Physiological sensors (such as HR and CL) provide real-time feedback on how students are responding to the immersive content, which can be used to adjust task complexity, interaction intensity, and learning pace. This adaptability could optimize cognitive load, helping students stay in a state of flow where they are sufficiently challenged but not overwhelmed, resulting in more effective learning.

The combination of real-time biometric data and adaptive content delivery can help create a personalized learning environment that fosters both engagement and efficiency. For example, as students with high curiosity become more engaged, the system could increase task complexity or offer new challenges, while for students with lower emotional regulation, the system could reduce task difficulty or offer emotional support to ensure that they stay engaged without becoming frustrated.

Ultimately, the results from Objective 3 confirm that disruptive technologies, when aligned with learners' cognitive and emotional profiles, can offer personalized learning experiences that enhance both performance and content retention. By adapting the learning environment to students' profiles, VR and similar technologies can help optimize cognitive processing and boost learning outcomes, ensuring that every learner is placed in an environment where they can thrive.

e·DIPLOMA

Funded by
the European Union