

Exploring age-related differences in immersive Virtual Reality experiences for cultural heritage

Bruno Rodriguez-Garcia · Ines Miguel-Alonso · David Checa · Andres Bustillo

Received: 18 September 2025 / Accepted: 5 June 2026

© The Author(s) 2026

Abstract

Immersive Virtual Reality (iVR) has reached a level of technological maturity that enables its use for the general public. However, further investigation is needed into the optimization of iVR applications and particularly, their design features in relation to various user characteristics such as age and previous iVR experience. To solve this issue, this study evaluates a Cultural Heritage dissemination experience in iVR, tested in a real-world setting, the Museum of Burgos (Spain), with participants from three age groups: children, young adults, and older adults. The iVR experience was structured in four levels of progressively complex interaction. A control group was formed to view a video containing the same dissemination content. A total of 206 participants completed a three-steps assessment measuring prior knowledge, User Experience (UX), and learning outcomes. The findings emphasized that iVR's strengths, immersion and interactivity, must be strategically optimized to outperform more well-established dissemination formats like video. The results showed that children responded positively to high interactivity; their motivation and rapid adaptation supported the use of interactive formats, provided cognitive demands are gradually introduced. Young adults exhibited the highest adaptability and fastest completion times, performing well across all interaction types, though with moderate UX ratings. Older adults preferred soft interaction and immersive aesthetics, achieving better learning outcomes in this context but extending the time to complete the iVR experience.

Keywords Virtual Reality · Human Computer Interaction · User experience · Children · Age · User Centered Design

1 Introduction

One of the main barriers to the adoption of new technologies is the level of difficulty that users face in adapting to them (Tawfik et al. 2021). Successful implementation requires attention to usability and interface design (Briones-Villafuerte et al. 2022), as personalized approaches such as User-Centered Design (UCD) (Partarakis & Zabulis 2024), and affordability (Ryokai et al. 2022). Without fulfilling those conditions, new technologies often remain limited to early adopters and research prototypes, complicating their real-world integration (Kouijzer et

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.



al. 2023). Even when the advantages are clear, implementation can still be obstructed by low levels of user acceptance and contextual barriers (Samala et al. 2025).

The release of Oculus Rift, an affordable Head-Mounted Display (HMD), was a significant milestone in the broader adoption of immersive Virtual Reality (iVR). iVR is primarily characterized by its interactivity, allowing users to alter and engage with the virtual environment at deeper levels, and by its capacity to induce a strong sense of presence or immersion, often described as the feeling of “being there” (Tcha-Tokey et al. 2016). The real-time interaction provides benefits such as operating hazardous machinery without risk (Dhalmahapatra et al. 2021). Likewise, the sense of presence enables users to explore past events (Slater et al. 2018) or long-lost historical environments from a first-person perspective (Ferdani et al. 2020).

Despite the great presence of iVR in literature (Lee et al. 2024; Silva & Teixeira 2022), its limitations, particularly interaction paradigms, continue to hinder adoption (Avilés-Castillo et al. 2025). Many issues arise not from hardware but from design shortcomings, which could be improved adapting iVR to different user characteristics, such as prior experience (Miguel-Alonso et al. 2024) or age (Cadet & Chainay 2021). Typically, User Experience (UX) in iVR has been studied primarily in young adult populations (M. J. Kim & Kang 2023), often neglecting other user groups in which iVR has demonstrated notable benefits, such as improving physical conditions in older adults (Schaumburg et al. 2025) or enhancing motivation of children in settings like museums (Tengfei & Guanjie 2024). While other studies have addressed these populations, such research remains relatively scarce (Forssell et al. 2024). Many studies involving older adults focus on users with dementia or other impairments (Baker et al. 2020), so their results are inapplicable to populations in good health (Healy et al. 2021). Additionally, sample sizes tend to be limited, often involving only around 20 participants per group (Abeele et al. 2021). Moreover, these studies typically compare only two age groups, one of which is almost always young adults, the most studied group in iVR research (M. J. Kim & Kang 2023), paired either with children (Cadet & Chainay 2021) or with older adults (Dong et al. 2023). Although such investigations yield valuable insights, they rarely compare multiple age groups simultaneously within the same experience, which would provide more comprehensive conclusions regarding UCD. Lastly, these studies generally focus exclusively on iVR experiences, with few comparisons to more established technologies such as video.

This study addresses the gap in user-adaptive iVR design, focusing on museums as a real-world context. This context was selected because it implies usually diverse background audiences (Gatto et al. 2025), different age groups (Tengfei & Guanjie 2024) and often with low prior iVR experience (Schofield et al. 2018). The high visitor numbers and limited time make museum-based iVR particularly challenging to implement (Rodriguez-Garcia et al. 2024). The research has been designed to answer the following Research Questions (RQ):

RQ1: What influence does age have on behavior, UX, and learning in iVR experiences?

RQ2: How should iVR experiences be adapted for different age groups?

To address these questions, this study presents an iVR experience specifically designed to evaluate UX, with four levels of increasing interactivity, and extract UCD guidelines for iVR experiences design. Its validation has been done on a large sample (N=206), that covers different age groups (children, young adults and elderly). The experimental design incorporates a wide range of variables that have been considered relevant in other studies related to UX in iVR: prior iVR experience (Miguel-Alonso et al. 2024), level of interaction (Kim et al. 2020), locomotion systems (Lim et al. 2022), and learning outcomes (Checa & Bustillo 2020). For this purpose, an iVR experience was created around the Isis of Clunia, a Roman sculpture from the Museum of Burgos (Spain) discovered in Clunia Sulpicia, one of the most important roman cities in the north of the Iberian Peninsula. This sculpture has been previously reconstructed and validated as a Virtual Agent (VA) (Rodriguez-Garcia et al. 2026). The iVR experience was used to disseminate knowledge about CH: Isis’s origins, Roman iconography, and the specific artistic features of the sculpture. Finally, a control group watched a video with the same content, in order to compare the learning outcomes with those of a more established technology. Although alternative approaches for immersive and non-immersive environments exist, such as stereoscopic displays (Printezis & Koutsabasis

2026), iVR through HMDs and video experiences were selected, as they represent some of the most accessible and widely adopted technologies, particularly in museum contexts (Gatto et al. 2025).

The structure of this research is as follows. Section 2 summarizes the most relevant features of the case study, the Isis of Clunia. Section 3 provides a detailed description of the design and development of the iVR experience and the video. Section 4 outlines the evaluation procedure followed in this research. Section 5 explains the study's results, while Sect. 6 provides the discussion, including best practices and study limitations. Finally, Sect. 7 provides a summary of the main conclusions and future lines of this research.

2 Study case: Isis of Clunia

The historical content about the goddess Isis, and specifically the Clunia sculpture, is broad and complex. This section summarizes the main components evaluated in both the iVR experience and the video: 1) Origins and cult, 2) Iconography, and 3) Formal characteristics of the Clunia sculpture.

Isis, an Egyptian goddess of fertility, motherhood, and magic, was the wife of Osiris and the mother of Horus. After Osiris's murder, she restored his life with her magic (Pinch 2004). Her cult endured Roman assimilation of the Egyptian pantheon and was adopted in Hispania, where it became linked to motherhood (Arroyo de la Fuente 2002). The Clunia sculpture, like others in Roman culture, reflects a Hellenistic rather than Egyptian style (Zaparaín Yáñez 2022). It included four attributes (Arroyo de la Fuente 2002): 1) the sistrum, an ancient rattle made of metal rods, usually held in her right hand, 2) the situla, a ritual pitcher typically carried in her left hand, 3) the crescent moon crown adorning her hair, and 4) the Isiac knot, which gathered her mantle around her chest. Figure 1b shows the virtual reconstruction of the sculpture highlighting the above attributes.

The sculpture that dates to the 2nd century AD, was discovered in 1853 by Salomón Remigio. Its deterioration over the centuries had left it without arms and most of the other attributes, except for the Isiac knot (Salomón, 1853). Its original location is thought to have been Clunia's theater (Muñiz Grijalvo 2021), shown in Fig. 1c. Clunia Sulpicia was one of the most important Roman cities in the north of the Iberian Peninsula, in the present-day Burgos province. The city is known for its forum, baths, and theater. Today the sculpture is exhibited at the Museum of Burgos (Fig. 1a), being one of the museum's key pieces.

3 Materials

First, the design and development of the iVR experience will be described, followed by a description of the video made for the control group. Both the iVR experience and the video may be viewed in Supplementary file 1.

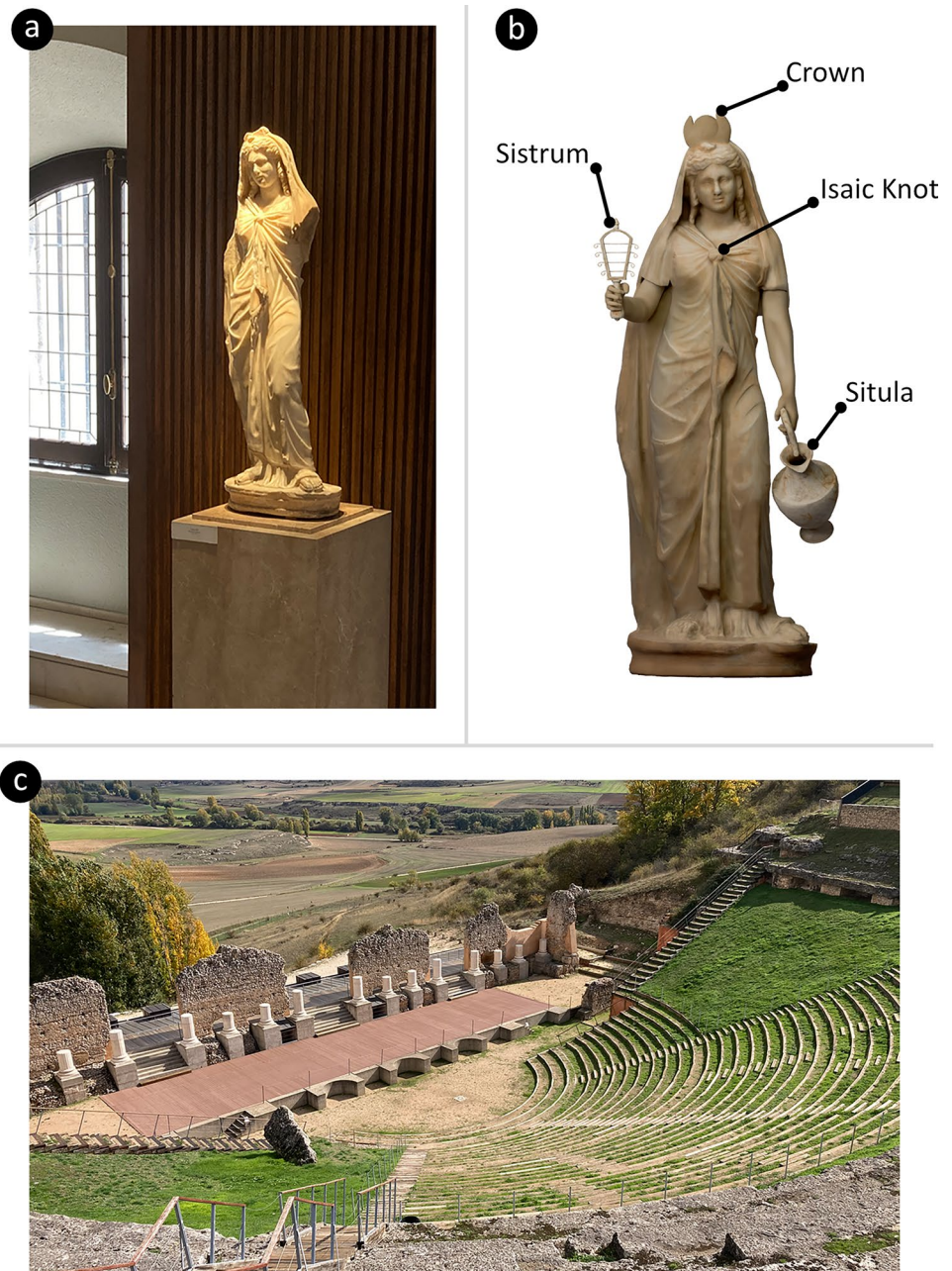
3.1 Design of the iVR experience

The design of the iVR experience and an overview of its levels are outlined below, after which its development and format-related aspects are addressed.

3.1.1 iVR design

The iVR experience is designed to convey the historical content from Sect. 2 to a broad museum audience. It aims to be accessible and enjoyable for children, young adults, and elderly users, as well as for those with low prior experience with iVR. The goal is to combine education with engagement, balancing enjoyment and content understanding. The following paragraphs describe the main design challenges addressed during its development: the research approach, the time, the user motivation, the novelty effect, the knowledge transmission and the adverse effects of iVR.

Fig. 1 **a** Image of the Isis sculpture at its current location in the Museum of Burgos, **b** Virtual reconstruction of the statue with each main attribute highlighted, and **c** Image of the Roman theater of Clunia, where the sculpture was discovered



The research approach was aimed at identifying the most suitable design patterns for different user profiles, which is a limitation compared to commercial products that rely on established patterns. To do so, progressively complex interactions were implemented in a level-based structure. This structure splits the iVR experience into two halves. One of soft interaction with Levels 1 and 2, and a more complex hard interaction with Levels 3 and 4. Each level records completion time and a behavioral indicator related to the main interaction, detailed in the following subsections. The design enhances the research potential in so far as users are exposed to a range of interactions within the iVR environment, facilitating behavioral measurements across scenarios. In this design, the aim was to include the most essential interactions in an iVR experience, namely locomotion and the manipulation of virtual objects.

The time users can spend in an iVR experience is a critical design factor that can vary depending on many other variables. Experimental evaluations of prolonged iVR experiences also pose significant challenges (Vlahovic et

al. 2022). The iVR experience was therefore designed to last approximately 6–7 minutes, which was considered an appropriate length of time for the context, even though slower users may need extra time.

User motivation is crucial for sustaining engagement during iVR experiences (Xiao & Hew 2024). It can be enhanced through the use of narrative elements (Schaumburg et al. 2025). In the iVR experience, the goddess Isis appears at the Roman Theatre of Clunia as a semiautonomous VA (Wei et al. 2025) who asks users to help reconstruct her sculpture by selecting between various options. VAs have been explored in previous research to boost engagement, especially among children (Bailey & Schloss 2024). Likewise, visual and auditory elements, including the goddess' voice, support historical learning and have proven satisfying for elderly users (Fiocco et al. 2021). The content is embedded within a coherent narrative, interweaving information on her origins, worship, and formal characteristics rather than presenting a linear account.

The novelty effect is the initial excitement of using new technology and can influence learning and motivation (Looi et al. 2009). It is of particular relevance to museum audiences whose prior iVR experience is often limited (Puig et al. 2020). The level-based game structure was implemented to reduce the impact of the novelty effect, so that users gradually acclimate as the level of intensity increases. Although tutorials can mitigate the novelty effect (Miguel-Alonso et al. 2022), they were excluded to avoid lengthening the iVR experience and to study user interaction in a real-world context for deriving adaptive design guidelines.

Knowledge transmission is a widely studied component of iVR experiences, often guided by such theories as the Cognitive Theory of Multimedia Learning (Richard E. Mayer 2014a and b), gamification (Faust 2021) or the Cognitive Affective Model of Immersive Learning (Makransky & Petersen 2021). Applying these theories to iVR experiences for a broad audience is challenging, as the iVR experience cannot only be focused on learning. Even so, storytelling, which is considered a form of gamification, was employed to enhance engagement (Staneva et al. 2023). The design leverages iVR immersive characteristics to convey knowledge, using the Roman Theatre of Clunia as a visual setting and spatialized audio for the differentiation of auditory cues. Interaction allows users to pick up and explore key visual elements, as the statue or an old drawing of it.

Adverse effects of iVR are an important consideration when designing for a varied audience, particularly comfort and cybersickness (Porcino et al. 2020). Usually designed for young adults, HMDs can cause fit and usability issues for children (Muravevskaia & Gardner-McCune 2023) and elderly users (Abeele et al. 2021). In addition, misalignment may lead to vision problems among children (Ryokai et al. 2022), though newer HMDs have since reduced that risk (Alrashidi et al. 2023). To solve this issue, modern HMDs with improved weight distribution and alternative straps were used to enhance ergonomics (Mehrfard et al. 2019). Specifically, the Meta Quest 2, Meta Quest 3, and Meta Quest Pro models were used. For the Quest 2 and Quest 3, the default elastic strap was replaced with a more rigid and adjustable one to better fit different users, particularly children. Moreover, controllers were used because they generally provide better usability than other solutions such as hand tracking (Baker et al. 2020). To further enhance usability, the interaction was designed to be operable with a single hand (Schaumburg et al. 2025). Active movement was employed to further reduce discomfort, especially for elderly users, as it is considered the most comfortable locomotion system (Brument et al. 2024). In this regard, only one level using an alternative locomotion system. Finally, the short duration of the iVR experience lowers the risk of adverse effects.

Summarizing, the iVR experience was designed to balance key factors without over-targeting a specific audience. The museum setting constraints were considered to create an enjoyable iVR experience for most users while conveying the sculpture's history and iconography. In Fig. 2 each level of the iVR experience is presented, while vertically, the most relevant characteristics of the levels are described.

3.1.2 Level 1 design

This level takes place in a generic Egyptian temple with dim lighting and the sound of fire. At the center stands the fully reconstructed statue of the Goddess Isis, as shown in Fig. 3a. After 5 seconds, the statue begins to speak,

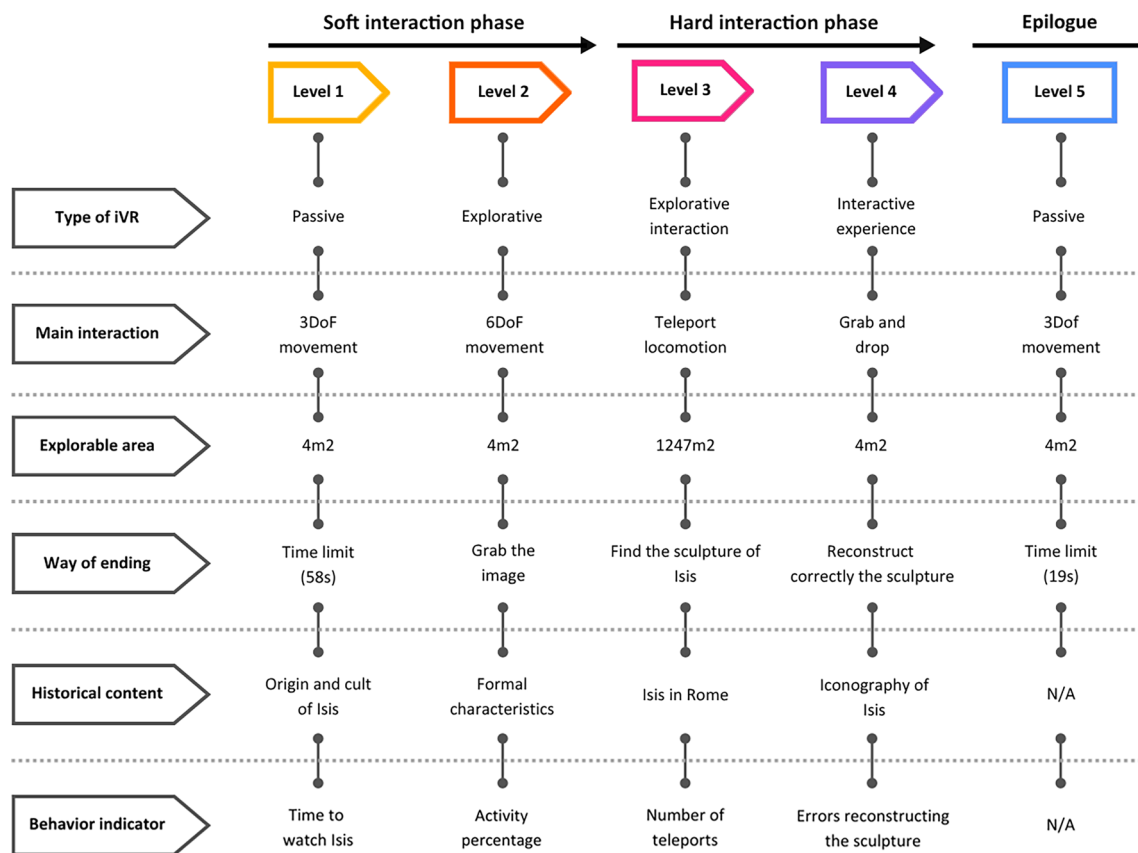


Fig. 2 Summary of the most relevant characteristics of the iVR experience levels. Levels are shown horizontally, while their characteristics are displayed vertically

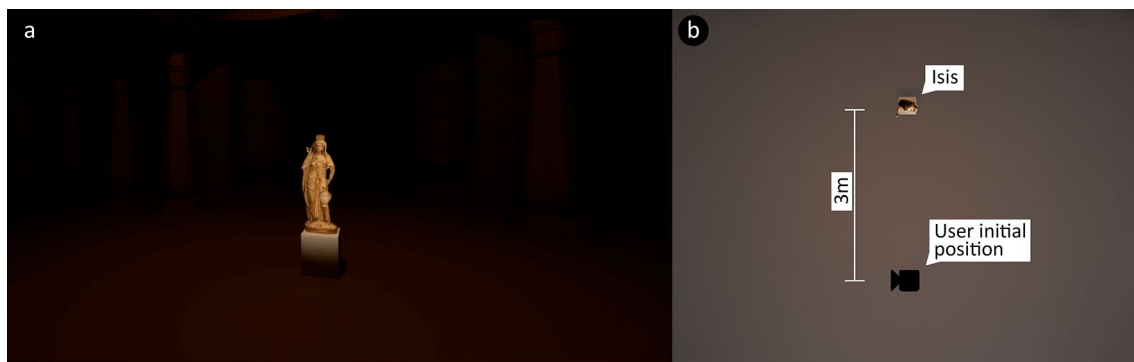


Fig. 3 **a** User's point of view looking at Isis in Level 1 and **b** Top-down view of Level 1 indicating the initial position of the user, Isis, and the distance between them

providing the user with context about the goddess's origins as well as the task to be completed in the experience: to locate and to reconstruct her statue in the theater of Clunia.

This level introduces the goddess's Egyptian origins and the narrative goal of finding and reconstructing her statue. From a usability point of view, it is a simple Passive iVR experience, with no direct interaction and very limited movement, though still offering a 6 Degrees of Freedom (6DoF) experience. Users start facing the statue's side and must turn 90°, guided by Isis's spatialized voice. Figure 3b shows a top-down view of the level layout. The aim of the level is to acclimate them through a basic spatial interaction like a 3 Degrees of Freedom (3DoF)

experience. This level ends automatically after 58 seconds without requiring any user action. The behavior indicator is the time users take to look at Isis, showing their reaction to the auditory cue and awareness of turning. The total completion time is also recorded, although it is identical for all users.

3.1.3 Level 2 design

Level 2 takes place at the center of the Roman theater of Clunia with ambient nature sounds, as shown in Fig. 4a from the user's point of view. The goddess speaks in a reverberated, undefined voice, explaining the location and task. To proceed, users must move around the stage to grab an image that reveals the statue's current appearance.

The narrative explains the sculpture's formal traits and its discovery, reinforced by Isis's voice, the theater setting, and the image. It advances the narrative by requiring users to first locate the reference image before finding the statue. From a usability perspective, this level introduces 6DoF interaction or Real-Walking (Boletsis & Chasanidou 2022) in an Explorative experience, where the user can explore the environment but with limited interaction. Unlike the previous level, where darkness and lack of guidance discourage movement, Level 2 motivates users to move around and to interact while searching for the image. The grabbing interaction is introduced at the end of the level and is therefore not strictly part of it. The end of the level is also the end of the soft interaction phase of the iVR experience, with 3DoF and 6DoF movement interactions, while the following levels introduce more complex mechanics in a hard interaction phase.

To complete the level, users must locate and grab the image placed 1.1 m away among low stones, requiring them to crouch. Additional physics-enabled stones surround it, moving when touched. The interaction occurs within a 4 m² area, adaptable to different installation spaces. Figure 4b shows a top-down view of the level layout. As the main interaction is 6DoF movement, the performance indicator is the percentage of time users remain active (over 10 cm/s). The level ends once the image is grabbed, with no fixed duration.

3.1.4 Level 3 design

This level continues in the Roman theatre of Clunia, right after the user retrieves the image. Isis instructs the user to teleport in the search for her statue, marking the end of the level. She also recounts her story.

The goddess's origins and her worship in Roman society are entirely conveyed through Isis's narration during this level. Narratively, it brings the user closer to the main task of finding her statue. In terms of usability, interaction during the level increases by combining the previously introduced grabbing mechanic with a new locomotion system: teleportation, activated by pressing a controller button. The teleportation system, the most common system in iVR experiences (Prithul et al. 2021), turns the level into an Explorative interaction experience where the user can both explore and interact with the environment. To maintain boundaries, teleportation is limited to



Fig. 4 **a** User's point of view looking towards the image in Level 2 and **b** Top-down view of Level 2 indicating the initial position of the user, the position of the image to be found, and the size of the explorable area

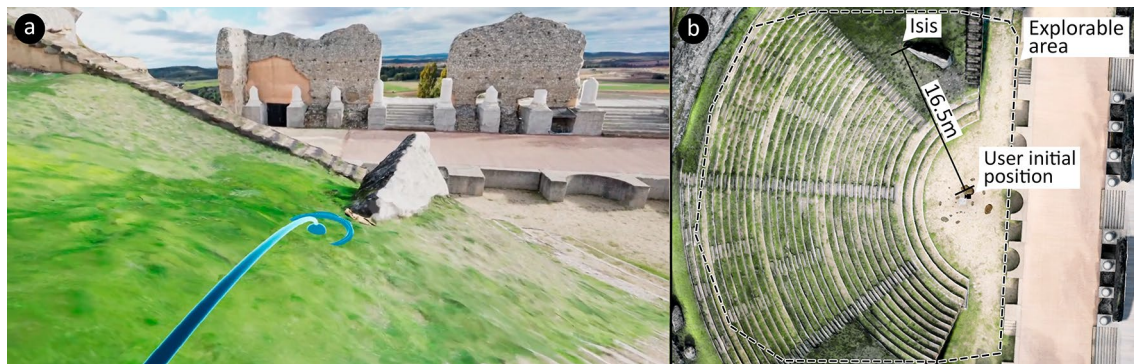


Fig. 5 **a** User's point of view near the Isis sculpture in Level 3 and **b** Top-down view of Level 3 showing the user's initial position, the location of Isis, the distance between them, and the size of the explorable area



Fig. 6 **a** User's point of view at the beginning of Level 4 and **b** Top-down view of Level 4 indicating the initial position of the user, the location of Isis and the table with the objects, and the size of the explorable area

the stage and seating area of the theater, covering about 1,250 m². Figure 5b shows a top-down view of the level with the designated movement area.

To finish the level, the user must approach the statue, with a spatialized audio cue triggered at 2m distance. Figure 5a shows the user's viewpoint close to the statue, teleporting towards it. The statue lies about 16.5m from the starting point, hidden behind a large rock visible earlier but then concealed, so as not to make the task obvious. Behavior is measured by the number of teleports used and the completion time. The level has no fixed duration, as it depends on the user searching for the statue.

3.1.5 Level 4 design

At the start of Level 4, the user is teleported to the center of the theater. The Isis statue stands in front of the user, with a table on the left displaying elements and reference images for reconstruction (Fig. 6a). Isis explains that the user must rebuild the statue using the correct parts, some of which are distractors, to complete the level.

This level explains the goddess' iconography using the reference images showing her common attributes. Interactivity is integrated into learning, as users examine images and place attributes correctly. Each correct placement prompts Isis to give the object's name and function. The level ends the narrative, with the statue fully reconstructed. In terms of usability, exploration is limited to a 4 m² area with no teleportation, but interaction is maximized, making this level an Interactive experience, where the user has a high degree of interaction but limited exploration.

The level is completed when the user places five items on the statue, of which two are incorrect. The statue, table, and objects are 1 m in front of the user, Fig. 6b shows a top-down view of the level layout. Behavior is measured by the number of reconstruction errors and completion time. The duration of the level depends on how long users take to finish, so it is not fixed.

3.1.6 Level 5 design

This level serves as an epilogue that concludes the narrative. The user returns to the same setting as in Level 1, but this time spawns facing the goddess. The statue says goodbye and informs them that it is currently located in the Museum of Burgos. Level 5 lasts 19 seconds, no data are collected, nor is the level included in any subsequent analysis.

3.2 Development of the iVR experience

The iVR experience was developed for standalone HMDs using Unreal Engine 5. The use of standalone HMDs simplifies public use by eliminating cables and external computers. A Windows PC version was also exported for versatility. Unreal Engine was chosen for the team's familiarity, ease of programming, and ability to create photorealistic environments (Bekele et al. 2018), which is an important factor in achieving a high Level of Detail (LoD) in iVR experiences.

The LoD defines how closely a 3D model resembles its real counterpart (Banfi 2021). In CH, accurate models are needed for precise observation, particularly when based on photogrammetry. Managing LoD is especially important on standalone HMDs, which have limited computational power, making optimization essential to avoid cybersickness (Schaumburg et al. 2025) and enhance user interaction (Ahn et al. 2023). The modeling procedure followed the reconstruction of the Isis of Clunia, with additional low-poly models added, notably the Clunia theatre captured via aerial photogrammetry and optimized. The Isis model was rigged for animations. Texturing used atlases to improve performance and audio was post-produced for volume equalization and effects. Finally, all assets were integrated into the development engine, using precomputed lighting for enhance performance.

The result is an optimized iVR experience that improves UX and faithfully represents heritage assets. The Blender add-on LoDCalculator, which provides an indicative score of the LoD of a 3D model, gave the models a score of 5/5 (Rodriguez-Garcia et al. 2025). Figure 7 summarizes the development and optimization process, showing phases from source acquisition to asset creation and implementation in the development engine, with sources and processes indicated by gray dashed rectangles.

3.3 Design and development of the video

The Isis of Clunia model was used to create a complementary video shown to the control group. The video has a resolution of 1920×1080 pixels and a duration of 4 minutes 10 seconds with English and Spanish subtitles for enhanced accessibility. It conveys the same knowledge as the iVR experience, previously described in Sect. 2.

The video format means that what users see and hear can be fully controlled, enabling condensed information and additional visual reinforcement. So, the video slightly expands on the historical content and reorganizes it linearly: origins and cult, iconography, and then the characteristics of the sculpture. In Fig. 8, some frames of the video can be seen. Advantages of the video format against the iVR experience include easier production, lower cost, and simpler use. However, interaction, immersion, satisfaction, and engagement are reduced compared to iVR. All video resources were created using a similar procedure to the one shown in Fig. 7, though final engine integration was replaced by editing the clips.

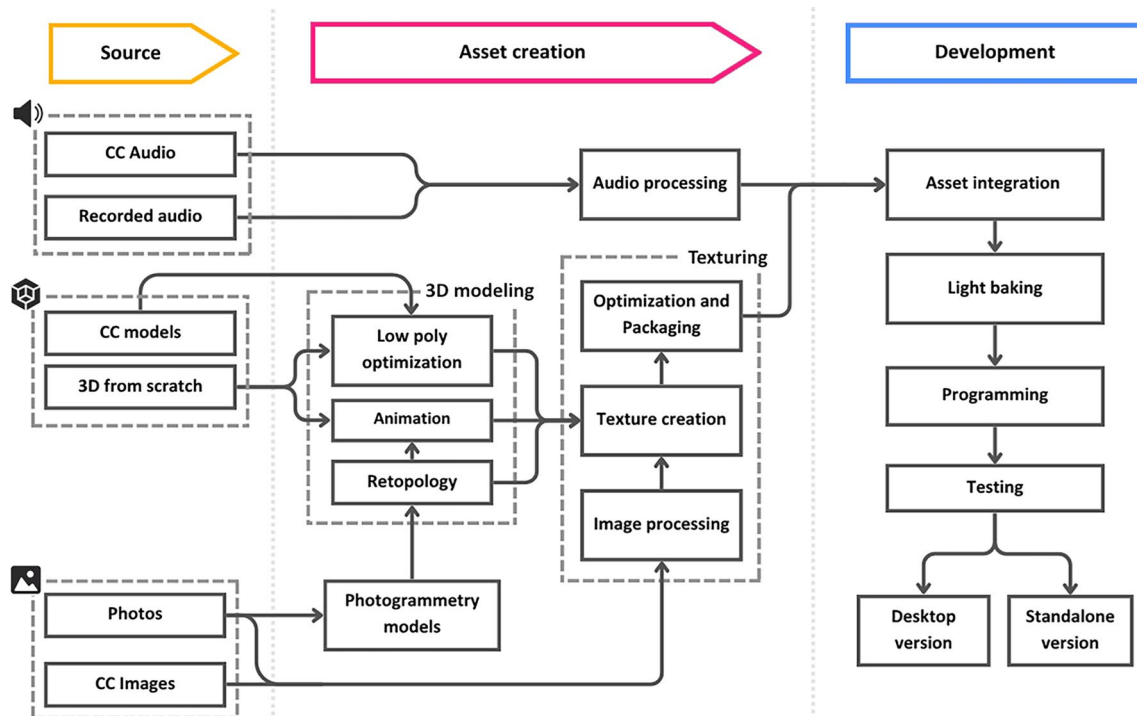


Fig. 7 Summary of the development procedure. The phases are shown horizontally. Gray dashed-line rectangles indicate the type of source

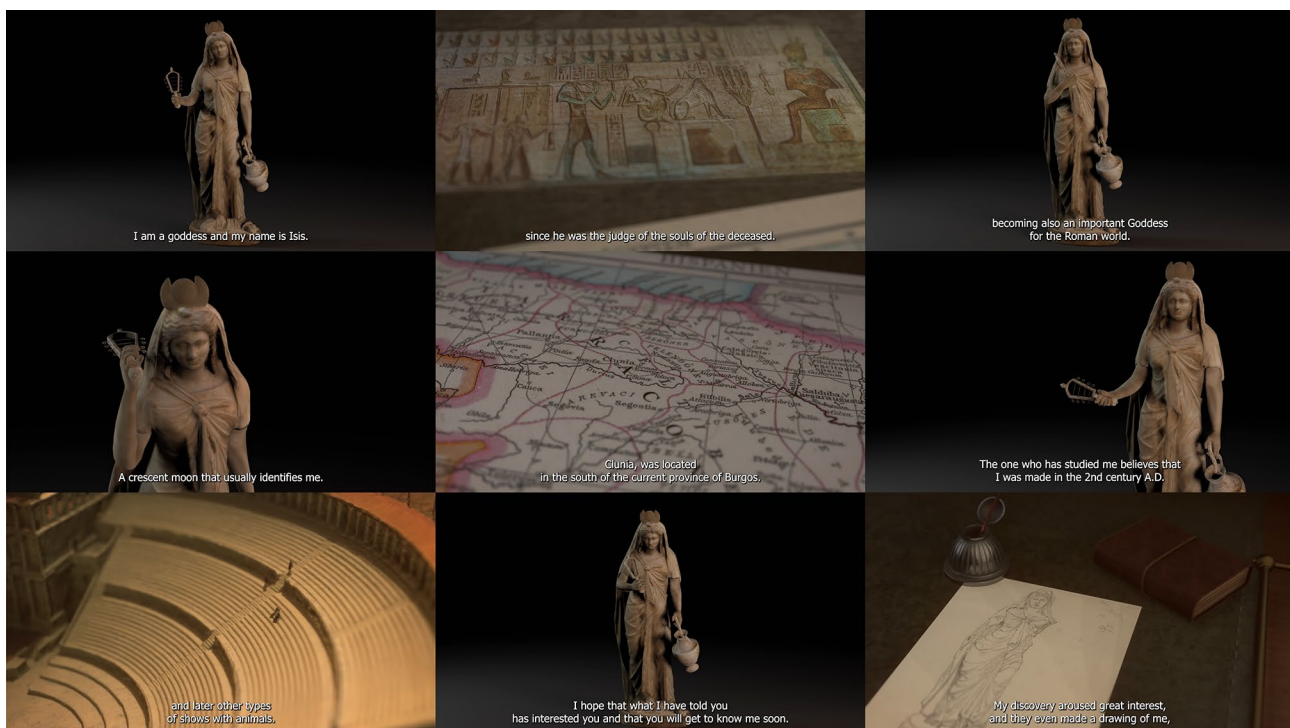
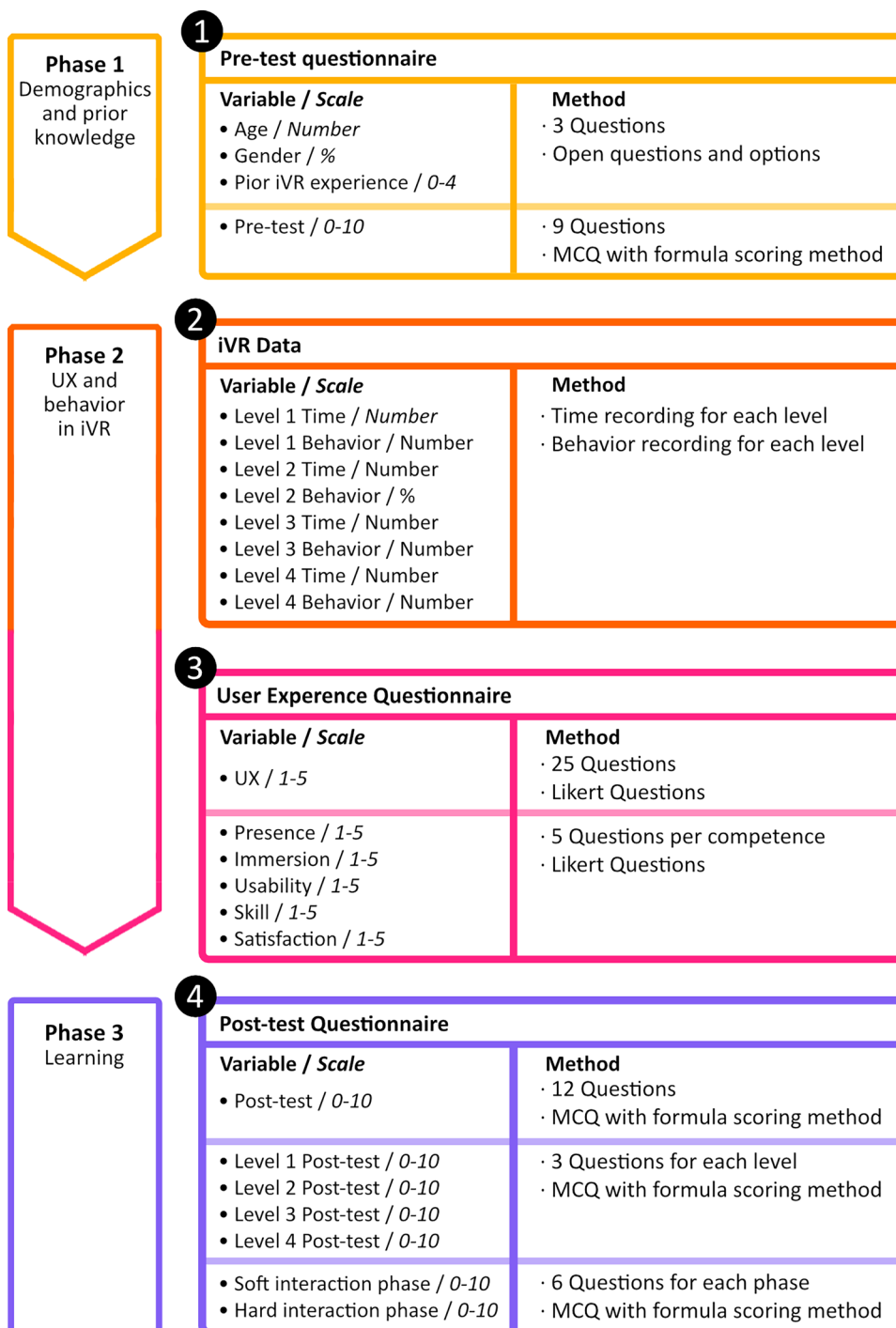


Fig. 8 Frames from the video of the Isis of Clunia with English subtitles

4 Evaluation procedure

The iVR experience and the video were evaluated using a three-phase procedure, collecting objective and subjective data before, during, and after the experience. The assessment was conducted in multiple sessions, due to the large, and demographically diverse sample. The procedure had been validated for iVR evaluation in prior research (Martinez et al. 2021). The procedures for this study were approved by the IO 41/2024 Bioethics Committee of the University of Burgos. The characteristics of each evaluation phase are summarized in Fig. 9. The characteristics

Fig. 9 Summary of the evaluation procedure. The test characteristics are shown horizontally showing the methods used in each phase, the recorded variables, and their corresponding scales, and the order of the tests and their related phases are shown vertically



of each instrument are described horizontally showing the methods used in each phase, the recorded variables, and their corresponding scales, while the order of the phases and the tests included in each one are presented vertically. In order to address the RQ, the three phases collect the key data required for their examination. Data from Phase 1 are used to demographically segment participants and assess their prior knowledge. Phase 2 gathers data related to behavior in iVR and UX. Finally, Phase 3 focuses on learning outcomes, enabling the examination of the influence of these factors across different demographic groups and their corresponding design needs. The 3 phases are described below in detail. The questionnaires used are available in Supplementary file 2. Lastly, the individual sessions are also explained.

4.1 Phase 1: demographics and prior knowledge

This phase took place just a few minutes before the iVR or video experience and collected demographic and prior knowledge data through a 12-item questionnaire. The first 3 questions covered age, gender, and prior iVR experience, with iVR experience assessed through four predefined options. The remaining 9 Multiple-Choice Questions (MCQ) tested knowledge on Isis's Egyptian origins, worship, and Roman culture. MCQs were chosen because of their suitability for assessing study-unit knowledge (Excell 2000). Formula scoring was applied to reduce guessing bias (Ndu et al. 2016), with correct answers adding 1 point and incorrect ones subtracting 1/3. Scores were normalized using Min-Max (0–10) normalization.

4.2 Phase 2: UX and behavior in iVR

During this phase, user behavior and UX data were collected exclusively for iVR participants. This phase took place during and few minutes after the iVR experience. Each of the first four levels (Level 5 served as an epilogue) was evaluated using two measures: time in seconds to complete the level and a specific behavioral indicator, resulting in a total of four behavior indicators:

- *Level 1*: The time each user took to look at the goddess. A shorter time was considered positive, as it indicated a faster reaction to the 3DoF Isis stimuli.
- *Level 2*: The percentage of activity time. Activity was defined as faster than 10 cm/s. A higher level of activity was considered positive, as it suggested a better understanding of 6DoF mechanics.
- *Level 3*: The number of teleports used. A higher number of teleports was considered positive, as it reflected greater spatial exploration, since users tended to perform a greater number of displacements with this technique (Brument et al. 2024).
- *Level 4*: The number of mistakes made when reconstructing the sculpture. Fewer mistakes were considered positive, as they indicated a better understanding of the task.

UX data were collected with a 25-item Likert questionnaire, with higher scores indicating more positive responses. This approach is commonly used for iVR usability and UX evaluation experiences (Rodriguez-Garcia et al. 2024). The questionnaire assessed five competencies; each measured with five questions: Presence, Immersion, Usability, Skill, and Satisfaction. The average of the five responses represented the value of each competence, while the overall UX score was calculated as the average of 25 responses. Questions were adapted from System Usability Scale (Brooke 1996), the Game Experience Questionnaire (Ijsselstein et al. 2023) and the Presence Questionnaire (Witmer & Singer 1998). These competencies were chosen because they reflect two key features of iVR, immersion and interactivity (Burdea & Coiffet 2003), while satisfaction is a transversal metric across different types of UX questionnaires, whether iVR-based or not. The direct use of standard questionnaires, although beneficial, would have drastically increased the number of questions required to assess all the competencies examined in this study, which represents a serious limitation in complex evaluation contexts such as museums (Rodriguez-Garcia et al. 2024).

4.3 Phase 3: learning

This phase occurred after the iVR or video experience, with timing varying by session. Learning was assessed using a 12 MCQ in a 0 to 10 scale, similar to the one used in Phase 1. In this case, each correct answer added one point out of 12, while each incorrect answer subtracted 1/3. The content of the questions was linked to the topics presented in Sect. 2: the origins and worship of the goddess, her iconography, and the formal characteristics of the Clunia sculpture. Scores were divided into four sub-scores corresponding to the historical content presented in each iVR level and in two sub-scores based on the two phases of the experience: soft interaction (Levels 1 and 2) and hard interaction (Levels 3 and 4); applying all the same formula scoring. Although the video slightly expanded content, the post-test only included material shared by both formats.

4.4 Sessions

A representative sample had to be obtained through several sessions. These sessions were organized as complementary activities within various educational programs related to history and aimed at different age groups, in order to deploy the experience in a real-world context and to balance participants' prior knowledge. Teachers integrated the activity into their curricula, introducing all participants to Roman culture prior to the experience. Participants were recruited through the teachers, who offered the activity to their students, allowing them to freely choose between the iVR or video modality. The only exception was the young age group, for whom the experience was integrated into a module on Roman history as part of their coursework. Nevertheless, participation in the study and the provision of collected data were entirely voluntary. No specific exclusion criteria were established, except for individuals who reported that they did not wish to participate in the research. Figure 10 shows images from each session, labeled with their corresponding number. The characteristics of these sessions are detailed below:

- *Session 1*: Held in November 2024 at the University of Burgos, involving 30 first-year students from the Bachelor in History and Heritage at the University of Burgos. Most of them experienced the iVR experience, while



Fig. 10 Images of users in each of the sessions. The image number corresponds to the session number

- those with some form of visual or motor disability that prevented them from fully enjoying the iVR experience (Baker et al. 2020) watched the video instead. Participation in this evaluation was part of their coursework. In this session, Phase 3 took place one week after Phase 2. This session provided data for the young demographic.
- *Session 2:* Held in March 2025 at the University of Burgos, involving 15 third-year students from the same bachelor program, also as part of their coursework. This group only experienced the iVR experience. In this session, Phase 3 was conducted one week after Phase 2. This session provided data for the young demographic.
 - *Session 3:* Held in March 2025 at a cultural center in the Burgos region, involving 42 volunteer students from art history courses for older adults who only experienced the iVR experience. In this session, Phase 3 took place a few minutes after Phase 2. This session provided data for the older demographic in the iVR group.
 - *Session 4:* Held in April 2025 at the Museum of Burgos over three consecutive days, involving 115 volunteer school students who experienced the iVR experience. A portion of the sample watched the video instead, including some with motor disabilities that prevented them from fully enjoying the iVR experience (Baker et al. 2020). In this session, Phase 3 was conducted a few minutes after Phase 2. This session contributed the data for the child demographic.
 - *Session 5:* Held in May 2025 at the Museum of Burgos, involving 8 volunteer students from art history courses for older adults who only watched the video. In this session, Phase 3 was conducted one week after Phase 2. This session provided data for the older demographic in the video group.
 - *Session 6:* Held in June 2025 at the University of Burgos, involving 5 volunteer students from the Bachelor in Video Game Design at the University of Burgos who only watched the video. In this session, Phase 3 took place a few minutes after Phase 2. This session provided data for the young demographic in the video group.
 - *Session 7:* Held in November 2025 at the University of Burgos, involving 21 first-year students from the Bachelor in History and Heritage at the University of Burgos. Most of them watched the video, while some experienced the iVR experience. Participation in this evaluation was part of their coursework. In this session, Phase 3 took place a few minutes after Phase 2 for video users and one week later for iVR users. This session provided data for the young demographic.
 - *Session 8:* Held in December 2025 at the Museum of Burgos involving 14 volunteer students from art history courses for older adults who only watched the video. In this session, Phase 3 was conducted a few minutes after Phase 2. This session provided data for the older demographic in the video group.

A one-week interval between experience and post-test was applied, whenever possible, as is recommended (Mayer 2014a and b). Initial sessions with university students meant better sample control, but other sessions occurred during one-off events, risking participant loss, particularly in Session 4 before the school holidays. Most of the sample (137 users) performed their experience at the Museum of Burgos, but some of the sessions were conducted at convenient locations for the participants, due to the logistical challenges of a multi-phase evaluation and the need to maximize user numbers. These logistical constraints prevented the same procedure from being applied consistently across all sessions, often requiring Phase 3 to be conducted only a few minutes after Phase 2. This represents a limitation that may affect the learning outcomes; therefore, the results related to learning should be interpreted with caution and in light of this constraint.

To balance the duration of the sessions across groups, the following strategy was adopted. Video sessions were conducted as group viewings, and questionnaire completion was also carried out collectively in classrooms or presentation rooms. VR sessions, by contrast, were organized in isolated rooms where participants entered in groups according to room capacity. Meanwhile, the remaining participants engaged in complementary activities related to historical dissemination.

Most video participants viewed the content using a Full HD projector with an approximate screen size of 90 inches, with audio played through external speakers. The only exception was the eight participants from Session 5, who watched the video on 24-inch Full HD monitors using headphones. All participants viewed the video with Spanish subtitles. Regarding iVR participants, the experience was tested using its desktop version in order to facilitate the collection of behavioral data. The least powerful system used was a laptop equipped with an

NVIDIA RTX 2060 graphics card, an Intel i7-8750H processor, and 16 GB of RAM. This configuration ensured a smooth experience at approximately 90 FPS for all users. All HMDs were initially calibrated using a standard interpupillary distance of approximately 63 mm. When the headset was fitted, participants were asked whether they could see clearly, and the headset height and interpupillary distance were adjusted, if necessary, particularly in the case of children. In this way, all participants were exposed to a comparable overall experience.

5 Results

The results were organized by age group, due to the age differences in the sample: Kid group (Session 4), Young group (Sessions 1, 2, 6 and 7) and Senior group (Sessions 3, 5 and 8). A middle-aged group was not included, as no participants from this demographic were available. The sample was also divided by device: iVR and video.

Of 250 evaluated users, 44 were excluded, leaving 206. Exclusions were due to missing data, copied answers, or irrational (troll) behavior, the latter identified as scoring 1 on all Satisfaction questions. Most exclusions in the Kid group resulted from data loss or copying, complicated by children's higher activity levels (Schloss et al. 2021). In this regard, participants with missing or incomplete data were excluded from the study. Troll behavior appeared in the Young group, that was called to participate as part of their class content, likely due to lack of intrinsic motivation (Ryan & Deci 2000). This phenomenon represents a limitation of the study.

Due to the large amount of data and the number of groups analyzed, and to facilitate readability, some tables are presented in a condensed format, while others are only referenced in the text. All tables are provided in Appendix and are cited as required throughout the manuscript. All tables reporting descriptive statistics consistently include the Mean (M), Standard Deviation (SD), and 95% Confidence Interval (CI). In contrast, all inferential statistics tables report the p-values (p), effect sizes (η^2), and the statistical analysis method employed. No corrections for multiple comparisons were applied due to the exploratory nature of the study based on the RQ. In any case, all inferential analyses, both p-values and effect sizes are reported and discussed to provide a more robust interpretation of the results. All data, including excluded participants and the reasons for their exclusion, are included in Supplementary file 3. Results are presented by evaluation phase.

5.1 Phase 1: demographics and prior knowledge

Table 1 shows total sample numbers, age, gender distribution, prior iVR experience, and pre-test results in terms of M and SD. iVR experience was rated from 0 to 4, while pre-test scored between 0 and 10. In its version in Appendix, the CI is also shown.

The age groups were homogeneous regarding device, with $\approx 25\%$ of users in the video condition. Although the iVR and video groups are unbalanced, this is because they are only compared in the learning outcomes, with

Table 1 Summary of demographic data divided by age group and device

Age group	Device	N	Age		Gender		Prior iVR experience		Pre-test	
			M	SD	M %	F %	M	SD	M	SD
Kid	Both	82	9.02	0.68	59%	41%	1.24	1.21	2.93	1.84
	iVR	63	8.98	0.71	54%	46%	1.33	1.26	2.94	1.92
	Video	19	9.16	0.60	74%	26%	0.95	1.03	2.90	1.60
Young	Both	59	20.40	2.62	68%	32%	1.12	1.04	6.46	1.96
	iVR	37	20.00	2.15	57%	43%	1.05	1.03	6.97	1.67
	Video	22	21.05	3.21	86%	14%	1.23	1.07	6.08	2.30
Senior	Both	65	67.80	7.40	22%	78%	0.48	0.62	5.49	2.42
	iVR	40	68.65	7.62	20%	80%	0.45	0.64	5.45	2.67
	Video	25	66.44	6.96	24%	76%	0.52	0.59	5.56	2.00

the iVR group carrying greater weight in the study. Age group sizes were sufficient for statistical analysis (≥ 20 participants), except for the video users in the Kid age group ($n = 19$).

Considering age, the average of the age groups effectively represents the three chosen demographics. Kid group averages around 9 years old. The average age of the Young group was around 20 years old, representing young adults (M. J. Kim & Kang 2023). Finally, the average age of the Senior group was over 60 years. The standard deviation increases with age, which is expected given the greater variability.

Clear trends regarding age differences emerged and are summarized in (online Fig. 11). It shows the number of users, gender distribution, average age, pre-test score, and prior iVR experience for each age group, with deviation lines indicating standard deviation. In Phase 1, standard deviation was prioritized for visualization, as the focus was on describing the sample. In subsequent analyses, figures will emphasize CI to highlight population-level patterns. Accordingly, figures will consistently show CIs, while tables will report standard deviations to complement the visual information.

Kid group showed a balanced gender ratio (41% male / 59% female), while Young group were mostly male (68%) and Senior group predominantly female (78%). Previous iVR experience is generally low across all age groups (close to 1 on the scale) meaning that participants have only used iVR once or twice before. Among the age groups, the Senior group stands out for having the lowest prior iVR experience, below 1. Finally, prior knowledge varies notably by age group, with Kids showing the lowest score, followed by Seniors, and Young participants exhibiting the highest level of prior knowledge.

5.2 Phase 2: UX and behavior in iVR

The UX results are presented in (online Fig. 12), separated by age groups. Only the iVR results are analyzed, as it was the only device evaluated in Phase 2. The X-axis first displays the overall UX score, followed by the scores for each competence. The Y-axis shows the average value of each score, with an expanded scale for better visualization. Error bars indicate the CI for each score. Significant differences between one age group and the other two, based on Analysis of Variance (ANOVA), are indicated with a diagonal line pattern. Additionally, significant differences within each age group, according to their prior iVR experience, are also marked with an “iVRXP” displayed above each corresponding bar.

The complete descriptive statistics can be found in Table 2 (Online appendix). Additionally, results of the ANOVA analysis are reported in Table 3 (Online appendix). Table 3 (Online appendix) shows the ANOVA results for age group comparisons and within-group differences by prior iVR experience. Welch’s ANOVA was employed, as Levene’s test results indicated that some variables violated the homogeneity of variance assumption (Keselman et al. 1998). Significant values are marked with an asterisk (< 0.05) and highly significant values with two asterisks (< 0.01).

Overall, UX scores were high (online Fig. 12), with a mean close to 4 out of 5 across competencies, particularly for Kid and Senior groups. The highest-rated competency across all groups was Satisfaction. In contrast, the lowest-rated competency overall was Skill, except in the case of the Kid group. The Young group was the only one that consistently exhibited lower values across all competencies, with UX scores only slightly above satisfactory (≈ 3.5). Many of those values were statistically significant with high effect sizes, as can be seen in (online Fig. 12).

These results showed that perceived UX was higher for both the Kid’s and the Senior groups, though with differing emphasis. Interaction-related competencies (Usability, Skill) were more positively rated in the Kid group, while immersion-related competencies (Immersion, and especially Presence with significant difference) were rated higher in the Senior group. The competencies most influenced by prior iVR experience were Presence, Skill and Satisfaction. The users of the Young group with less prior iVR experience significantly increased their Presence and Satisfaction, as did the members of the Senior group with Skill, in both cases with very high effect sizes.

Analysis of prior iVR experience further characterizes the age groups, showing significant differences with high effect sizes across all UX competencies, as shown in Table 3 (Online appendix). Those differences reflected

a general trend: a preference for interaction for the Kid group, and for immersion for the Senior group. The Young group, while more balanced across competencies, reported an overall lower perception of UX.

The second analysis refers to the results of behavior indicators in the iVR experience. Table 2 displays the mean and standard deviation for each age group based on level-specific behavior indicators and post-test scores. Although post-test data are not properly a behavior indicator, level-specific scores are included here due to their relevance in relation to behavior. Overall iVR results are also reported, except for behavior, given the distinct nature of these measures. In its version in Appendix, the CI is also shown. Additionally, results of the ANOVA analysis are reported in Table 5 (Online appendix). Welch's ANOVA was used due to violations of the homogeneity of variance assumption in some cases. Table 5 (Online appendix) includes the p-values and the effect sizes, for all possible age group combinations. No time data were analyzed for Level 1, as the duration was fixed for all users. Similarly, no analysis was performed on the total behavior due to the heterogeneous nature of the behavioral metrics.

The results of the iVR experience indicators are shown in (online Fig. 13) using a forest plot. For each level, the post-test, behavior, and time scores of the age groups are presented. Values were computed using Z-score normalization, represented on the X-axis, allowing indicators of different types to be compared by expressing the number of standard deviations from the mean (Pandey & Jain 2017). Given the diverse nature of the values, the scale of some indicators was reversed to standardize the interpretation in (online Fig. 13). Higher values (represented as positive σ to the right of the X-axis) were interpreted as follows:

- In the post-test, higher scores indicated greater learning outcomes and were therefore considered positive (Bichi 2016).
- For the time indicators, shorter completion times were considered favorable, as they suggest more efficient problem-solving (Checa et al. 2021).
- The behavior indicators were interpreted according to the descriptions provided in Sect. 4.2.

Statistically significant differences of a group compared to the others are marked by a diagonal line pattern.

The three age groups showed clear differences across all levels, most of statistical significance with high effect sizes, supporting earlier findings on UX. The average iVR completion time was about 8 minutes, with each level taking progressively longer due to increasingly complex interactions. Behavior indicators differed between age groups, but user behavior became more similar in the hard interaction phase (Levels 3 and 4). Also, post-test scores varied: Level 2 had lower scores than Level 1, Level 3 had higher scores, which dropped again in Level 4.

The Kid group completed levels quickly and showed high behavior indicators, especially in the soft interaction phase, but had low post-test scores. An exception occurred in Level 4, where behavior decreases but post-test performance improved compared to previous levels. Considering gender, ANOVA revealed significant gender differences in behavior for boys and girls within those levels, with boys being more active, supported by moderate effect sizes. Both the p-values and the effect size of the ANOVA analysis for behavioral indicators by gender within each age group are presented in Table 6 (Online Appendix 1). In the Young group, gender differences

Table 2 M and SD values for each age group across all level indicators, as well as the overall results of the entire experience

Age group	Indicator	Level 1		Level 2		Level 3		Level 4		Total	
		M	SD	M	SD	M	SD	M	SD	M	SD
Kid	Time	58.0	0.00	93.6	59.9	110	57.5	137	56.4	417	98.2
	Behavior	14.6	9.42	0.37	0.16	12.5	10.1	0.94	1.44	N/A	N/A
	Post-test	4.29	3.86	2.79	2.27	3.89	3.45	3.67	3.30	2.95	2.06
Young	Time	58.0	0.00	79.0	58.7	139	76.2	119	53.0	414	139
	Behavior	11.2	3.73	0.48	0.17	18.1	16.2	0.46	0.93	N/A	N/A
	Post-test	8.20	2.86	5.59	3.11	7.43	3.39	5.02	3.52	6.26	1.88
Senior	Time	58.0	0.00	142	88.6	178	113	226	127	623	212
	Behavior	19.5	14.0	0.29	0.15	12.8	15.7	0.63	1.21	N/A	N/A
	Post-test	7.92	3.00	4.45	2.93	5.04	3.91	4.42	3.62	5.04	2.12

at Level 1 were not robust due to few female participants. The Kid group's low post-test scores reflect their prior knowledge, while improved results in Level 4 may result from repeated reconstruction attempts and higher interactivity.

The Young group shows notable results across all measures, as can be seen in (Online Fig. 13); with the largest inter-group differences at Level 2. Those differences decreased at later levels, particularly by Level 4. While Level 3 showed some significant differences, these are mostly due to post-test scores with minimal effect sizes between Young and Senior groups. As the differences between the levels diminished, progression through to higher levels reduced disparities between groups.

The Senior group is characterized by lower behavior indicators and longer completion times. Although with good post-test results, they were slightly lower than the results of the Young group. In any case, despite the interaction challenge for this age group, the learning rates were not significantly affected. The largest differences with other groups again occur at Level 2, but these had diminished by Level 4 with the progression of the experience.

Considering the impact of Level 2 in the behavior of the different age groups, an Analysis of Covariance (ANCOVA) was performed to evaluate the correlation between UX and behavior indicators. The results are summarized in Table 7 (Online appendix) where age group is a fixed factor across behavior indicators and UX competencies, with Level 2 being the only level showing significant correlations with UX and Immersion.

5.3 Phase 3: learning

Learning was evaluated by comparing pre and post-tests results and searching for differences between the first (soft interaction) and the second (hard interaction) phases of the iVR experience. The learning outcomes are summarized in Tables 3 and 9 (Online appendix) and (Online Fig. 14). Table 3 (expanded in online appendix with CI values) shows the average and standard deviation for the pre-test, post-test, soft interaction phase (Levels 1–2), hard interaction phase (Levels 3–4), and the absolute learning gain relative to the pre-test. The post-test was not divided into four sections, as too few questions would produce unreliable results, especially with formula scoring. The results are presented all together and by age group, distinguishing by device (iVR or video). Table 9 (Online appendix) presents ANOVA results (p and η^2) across age group and device combinations, using Fisher's ANOVA, as the Levene's test confirmed homogeneity of variance. Finally, (Online Fig. 14) presents Phase 3 results by age group and the total. The X-axis shows columns for post-test, soft interaction, and hard interaction scores, separated by device (iVR left, video right), with pre-test scores marked by a black dashed line. The Y-axis represents the test scores and error bars indicate the CI for each score.

Some interesting conclusions can be drawn from this statistical analysis. Except for some differences between the Young and Senior groups, all post-test results showed significant differences with high effect sizes, as can be seen in Table 9 (Online appendix). Overall, scores were slightly lower than the pre-test with the soft interaction phase generally yielding higher scores than the hard interaction phase in both iVR and video conditions, possibly reflecting a higher difficulty in the latter phase. Finally, total post-test scores for iVR users were similar to the video condition overall, except where the video users scored better in the Senior group.

Absolute learning gains were positive in the soft interaction phase but turned negative in the hard interaction phase, explaining the overall slightly lower iVR results, likely due to the greater interactivity of the latter phase. However, in the hard interaction phase, Kid and Young groups still showed better learning in iVR, especially Kid group, whereas Senior group participants performed better with the video, showing even larger differences than in the soft interaction phase. The results of the ANOVA analysis pointed to more significant learning in the Kid group than in the Senior group and the Young group, though with a small effect size, as can be seen in Table 9 (Online appendix). Online Figure 14 shows that the Kid group maintained similar iVR scores across phases, while the video users had significantly lower scores, although initially better in soft interaction. Senior group users showed lower iVR outcomes, especially in the hard interaction phase.

An additional linear regression analysis was conducted to further explore the differences between both phases of the post-test, focusing especially on the iVR users, among whom the most pronounced differences and largest

Table 3 M and SD values for the total sample and for the combinations of age groups and device used in the Phase 3 results

Age group	Device	Pre-test		Post-test		Soft interaction		Hard interaction		Learning total		Learning Soft interaction		Learning Hard interaction	
		Pre-test		Post-test		Soft interaction		Hard interaction		Learning total		Learning Soft interaction		Learning Hard interaction	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Total	Both	4.80	2.60	4.92	2.71	5.67	3.03	4.33	2.91	0.12	2.64	0.87	2.92	-0.47	2.98
	iVR	4.72	2.70	4.43	2.46	5.06	2.94	3.92	2.68	-0.30	2.52	0.34	2.92	-0.80	2.88
	Video	4.97	2.39	5.97	2.93	6.97	2.82	5.19	3.20	1.00	2.70	2.00	2.62	0.23	3.09
Kid	Both	2.93	1.84	2.89	2.11	3.48	2.50	2.67	2.40	-0.04	2.58	0.54	2.72	-0.26	2.96
	iVR	2.94	1.92	2.95	2.06	3.26	2.52	2.89	2.39	0.01	2.57	0.32	2.82	-0.05	2.95
	Video	2.90	1.60	2.69	2.32	4.18	2.36	1.93	2.37	-0.21	2.66	1.28	2.26	-0.97	2.98
Young	Both	6.64	1.96	6.40	1.91	7.26	2.38	5.56	2.29	-0.24	2.28	0.62	2.78	-1.08	2.54
	iVR	6.97	1.67	6.26	1.88	6.89	2.48	5.63	2.22	-0.71	2.18	-0.08	2.72	-1.34	2.46
	Video	6.08	2.30	6.63	1.99	7.88	2.11	5.43	2.44	0.55	2.29	1.80	2.52	-0.65	2.65
Senior	Both	5.49	2.42	6.13	2.41	6.99	2.48	5.31	3.04	0.65	2.97	1.50	3.23	-0.18	3.31
	iVR	5.45	2.67	5.04	2.12	6.18	2.33	3.96	2.74	-0.40	2.73	0.74	3.24	-1.49	2.90
	Video	5.56	2.00	7.88	1.73	8.29	2.17	7.47	2.13	2.32	2.57	2.73	2.87	1.91	2.87

sample size were observed. Online Figure 15 presents two scatter plots depicting the normalized learning gain (Hake 1998) of the iVR users. The left plot illustrates learning outcomes in the soft interaction post-test, while the right plot corresponds to the hard interaction phase post-test. In each plot, the X-axis represents the pre-test score on a scale from 0 to 1, and the Y-axis shows the percentage improvement from pre-test to post-test in the different phases. Trend lines for each group were fitted using second-degree polynomial curves. A jitter of 3% was applied to enhance data visualization. The visual presentation of the scatter plots and the fit indicators of the analysis are presented in Table 10 (Online appendix), including R , R^2 , and p -values for each analysis according to age and device group. Regression analyses were conducted from pre-test to soft interaction phase, pre-test to hard interaction phase, and between both phases.

Some clear conclusions can be extracted from this linear regression analysis. In the soft interaction phase, Young and Senior groups showed a steeply decreasing trend, while Kid group declined only slightly, indicating that users with lower prior knowledge, especially in Young and Senior groups, achieved higher learning gains. Most of the curves remain above zero on the Y-axis, reflecting positive outcomes. In the hard interaction phase, Kid and Young groups patterns remained similar, but the Senior group's curve flattens, reducing gains. Linear regression confirms this shift as statistically significant for Senior iVR users. Overall, curves move closer to zero with more users below it, showing reduced learning gains. These results reinforce the previous conclusions about learning patterns in age groups. Learning outcomes tend to decline in the hard interaction phase, with this change being significant in both iVR and video users according to the linear regression analysis. While Kid and Young groups maintained their soft-interaction phase patterns, Senior group differed, with lower prior knowledge linked to poorer results and higher pre-test scores associated with better outcomes.

6 Discussion

The discussion will be organized according to the three evaluation phases. Subsequently, a set of best practices will be presented, taking into account the age groups and the outcomes of this study. Finally, the limitations of the research will be discussed.

6.1 Phase 1: demographics and prior knowledge

The three age groups adequately represent the selected demographics. First, the Kid group is situated within the “concrete operational” cognitive stage, characterized by sufficient logical and cognitive abilities but a need for many sensory stimuli (Tengfei & Guanjie 2024), capable of using iVR devices, although supervision is required (Rauschenberger & Barakat 2020). Gender distribution is balanced, which is expected given that participants were recruited from a primary school class, and their low level of prior experience with iVR is consistent with their young age. Finally, their few prior knowledge aligns with the educational stage to which they belong.

The age of the Young group is centered around the early twenties, although it shows a gender imbalance in favor of male participants. This imbalance is likely due to recruitment from a smaller and more specialized class, where gender bias resulting from class size is more probable. Their low level of iVR experience, together with their higher level of prior knowledge, can be explained by the educational context of the group. As undergraduate students in the humanities, they are expected to possess greater prior knowledge and might express less interest in iVR technology than students studying in other fields (Assem & Morsi 2025).

Finally, the age of the Senior group is centered around 60 years, which according to the World Health Organization can be classified as elderly (Lin et al. 2018). The gender imbalance toward female participants, as well as their level of prior knowledge, can also be explained by the educational context of the group, as participants were recruited in courses for older adults which tended to be mostly female students (FORMOSA, 2014). Finally, their low level of prior experience with iVR is not unexpected, as older adults tend to exhibit greater skepticism toward new technologies (Al Khalifah et al. 2025).

6.2 Phase 2: UX and behavior in iVR

UX results in this phase were positive, particularly regarding satisfaction, which aligns with common findings in iVR research (Krupljanin et al. 2025). Despite this, Skill was the lowest-rated competence, likely because it reflects the frequency and variety of interactions within the iVR experience, which may be challenging for some users who require greater adaptation time (Abeele et al. 2021). Regarding behavior, one result shared across groups was that the time spent at each level increased progressively, probably due to increasingly complex interactions. However, user behavior became more similar during the hard interaction phase (Levels 3 and 4), likely as a consequence of a waning novelty effect (Koch et al. 2018). In addition to these common results, each age group exhibited distinct patterns in both UX and behavior within the iVR experience.

The Kid age group showed a clear preference for interactivity, rating interaction-related competencies, Usability and Skill, more positively. Their particularly high value in Skill is noteworthy, as this was the lowest-rated competency overall. This is not surprising given children's greater tendency toward interaction (Bailey & Schloss 2024) and exploration (Muravevskaia & Gardner-McCune 2023). These tendencies were reflected in their iVR behavior, as they completed and adapted to the levels quickly, exhibiting high behavioral indicators, especially in the earlier, movement-based levels. This higher level of activity, particularly among boys, has been supported by previous research reporting similar findings (van Dun et al. 2021). Although their post-test results were generally lower, their improved performance at Level 4 may result from repeated reconstruction attempts, which can enhance learning through agency and interaction (Makransky & Petersen 2021).

The Young group stood out as the only group that consistently exhibited lower values across all competencies, with UX scores only slightly above satisfactory. Several factors may explain this divergence. First, technology anxiety may have been more pronounced in this group, as participants reported lower levels of prior iVR experience (Al Khalifah et al. 2025). In addition, this demographic typically shows lower motivational levels compared to children (Cadet & Chainay 2021) and older adults (Dong et al. 2023). Finally, UX perceptions among the Young group, who participated as part of their coursework, may have been negatively influenced by a lack of intrinsic motivation (Ryan & Deci 2000). On the other hand, within this group, lower prior experience with iVR appeared to favor higher Presence and Satisfaction scores as a result of a novelty effect, which tends to benefit more passive competencies due to fewer interaction difficulties compared to other age groups (Dong et al. 2023). Regarding iVR behavior, this group exhibited short completion times and high behavioral indicators across all levels, particularly standing out at Level 2. These results are consistent with previous research indicating that young adults generally face fewer difficulties when interacting in iVR environments (Dong et al. 2023), partly because they represent the most extensively studied demographic in iVR research (M. J. Kim & Kang 2023).

The peak differences observed at Level 2 reflect a strong novelty effect, highlighting the importance of gradually introducing new interactions to support user acclimation (Abeele et al. 2021). However, frequent changes in interaction types appeared to trigger repeated novelty effects, suggesting that limiting the number of interaction systems may be beneficial.

Finally, the Senior age group stood out for a stronger preference for passive competencies, particularly Immersion and Presence. These findings align with prior research highlighting older users' preference toward Immersion and Presence at the expense of more interactive competencies (Lorenz et al. 2023), especially in environments with naturalistic aesthetics such as the theater of Clunia (Abeele et al. 2021). In addition, prior iVR experience in the Senior group, who already exhibited high levels of Presence (Schaumburg et al. 2025), had a stronger impact on competencies related to interactivity. Regarding iVR behavior, this group was slower in completing the levels and showed lower behavioral indicators, which improved with increased time spent in the iVR experience. Although their post-test results were good, they were slightly lower than those of the Young group. This outcome was expected, as higher prior knowledge generally leads to better learning outcomes (Dengel & Magdefrau 2020); however, the results indicate that despite interaction-related challenges in iVR, the potential for improvement among older users remains high (Chen & Or 2017). Taken together, these findings suggest that older users can achieve favorable outcomes in hard interaction settings, but they require careful introduction and adequate

acclimatization time (Ijaz et al. 2022). For shorter iVR experiences, a simpler design emphasizing presence and immersion appears more suitable.

6.3 Phase 3: learning

Learning outcomes were satisfactory, with post-test scores comparable to those of the pre-test, which can be considered a positive result. This was foreseeable, as the post-test covered more advanced topics, such as Isis's iconography, whereas the pre-test focused on her origins and general aspects of Roman culture (Bichi 2016). In addition, scores in the soft interaction phase were generally higher than in the hard interaction phase, most likely due to the increased complexity of Levels 3 and 4 and a renewed novelty effect (Miguel-Alonso et al. 2024). Furthermore, questions in the soft interaction phase often addressed Isis's Egyptian origins, a more familiar topic, thereby introducing a difficulty bias. Regarding learning outcomes by device, clear differences were observed, with generally higher scores in the video condition, particularly during the soft interaction phase. However, this difference was much less pronounced in the hard interaction phase. It is important to consider the differences in exposure time across phases, as these may have contributed to the observed device-related differences, which were approximately one point on a ten-point scale. In any case, this pattern, like the other observed differences, was more strongly modulated by age group.

To illustrate these differences and link them to the previous results, (Online Fig. 16) presents a series of interaction plots. The bottom of the figure shows the interaction between learning outcomes in the soft and hard interaction phases, separated by device, for each age group. The Y-axis represents learning scores. The top of the figure illustrates the interaction between behavioral indicators and learning outcomes across the soft and hard interaction phases for each age group. Z-score normalization was applied to both variables to enable comparison. In all interaction plots, the CI range is displayed using error bars. The following paragraphs further explore these age-related differences.

The Kid group exhibited relatively low learning outcomes, which is expected given their young age. Although their overall learning gains were modest, they started from a similarly low baseline. However, this group showed higher learning outcomes in the hard interaction phase under the iVR condition, highlighting the benefits of increased interactivity for this age group and supporting prior findings on children's enhanced learning through interactive experiences (Zhao et al. 2019). In contrast, learning outcomes in the soft interaction phase were lower for this group, with better results observed in the video condition. As shown in (Online Fig. 16), under the iVR condition learning remained relatively stable across phases, whereas a sharp decrease was observed in the video condition, indicating an interaction effect between phase and device. Furthermore, when behavioral indicators are considered, behavior decreased during the soft interaction phase but increased substantially during the hard interaction phase, a pattern that differs from that observed in the other age groups and highlights the importance of learning through interaction and agency for this age group (Makransky & Petersen 2021).

The Young group achieved the highest post-test scores; however, their learning gains were not the highest overall, remaining slightly below those of the Senior group. Nevertheless, it should be considered that improvement is more difficult when starting from a higher baseline. This group followed the general pattern of better learning outcomes in the soft interaction phase and lower performance in the hard interaction phase, with higher learning scores in the video condition; particularly during the soft interaction phase. In this specific case, these results should be interpreted with caution, as iVR participants in this group completed the post-test one week after the experience, whereas video participants completed it after only a few minutes. This timing difference may have influenced the results, potentially favoring the video condition. As shown in (Online Fig. 16), a slight interaction between device and interaction phase can be observed, with a more pronounced decline in learning outcomes for the video condition across phases. Regarding the interaction between behavior and learning across phases, both variables exhibited a relatively similar decline, suggesting that interaction did not have a clear effect on learning for this group, which can also be observed in (Online Fig. 15). This may be explained by young adults' high adaptability to iVR environments (M. J. Kim & Kang 2023). In this sense, both interaction phases appear suitable

for this age group, with observed differences primarily attributable to the increased difficulty of the hard interaction phase (Bichi 2016).

The Senior group achieved the second-highest post-test scores, very close to those of the Young group, while showing the highest learning gains overall. This outcome was mainly driven by participants in the video condition, who achieved the highest learning scores in both the soft and hard interaction phases, indicating that this technology is particularly suitable for users within this demographic. When examining iVR users within the Senior group, positive learning outcomes were observed in the soft interaction phase, which became negative in the hard interaction phase, consistent with known interaction-related challenges among older adults (Chen & Or 2017). Nevertheless, their stronger performance during the iVR soft interaction phase can be explained by a preference for passive, presence-focused experiences (M. J. Kim & Kang 2023), suggesting that the first half of the iVR experience was more appropriate for this group. Interaction plots in (online Fig. 16) indicate a steeper decline from the soft to the hard interaction phase for iVR users. While behavioral indicators showed a slight increase during the soft interaction phase, they decreased during the hard interaction phase. Furthermore, for the Senior group, the convergence in performance observed in the later phases (Online Fig. 15), together with the linear regression analysis of participants with lower prior knowledge, suggests a homogenization of the group. This pattern is likely due to limited learning in iVR, causing post-test performance to depend more strongly on prior knowledge and reducing the novelty effect. This stabilization led to lower motivation among participants with less prior knowledge, whereas those with higher prior knowledge adapted more effectively to the interaction. In any case, higher prior knowledge accentuated the differences observed among older participants during levels involving greater interaction demands. In any case, low levels of interaction appear to be beneficial for learning in this age group, whether through less interactive iVR experiences or via other formats such as video.

6.4 Best practices

The iVR experience revealed some usability issues linked to its experimental design, which prioritized research requirements over commercial polish, despite being intended for a museum-like context. These issues were mainly linked to the repeated introduction of new interaction systems, creating a strong novelty effect, especially evident in Level 2, where differences were most pronounced. A more accessible design would rely on fewer, well-introduced interaction types. Even so, user UX scores were high, and post-test results were favorable, often surpassing those of the video condition, indicating that the design was successful and effective for knowledge dissemination. However, for greater effectiveness, the design should also consider the specific needs of different age groups (Hughes et al. 2017). The three age groups showed significant differences across all evaluation phases (previously discussed), but all shared low of prior iVR experience, an important factor for mass audiences, who usually lack such exposure (Puig et al. 2020). Nonetheless, as the iVR experience advanced, group differences tended to diminish, suggesting that in longer sessions the novelty effect fades (Koch et al. 2018).

Additionally, the iVR experience revealed significant differences compared to the video, consistent with findings from previous research (Wyssenbach et al. 2025). These differences were most pronounced when leveraging iVR's defining features, immersion and interactivity, with the appropriate age groups, namely Kid and Senior groups. It highlights the importance of tailoring experiences to specific demographics groups and ensuring that iVR content is developed with high LoD and optimization to maximize immersion (Dengel & Magdefrau 2020) and interactivity (Ahn et al. 2023). Since the lack of realism in iVR experiences is considered one of the shortcomings in their design (Tuena et al. 2020). When these conditions are not adequately met, more established and cost-effective formats, such as video, may prove more suitable. In fact, this study found that learning outcomes in the soft interaction phase were generally higher with video. Although iVR is often associated with greater user satisfaction, video outperformed it in certain aspects, such as soft interaction learning. Given that 3D models, commonly employed as foundational resources, can serve multiple purposes, it is advisable to provide a video alternative (or another inclusive format) alongside iVR experiences designed for broad audiences, particularly for reasons of accessibility. In this study, 5% of participants had to rely on video due to disabilities preventing the use

of iVR. While more accessible iVR designs are possible, they often diverge substantially from traditional formats intended for mass audiences (Pedroza-Santiago et al. 2025).

The main findings are summarized in (Online Fig. 17) using a forest plot. For the iVR experience, scores for time, learning in hard and soft interaction, UX and behavior are shown separated by age group are presented. Values were computed using Z-score normalization, similarly to (Online Fig. 13). Given the diverse nature of the values, the scale of some indicators was reversed to standardize the interpretation. Positive values are represented as positive σ to the right of the X-axis.

Considering the results of this research, the following series of best practices have been identified for the design and development of iVR experiences aimed at a mass audience and adapted to different age groups:

- When working with 3D models, it is essential to leverage their flexibility to develop complementary and more accessible experiences.
- Using iVR experiences with a mass audience requires more extensive planning due to the increased number of challenges involved.
- For an iVR experience to be effective and preferable over more affordable and accessible technologies, it is necessary to capitalize on its immersive and interactive features.
- It is advisable to focus the iVR experience on a few well-introduced mechanics to avoid inducing novelty effects in users.
- Immersion has proven to be a key factor for all age groups, making it essential to ensure high LoD and optimization in iVR experiences.
- The longer the duration of an iVR experience, the less noticeable the differences between users become, in terms of UX perception and learning outcomes.
- User-adapted design is more effective than a universal approach; however, effective universal experiences can still be created by carefully considering general usability principles.
- High levels of interaction are suitable for children, but due to their agitation, it is necessary to moderate interaction and introduce educational content only after they have acclimated to the experience, especially among boys.
- Young people can adapt to different interaction systems, so the design can be more flexible with this age group.
- Older adults take significantly longer to complete experiences, so it is preferable to design shorter experiences while accounting for their longer completion times.
- Older users prefer and learn more from soft interaction experiences with naturalistic aesthetics; nevertheless, if sufficient time is provided for proper adaptation, more complex experiences can be developed.

The results of this study are not only relevant within the field of CH, which served as the application domain, but also have broader applicability to any area aiming to engage multiple demographics. While they are particularly useful in CH due to the wide audience targeted by such applications (Gatto et al. 2025), the findings can be applicable to other domains as well. Beyond CH, fields with high potential for these insights, due to their frequent use with children, young adults and older adults, include education (Pellas et al. 2021) and rehabilitation (Tuena et al. 2020). In education, the findings on experience design and its impact on learning have direct application, particularly for child and young adult demographics, which are often the primary target of such interventions (Pellas et al. 2021). Conversely, insights regarding older adults may be especially valuable in rehabilitation, where applications are typically designed for this population and often rely on non-interactive experiences despite users' preferences for interactivity (Ijaz et al. 2022). The guidelines derived from this study to improve user experience are particularly relevant in this context, especially regarding the integration of interactivity, one of the main challenges for this age group (Chen & Or 2017). It should be noted, however, that this study was conducted with healthy participants. Many applications in these fields are designed for individuals with advanced health conditions, for whom these results may not be directly applicable (Baker et al. 2020).

6.5 Limitations

Despite the contributions of this study, several limitations need to be considered when interpreting the findings.

First, although an effort was made to cover a wide demographic range in terms of age, certain age groups were not included and therefore not analyzed. Specifically, adolescents and middle-aged adults were excluded due to the inability to recruit participants from these age ranges. Including these groups could have strengthened the study's results and provided a more comprehensive understanding of age-related differences in iVR experiences.

Additionally, a potential gender bias should also be considered. This study focused on analyzing demographic groups separately and comparing their results. For this reason, although the overall sample was relatively balanced in terms of gender, clear imbalances were present in the Young and Senior groups. These disparities are attributable to the educational context of the groups, which is often gender-skewed (FORMOSA, 2014), and represent a demographic limitation that should be addressed in future research. Although no clear gender differences were detected in this study, this may be due to the uneven distribution of participants across groups.

Moreover, the use of the VA, while useful in this study for constructing the narrative of the experiences, could have been enhanced by employing an autonomous agent (Wei et al. 2025). Such agents have been shown to provide greater immersion and encourage higher engagement, even among less participative users (Oger et al. 2025). This limitation arises from the level design, which was primarily focused on exploring the most common mechanics and interactions in iVR experiences. In any case, this aspect of the design was not fully leveraged, and its impact on different demographic groups was not explored.

The difference in timing during Phase 3 of the evaluation represents a major limitation of this study, as administering the learning questionnaire either minutes or a week after the experience can significantly affect the results. This bias was taken into account when interpreting the data. It arose due to logistical constraints of the evaluation: since the study was conducted in real-world environments, the date, access time, and location were limited by the participants' availability, and most sessions took place outside the laboratory (Quirós-Ramírez et al. 2025). This presented challenges for conducting multiple phases consistently.

For similar reasons, an adapted questionnaire was prioritized over standard UX questionnaires. The study analyzed multiple competencies, which would have required several separate questionnaires if standard instruments had been used. This could have led to participant survey fatigue (Jeong et al. 2023), in addition to further complicating the logistics of the evaluation, which was sometimes conducted in non-laboratory settings with time constraints.

7 Conclusions

In this study, the design and evaluation of a CH iVR experience at the Museum of Burgos has been presented. The experience was directed at children, young adults, and elderly, with a video group serving as a control. The iVR experience included four progressively complex levels, so that users could gradually engage with various interaction systems. The design balanced duration, motivation, novelty effect, and potential iVR side effects, supported by a narrative led by the VA, the goddess Isis of Clunia, and careful optimization. The iVR experience and the video were evaluated using a sample of 206 participants distributed across three age groups (Kid, Young, and Senior) through a three-phase assessment: prior knowledge and demographic characteristics, perceived UX and data obtained from the iVR experience and post-experience learning outcomes. The evaluation yielded the following findings.

- The Kid group responded enthusiastically to high interactivity, with particularly strong behavioral engagement among the boys rather than the girls. They completed tasks rapidly and interacted confidently. Although their learning outcomes with soft interaction were higher with the video, they ultimately benefited from greater

- interactivity in terms of both learning and satisfaction. Their high motivation and quick adaptation suggest that interactive formats are suitable for this group, provided that cognitive demands are introduced progressively.
- The Young group showed the highest adaptability and the shortest completion times across all levels, demonstrating consistent behavioral performance regardless of interaction type. These attributes indicate a strong potential for engaging with a wide range of iVR designs. However, their UX ratings were more moderate, possibly due to reduced novelty perception or a less involved attitude during testing.
 - The Senior group displayed a clear preference for soft interaction and tended to place greater value on immersive rather than interactive competencies, which aligned with their higher learning outcomes in such contexts. Although they required more time to complete tasks, they showed notable improvements over time, underscoring the importance of slower paced design. Older users can benefit substantially when iVR is properly adapted to their preferences, even in more interactive formats.
 - Comparisons between iVR and video indicated that, if the advantages of more accessible and familiar formats such as video are to be surpassed, then the key strengths of iVR, namely immersion and interactivity, must be intentionally and effectively leveraged through optimization and design. Video remains a valuable and practical alternative whenever those conditions cannot be ensured, or when addressing audiences with accessibility constraints. Nevertheless, tailoring experiences to user characteristics will yield better learning outcomes and engagement. Universally designed experiences can also succeed when they are grounded in strong usability principles and carefully considered design strategies.

These guidelines are not only relevant for Cultural Heritage, but also support the broader dissemination of iVR, particularly in fields such as education and rehabilitation, which often target the same demographics analyzed in this study. Future research will be focus on exploring the design patterns identified within each age group to develop experiences that are specifically tailored to their needs. In that regard, further investigation will also address strategies for the rapid reduction of the novelty effect during iVR experiences, a factor that has proven to be highly relevant for facilitating user adaptation across different demographic profiles. To this end, the use of fuzzy logic to adapt iVR environments based on performance-related factors can be highly valuable for developing systems that adjust to users' varying needs in real time (Maroungkas et al. 2025).

Appendix

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10055-026-01409-z>.

Acknowledgements This work is part of the Teaching Innovation Group "Patrimonio internacional y nuevas tecnologías" (PAINT) of the University of Burgos, whose members have helped in data acquisition. Special thanks to the Museum of Burgos for sharing their historical expertise and facilitating the evaluation of the experience.

Author contributions All authors contributed to the conception and design of the study; D. C. and A. B. were responsible for the methodology and provided guidance; I. M.A. contributed to preparing the material and collecting the data; B. R.G. performed the data analysis and visualization. All authors wrote, reviewed, and approved the final manuscript.

Funding Funding was provided by the Ministerio de Ciencia e Innovación, TED2021-129485B-C43, CPP2022-009724, NextGenerationEU, CPP2022-009724, TED2021-129485B-C43

Data availability All Supplementary files are provided at the following link: <https://xrailab.es/2025/09/18/annex/>. Supplementary file 1 includes a video with the iVR experience and the control group video. Supplementary file 2 includes the questionnaires used in the research in a .docx file. Supplementary file 3 includes the data used in this research in a spreadsheet.

Declarations

Conflict of interest The authors declare no competing interests.

Consent to participate All participants provided their informed consent to participate, in accordance with the standards of the relevant ethics committees (Reference Number: IO 41/2024 Bioethics Committee of the University of Burgos).

Ethical approval All procedures involving human participants were in accordance with the ethical standards of the Institutional Review Board of the U.S. Army Research Laboratory and with the 1964 Helsinki declaration and its later amendments.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abeele VV, Schraepen B, Huygelier H, Gillebert C, Gerling K, Van Ee R (2021) Immersive virtual reality for older adults. *ACM Trans Access Comput.* 14(3):1–30. <https://doi.org/10.1145/3470743>
- Ahn D-K, Bae B-C, Kim Y (2023) User experience of a digital fashion show: exploring the effectiveness of interactivity in virtual reality. *Appl Sci* 13(4):2558. <https://doi.org/10.3390/app13042558>
- Al Khalifah E, Hammady R, Abdelrahman M, Al-Shamaileh O, Marghany M, El-Jarn H, Darwish A, Kurt Y (2025) Technology anxiety in virtual reality adoption: examining the impact of age, past experience, and cybersickness. *IEEE Access* 13:71858–71879. <https://doi.org/10.1109/ACCESS.2025.3562383>
- Alrashidi M, Wadey CA, Tomlinson RJ, Buckingham G, Williams CA (2023) The efficacy of virtual reality interventions compared with conventional physiotherapy in improving the upper limb motor function of children with cerebral palsy: a systematic review of randomised controlled trials. *Disabil Rehabil* 45(11):1773–1783. <https://doi.org/10.1080/09638288.2022.2071484>
- Arroyo de la Fuente MA (2002) EL CULTO ISIACO EN EL IMPERIO ROMANO. CULTOS DIARIOS Y RITUALES INICIÁTICOS: ICONOGRAFÍA Y SIGNIFICADO. *Boletín De La Asociación Española De Egiptología* 12:207–232
- Assem H, Morsi W (2025) A framework for the design of effective learning environments in the AI era: Egyptian students' preferences from humanities versus engineering disciplines. *J Eng Sci* 53(3):319–345. <https://doi.org/10.21608/jesaun.2025.347140.1397>
- Avilés-Castillo F, Buele J, Palacios-Navarro G (2025) Virtual reality and user experience: current trends and future challenges. *IEEE Access* 13:55939–55956. <https://doi.org/10.1109/ACCESS.2025.3554434>
- Bailey JO, Schloss JI (2024) Knowing versus doing: Children's social conceptions of and behaviors toward virtual reality agents. *Int J Child-Comput Interact.* 40:100647. <https://doi.org/10.1016/j.ijcci.2024.100647>
- Baker S, Waycott J, Robertson E, Carrasco R, Neves BB, Hampson R, Vetere F (2020) Evaluating the use of interactive virtual reality technology with older adults living in residential aged care. *Inf Process Manag* 57(3):102105. <https://doi.org/10.1016/j.ipm.2019.102105>
- Banfi F (2021) The evolution of interactivity, immersion and interoperability in HBIM: digital model uses, VR and AR for built cultural heritage. *ISPRS Int J Geo-Inf.* 10(10):685. <https://doi.org/10.3390/ijgi10100685>
- Bekele MK, Pierdicca R, Frontoni E, Malinverni ES, Gain J (2018) A survey of augmented, virtual, and mixed reality for cultural heritage. *J Comput Cult Herit.* 11(2):1–36. <https://doi.org/10.1145/3145534>
- Bichi, A. A. (2016). CLASSICAL TEST THEORY: An introduction to linear modeling approach to test and item analysis. *Int J Soc Stud*, 02(09), 27–33. <https://edupediapublications.org/journals>
- Boletsis C, Chasanidou D (2022) A typology of virtual reality locomotion techniques. *Multimodal Technol Interact.* 6(9):72. <https://doi.org/10.3390/mti6090072>
- Briones-Villafuerte, G., Naula-Bone, A., Vaca-Cardenas, M., & Vaca-Cardenas, L. (2022). User Interfaces Promoting Appropriate HCI: Systematic Literature Review. *Revista Ibérica de Sistemas e Tecnologias de Informação*, 61–76.
- Brooke, J. (1996). SUS: A “Quick and Dirty” usability scale. In *Usability Evaluation In Industry*. <https://www.researchgate.net/publication/228593520>

- Brument, H., Zhang, R., & Kaufmann, H. (2024). Influence of Virtual Reality Setup on Locomotion Technique Usage during Navigation with Walking, Steering and Teleportation. *ICAT-EGVE 2024 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. <https://doi.org/10.2312/egve.20241377>
- Burdea, G., & Coiffet, P. (2003). Virtual reality technology. *presence: teleoperators and virtual environments*, 12(6), 663–664. <https://doi.org/10.1162/105474603322955950>
- Cadet LB, Chainay H (2021) How preadolescents and adults remember and experience virtual reality: The role of avatar incarnation, emotion, and sense of presence. *Int J Child-Comput Interact*. 29:100299. <https://doi.org/10.1016/j.ijcci.2021.100299>
- Checa D, Bustillo A (2020) Advantages and limits of virtual reality in learning processes: Briviesca in the fifteenth century. *Virtual Real* 24(1):151–161. <https://doi.org/10.1007/s10055-019-00389-7>
- Checa D, Miguel-Alonso I, Bustillo A (2021) Immersive virtual-reality computer-assembly serious game to enhance autonomous learning. *Virtual Real*. <https://doi.org/10.1007/S10055-021-00607-1>
- Chen J, Or C (2017) Assessing the use of immersive virtual reality, mouse and touchscreen in pointing and dragging-and-dropping tasks among young, middle-aged and older adults. *Appl Ergon* 65:437–448. <https://doi.org/10.1016/j.apergo.2017.03.013>
- Dengel, A., & Magdefrau, J. (2020). Immersive Learning Predicted: Presence, Prior Knowledge, and School Performance Influence Learning Outcomes in Immersive Educational Virtual Environments. *2020 6th International Conference of the Immersive Learning Research Network (ILRN)*, 163–170. <https://doi.org/10.23919/iLRN47897.2020.9155084>
- Dhalmahapatra K, Maiti J, Krishna OB (2021) Assessment of virtual reality based safety training simulator for electric overhead crane operations. *Saf Sci* 139:105241. <https://doi.org/10.1016/j.ssci.2021.105241>
- Dong Y, Liu X, Tang M, Huo H, Chen D, Du X, Wang J, Tang Z, Qiao X, Guo J, Fan L, Fan Y (2023) Age-related differences in upper limb motor performance and intrinsic motivation during a virtual reality task. *BMC Geriatr* 23(1):251. <https://doi.org/10.1186/s12877-023-03970-7>
- Excell PS (2000) Experiments in the use of multiple-choice examinations for electromagnetics-related topics. *IEEE Trans Educ* 43(3):250–256. <https://doi.org/10.1109/13.865196>
- Faust, A. (2021). The effects of gamification on motivation and performance. In *The Effects of Gamification on Motivation and Performance*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-35195-3>
- Ferdani D, Fanini B, Piccioli MC, Carboni F, Vigliarolo P (2020) 3D reconstruction and validation of historical background for immersive VR applications and games: The case study of the Forum of Augustus in Rome. *J Cult Herit* 43:129–143. <https://doi.org/10.1016/j.culher.2019.12.004>
- Fiocco AJ, Millett G, D'Amico D, Krieger L, Sivashankar Y, Lee SH, Lachman R (2021) Virtual tourism for older adults living in residential care: A mixed-methods study. *PLoS ONE* 16(5):e0250761. <https://doi.org/10.1371/journal.pone.0250761>
- FORMOSA, M. (2014) Four decades of Universities of the Third Age: past, present, future. *Ageing Soc* 34(1):42–66. <https://doi.org/10.1017/S0144686X12000797>
- Forssell, M., Paavilainen, J., Hassan, L., & Turunen, M. (2024). Older adults' playful experiences of VR gaming. *ACM International Conference on Interactive Media Experiences*, 182–191. <https://doi.org/10.1145/3639701.3656309>
- Gatto C, Barba MC, Chiarello S, Corchia L, Faggiano F, Nuzzo BL, Sumerano G, De Luca V, De Paolis LT (2025) Breaking the barriers: Extended reality and innovative technologies for enhanced accessibility of the Ceramics Museum of Cutfiano. *Digit Appl Archaeol Cult Herit*. 36:e00400. <https://doi.org/10.1016/j.daach.2025.e00400>
- Hake RR (1998) Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am J Phys* 66(1):64–74. <https://doi.org/10.1119/1.18809>
- Healy D, Flynn A, Conlan O, McSharry J, Walsh J (2021) Older adults' experiences and perceptions of immersive virtual reality: a protocol for a systematic review and thematic synthesis. *Int J Qual Methods*. <https://doi.org/10.1177/16094069211009682>
- Hughes S, Warren-Norton K, Spadafora P, Tsotsos L (2017) Supporting optimal aging through the innovative use of virtual reality technology. *Multimodal Technol Interact*. 1(4):23. <https://doi.org/10.3390/mti1040023>
- Ijaz K, Tran TTM, Kocaballi AB, Calvo RA, Berkovsky S, Ahmadpour N (2022) Design considerations for immersive virtual reality applications for older adults: a scoping review. *Multimodal Technol Interact* 6(7):60. <https://doi.org/10.3390/mti6070060>
- Ijsselstein, W., De Kort, Y., & Poels, K. (2023). The Game Experience Questionnaire.
- Jeong D, Aggarwal S, Robinson J, Kumar N, Spearot A, Park DS (2023) Exhaustive or exhausting? Evidence on respondent fatigue in long surveys. *J Dev Econ* 161:102992. <https://doi.org/10.1016/j.jdeveco.2022.102992>
- Keselman HJ, Huberty CJ, Lix LM, Olejnik S, Cribbie RA, Donahue B, Kowalchuk RK, Lowman LL, Petoskey MD, Keselman JC, Levin JR (1998) Statistical practices of educational researchers: an analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Rev Educ Res* 68(3):350–386. <https://doi.org/10.3102/00346543068003350>
- Kim MJ, Kang Y (2023) Older adults' user experience of virtual tourism: exploring presence and experiential value with respect to age difference. *Virtual Real* 27(4):2967–2987. <https://doi.org/10.1007/s10055-023-00849-1>
- Kim YM, Rhiu I, Yun MH (2020) A systematic review of a virtual reality system from the perspective of user experience. *Int J Hum Comput Inter* 36(10):893–910. <https://doi.org/10.1080/10447318.2019.1699746>

- Koch, M., Von Luck, K., Schwarzer, J., & Draheim, S. (2018). The Novelty Effect in Large Display Deployments-Experiences and Lessons-Learned for Evaluating Prototypes. *European Conference on Computer-Supported Cooperative Work*. https://doi.org/10.18420/ecscw2018_3
- Kouijzer MMTE, Kip H, Bouman YHA, Kelders SM (2023) Implementation of virtual reality in healthcare: a scoping review on the implementation process of virtual reality in various healthcare settings. *Implement Sci Commun*. 4(1):67. <https://doi.org/10.1186/s43058-023-00442-2>
- Krupljanin N, Alink LRA, van der Voort A, Struijk Wilbrink MR, Bergwerff CE (2025) The feasibility of an immersive interactive virtual reality task for children and adolescents. *Int J Child-Comput Inter*. 43:100722. <https://doi.org/10.1016/j.ijcci.2025.100722>
- Lee L-K, Wei X, Chui KT, Cheung SKS, Wang FL, Fung Y-C, Lu A, Hui YK, Hao T, U LH, Wu N-I (2024) A Systematic Review of the Design of Serious Games for Innovative Learning: Augmented Reality, Virtual Reality, or Mixed Reality? *Electronics (Basel)* 13(5):890. <https://doi.org/10.3390/electronics13050890>
- Lim, D., Shirai, S., Orlosky, J., Ratsamee, P., Uranishi, Y., & Takemura, H. (2022). Evaluation of User Interfaces for Three-Dimensional Locomotion in Virtual Reality. *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, 1–9. <https://doi.org/10.1145/3565970.3567693>
- Lin C-S, Jeng M-Y, Yeh T-M (2018) The elderly perceived meanings and values of virtual reality leisure activities: a means-end chain approach. *Int J Environ Res Public Health* 15(4):663. <https://doi.org/10.3390/ijerph15040663>
- Looi C-K, Wong L-H, So H-J, Seow P, Toh Y, Chen W, Zhang B, Norris C, Soloway E (2009) Anatomy of a mobilized lesson: Learning my way. *Comput Educ* 53(4):1120–1132. <https://doi.org/10.1016/j.compedu.2009.05.021>
- Lorenz M, Brade J, Klimant P, Heyde C-E, Hammer N (2023) Age and gender effects on presence, user experience and usability in virtual environments—first insights. *PLoS ONE* 18(3):e0283565. <https://doi.org/10.1371/journal.pone.0283565>
- Makransky G, Petersen GB (2021) The cognitive affective model of immersive learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. *Educ Psychol Rev* 33(3):937–958. <https://doi.org/10.1007/s10648-020-09586-2>
- Marougkas A, Troussas C, Krouska A, Sgouropoulou C (2025) An adaptive virtual reality game for programming education using fuzzy cognitive maps and pedagogical models. *Smart Learn Environ*. 12(1):62. <https://doi.org/10.1186/s40561-025-00392-3>
- Martinez K, Menéndez-Menéndez MI, Bustillo A (2021) Awareness, prevention, detection, and therapy applications for depression and anxiety in serious games for children and adolescents: systematic review. *JMIR Serious Games* 9(4):e30482. <https://doi.org/10.2196/30482>
- Mayer, R. E. (2014). *Computer Games for Learning*. In *Journal of Pedagogic Development* (Vol. 5, Number 2). The MIT Press, Cambridge. <https://doi.org/10.7551/mitpress/9427.001.0001>
- Richard E. Mayer. (2014). *The Cambridge Handbook of Multimedia Learning* (R. E. Mayer, Ed.). Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9781139547369>
- Mehrfard A, Fotouhi J, Taylor G, Forster T, Navab N, Fuerst B (2019) Virtual reality technologies for clinical education: evaluation metrics and comparative analysis. *Comput Methods Biomech Biomed Eng Imaging vis*. <https://doi.org/10.1080/21681163.2020.1835559>
- Miguel-Alonso I, Checa D, Guillen-Sanz H, Bustillo A (2024) Evaluation of the novelty effect in immersive Virtual Reality learning experiences. *Virtual Reality* 28(1):27. <https://doi.org/10.1007/s10055-023-00926-5>
- Miguel-Alonso, I., Rodriguez-Garcia, B., Checa, D., & De Paolis, L. T. (2022). Developing a tutorial for improving usability and user skills in an immersive virtual reality experience. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*: 13446 LNCS. https://doi.org/10.1007/978-3-031-15553-6_5
- Muñiz Grijalvo E (2021) El lugar del culto de Isis en las ciudades de la península ibérica. *Hispania Sacra* 73(147):11–17. <https://doi.org/10.3989/hs.2021.001>
- Muravevskaia E, Gardner-McCune C (2023) Designing a virtual reality empathy game framework to create empathic experiences for children. *Int J Child-Comput Inter*. 35:100561. <https://doi.org/10.1016/j.ijcci.2022.100561>
- Ndu IK, Ekwochi U, Di Osuorah C, Asinobi IN, Nwaneri MO, Uwaezuoke SN, Amadi OF, Okeke IB, Chinawa JM, Orjioko CJG (2016) Negative marking and the student physician—a descriptive study of Nigerian medical schools. *J Med Educ Curric Dev*. <https://doi.org/10.4137/JMECD.S40705>
- Oger A, Gorisse G, Fleury S, Christmann O (2025) MemorIA, an architecture for creating interactive AI historical agents in educational contexts. *Comput Anim Virtual Worlds*. <https://doi.org/10.1002/cav.70032>
- Pandey A, Jain A (2017) Comparative analysis of KNN algorithm using various normalization techniques. *Int J Comput Netw Inf Secur*. 9(11):36–42. <https://doi.org/10.5815/ijcnis.2017.11.04>
- Partarakis N, Zabulis X (2024) A review of immersive technologies, knowledge representation, and AI for human-centered digital experiences. *Electronics* 13(2):269. <https://doi.org/10.3390/electronics13020269>
- Pedroza-Santiago EA, Quiroz-Ibarra JE, Bojorges-Valdez ER, Padilla-Castañeda MÁ (2025) Comparison of manual, automatic, and voice control in wheelchair navigation simulation in virtual environments: performance evaluation of user and motion sickness. *Sensors* 25(2):530. <https://doi.org/10.3390/s25020530>

- Pellas N, Mystakidis S, Kazanidis I (2021) Immersive virtual reality in K-12 and higher education: a systematic review of the last decade scientific literature. *Virtual Reality* 25(3):835–861. <https://doi.org/10.1007/s10055-020-00489-9>
- Pinch G (2004) *Egyptian myth: a very short introduction*. Oxford University Press. <https://doi.org/10.1093/actrade/9780192803467.001.0001>
- Porcino, T., Trevisan, D., & Clua, E. (2020). Minimizing cybersickness in head-mounted display systems: causes and strategies review. *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*, 154–163. <https://doi.org/10.1109/SVR51698.2020.00035>
- Printezis P, Koutsabasis P (2026) Spatial augmented reality storytelling in arts and culture: a critical review from an interaction design perspective. *Heritage* 9(1):20. <https://doi.org/10.3390/heritage9010020>
- Prithul A, Adhanom IB, Folmer E (2021) Teleportation in virtual reality; a mini-review. *Frontiers in Virtual Reality*. <https://doi.org/10.3389/frvir.2021.730792>
- Puig A, Rodríguez I, Arcos JL, Rodríguez-Aguilar JA, Cebrián S, Bogdanovych A, Morera N, Palomo A, Piqué R (2020) Lessons learned from supplementing archaeological museum exhibitions with virtual reality. *Virtual Real* 24(2):343–358. <https://doi.org/10.1007/s10055-019-00391-z>
- Quirós-Ramírez MA, Feineisen A, Streuber S, Reips U-D (2025) Virtual reality experiments in the field. *PLoS ONE* 20(4):e0318688. <https://doi.org/10.1371/journal.pone.0318688>
- Rauschenberger, R., & Barakat, B. (2020). Health and safety of VR use by children in an educational use case. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 878–884. <https://doi.org/10.1109/VR46266.2020.00010>
- Rodríguez-García B, Guillen-Sanz H, Checa D, Bustillo A (2024) A systematic review of virtual 3D reconstructions of cultural heritage in immersive virtual reality. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-024-18700-3>
- Rodríguez-García B, Miguel-Alonso I, Guillen-Sanz H, Bustillo A (2025) LoDCalculator: A level of detail classification software for 3D models in the Blender environment. *SoftwareX* 30:102107. <https://doi.org/10.1016/j.softx.2025.102107>
- Rodríguez-García B, Corchia L, Zaparaín Yáñez MJ, De Paolis LT (2026) The virtual reconstruction of a statue of isis unearthed at the Roman City of Clunia (Spain): A 3D model for interactive virtual reality experiences. *J Comput Cult Herit*. 19(1):1–23. <https://doi.org/10.1145/3774911>
- Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol* 55(1):68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Ryokai K, Jacobo S, Rivero E, Park J (2022) Examining children’s design processes, perspective-taking, and collaboration when using VR head-mounted displays. *Int J Child-Comput Interact*. 33:100451. <https://doi.org/10.1016/j.ijcci.2021.100451>
- Salomón, R. (1853). *Estatua romana de Clunia*. https://bibliotecavirtualmadrid.comunidad.madrid/bvmadrid_publicacion/e/catalogo_imagenes/grupo.do?path=1031511
- Samala, A. D., Rawas, S., Rahmadika, S., Criollo-C, S., Fikri, R., & Sandra, R. P. (2025). Virtual reality in education: global trends, challenges, and impacts—game changer or passing trend? In *Discover Education* (Vol. 4, Number 1). Discover. <https://doi.org/10.1007/s44217-025-00650-z>
- Schaumburg M, Imtiaz A, Zhou R, Bernard M, Wolbers T, Segen V (2025) Immersive virtual reality for older adults: Challenges and solutions in basic research and clinical applications. *Ageing Res Rev* 109:102771. <https://doi.org/10.1016/j.arr.2025.102771>
- Schloss, I., O. Bailey, J., & Tripathi, S. (2021). “I’m in his belly!”: Children’s Responses to Different Types of Characters in Virtual Reality. *Interact Des Child*. <https://doi.org/10.1145/3459990.3460723>
- Schofield, G., Beale, G., Beale, N., Fell, M., Hadley, D., Hook, J., Murphy, D., Richards, J., & Thresh, L. (2018). Viking VR: Designing a Virtual Reality Experience for a Museum. *Proceedings of the 2018 Designing Interactive Systems Conference*, 805–815. <https://doi.org/10.1145/3196709.3196714>
- Silva, M., & Teixeira, L. (2022). *eXtended Reality (XR) Experiences in Museums for Cultural Heritage: A Systematic Review* (pp. 58–79). https://doi.org/10.1007/978-3-030-99188-3_5
- Slater M, Navarro X, Valenzuela J, Oliva R, Beacco A, Thorn J, Watson Z (2018) Virtually being Lenin enhances presence and engagement in a scene from the Russian Revolution. *Front Robot AI*. <https://doi.org/10.3389/frobt.2018.00091>
- Staneva, A., Ivanova, T., Rasheva-Yordanova, K., & Borissova, D. (2023). Gamification in education: building an escape room using VR Technologies. *2023 46th MIPRO ICT and Electronics Convention (MIPRO)*. <https://doi.org/10.23919/MIPRO57284.2023.10159923>
- Tawfik DS, Sinha A, Bayati M, Adair KC, Shanafelt TD, Sexton JB, Profit J (2021) Frustration with technology and its relation to emotional exhaustion among health care workers: cross-sectional observational study. *J Med Internet Res* 23(7):e26817. <https://doi.org/10.2196/26817>
- Tcha-Tokey, K., Christmann, O., Loup-Escande, E., Richir -, S., & Richir, S. (2016). Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *Int J Virtual Real*. 16, 33–48.
- Tengfei, X., & Guan jie, Y. (2024). The Interdisciplinary Integration of the Concept of Child-friendly Museums and Virtual Reality Technology. *9th International Conference on Economics, Management, Law and Education (EMLE 2023)*. <https://doi.org/10.26914/c.cnkihy.2023.081371>
- tom Dieck, M. C., Jung, T., & Michopoulou, E. (2019). *Experiencing Virtual Reality in Heritage Attractions: Perceptions of Elderly Users* (pp. 89–98). https://doi.org/10.1007/978-3-030-06246-0_7

- Tuena C, Pedroli E, Trimarchi PD, Gallucci A, Chiappini M, Goulene K, Gaggioli A, Riva G, Lattanzio F, Giunco F, Stramba-Badiale M (2020) Usability Issues of Clinical and Research Applications of Virtual Reality in Older People: A Systematic Review. *Front Hum Neurosci*. <https://doi.org/10.3389/fnhum.2020.00093>
- van Dun C, van Kraaij A, Wegman J, Kuipers J, Aarts E, Janzen G (2021) Sex differences and the role of gaming experience in spatial cognition performance in primary school children: an exploratory study. *Brain Sci* 11(7):886. <https://doi.org/10.3390/brainsci11070886>
- Vlahovic S, Suznjevic M, Skorin-Kapov L (2022) A survey of challenges and methods for Quality of Experience assessment of interactive VR applications. *J Multimodal User Interfaces* 16(3):257–291. <https://doi.org/10.1007/s12193-022-00388-0>
- Wei S, Freeman D, Rovira A (2025) Virtual humans in virtual reality mental health research: systematic review. *JMIR XR Spat Comput* 2:e75087–e75087. <https://doi.org/10.2196/75087>
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 225–240. <https://doi.org/10.1162/105474698565686>
- Wyssenbach T, Kaufmann K, Schwaninger A (2025) Segmentation influences learning: a study of knowledge acquisition through virtual reality and 2D video with airport security screeners. *Virtual Real* 29(3):107. <https://doi.org/10.1007/s10055-025-01186-1>
- Xiao Y, Hew KFT (2024) Intangible rewards versus tangible rewards in gamified online learning: Which promotes student intrinsic motivation, behavioural engagement, cognitive engagement and learning performance? *Br J Educ Technol* 55(1):297–317. <https://doi.org/10.1111/bjet.13361>
- Zaparaín Yáñez, M. J. (2022). *ESTATUA DE ISIS. CLUNIA*. <https://catedrapatrimonioubu.com/project/estatua-de-isis/>
- Zhao, W., Devine, K., & Gardner, H. (2019). Fantasy Gaming and Virtual Heritage. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1279–1280. <https://doi.org/10.1109/VR.2019.8797715>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Bruno Rodriguez-Garcia¹ · Ines Miguel-Alonso¹ · David Checa¹ · Andres Bustillo¹

✉ Andres Bustillo
abustillo@ubu.es

Bruno Rodriguez-Garcia
brunorg@ubu.es

Ines Miguel-Alonso
imalonso@ubu.es

David Checa
dcheca@ubu.es

¹ Department of Computer Engineering, University of Burgos, Avda Cantabria s/n, 09006 Burgos, Spain