A Literature Review of User Studies in Extended Reality Applications for Archaeology

Michele De Bonis, Huyen Nguyen, Patrick Bourdot

A Literature Review of User Studies in Extended Reality Applications for Archaeology

Michele De Bonis Huyen Nguyen Patrick Bourdot

Université Paris-Saclay, CNRS, LISN, VENISE team, Orsay, France

ABSTRACT

In the present study we conducted a systematic review on user studies for Archaeology in eXtended Reality of the last 10 years. After a screening and selection process, 52 articles were selected for an in-depth analysis. Their classification follows different axes: devices, location dependency, type of users, interaction and collaboration. We also organised the existing user studies according to tasks, evaluation measurements, number of participants, and how the study was conducted (pre-test and/or post-test, formative and summative evaluation, quantitative and qualitative data). We found an intertwined relation between Archaeology and Cultural Heritage, which is reflected in the vast presence of applications for museum exhibitions and tours on archaeological sites. Similarities between systems developed for archaeologists and for general public were also investigated. Our purpose was to find a common ground between different user studies that could help designers of the next systems have a base on which they can build their system. We also highlighted which would be the preferred and most suitable evaluation techniques, when they are needed, with the type of users to address. The results show a heterogeneity of measurable variables and possible choices, but some guidelines could be derived.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality / Mixed or augmented reality; Applied computing—Physical sciences and engineering—Archaeology

1 INTRODUCTION

The areas of application of Virtual and Augmented Reality (VR/AR) have grown extensively in the last two decades. In the field of archaeology there are numerous benefits that could derive from the use of these immersive technologies. All the phases of the archaeological process can be supported: from the excavation phase [69] to the test of plausible hypotheses [42], or from the reconstruction of old buildings and artefacts, to their presentation to the public [19].

Many VR/AR applications in archaeology aim to enrich and enhance visitors' experience during a museum exhibition or an excursion in place. In addition, they make it possible for the visitors to explore inaccessible, hard-to-reach places [13, 26, 34, 41, 44, 47, 53, 56, 68, 75], and to recreate sites and artefacts that were destroyed long years ago [35].

There are not only occasional visitors and tourists who would benefit from VR/AR technologies, but also local people who can gain a better knowledge of the past of the area they live in. Even archaeologists could use these platforms to train themselves before the actual expedition or to re-explore the ruins after the field work. Indeed, archaeological excavations are a destructive process, so documenting every step in the first place can be useful to formulate hypotheses at a later time thanks to immersive exploration of some 3D simulations. Therefore, VR/AR platforms can be fruitful for exploration, learning and teaching, as a more immediate way to create a relationship with a distant past. They are thus helpful to better apprehend history of a ruin or an archaeological site and to increase the users' knowledge about it.

The way of interacting in a VR/AR application can determine its success or failure. So, it is essential to determine whether using VR/AR technology facilitates immersive experience and, more generally, whether its allows the application to fit better with the users' needs via user studies. However, conducting an analysis of a large number of VR/AR applications in archaeology and gaining insights into their features as well as their qualities and drawbacks are a complex task. There are many factors that play the roles in the final performance of a VR/AR system. For instance, there are interaction design and its implementation, adaptability of VR/AR to the purpose of such technology in interacting with archaeological artefacts, experimental design, and user study conduct. Finding a common ground from the previously conducted user studies in VR/AR for archaeology could help the future designers and developers to have a guideline to follow and to be sure that they would lead to a significant result. Since the users are almost all the time at the centre of the investigation, it is important to clarify which kind is the designated one for VR/AR systems, as well as the contiguous fields of study that are covered, when designing a VR/AR system for archaeology. For clarification, this paper uses the term eXtended Reality (XR) to target either VR or Mixed Reality immersive experiences, according to Milgram's reality-virtuality continuum [55].

Our ultimate goals are to give an overview of the current research in XR for archaeology by highlighting papers that make an important impact from different respective factors. In order to identify research gaps and future research opportunities, we reformulate the aforementioned issues in the following research questions:

- RQ1: What are the related fields of study or other subjects to be aware of in the design and development of an XR-based system for archaeology?
- RQ2: What is the best way or the best recommendable methodology to evaluate an XR archaeology system? Could a standardised model of analysis be proposed for evaluation of XR systems for archaeology?
- RQ3: Are there interaction techniques or platforms that can target different types of users (general public and experts)?

These questions are formulated to orient the literature reviewing process, and provide significant insights into the current state of research, the related areas, the most used methodology, and the type of users which are addressed. They should give an overall direction for the future work in XR system for archaeology.

To give an answer to these questions, we conducted a systematic review of the published scientific literature of the last 10 years (2012-2022), following the PRISMA guidelines [60]. The period of the last decade was judged sufficient to have a comprehensive look on the evolution of the field and exclude obsolete techniques.

The paper is organised as follows. Section 2 introduces the context of this literature review and describes the method taken in collecting the data for this paper. Section 3 details the results from the analysis of the existing XR applications in archaeology, dwelling carefully on user studies. Finally, we derive some guidelines and discuss open problems in Section 4 and 5.

2 CONTEXT AND METHOD

The advances in XR technology have dramatically changed the human-machine interaction design methodology in the field of

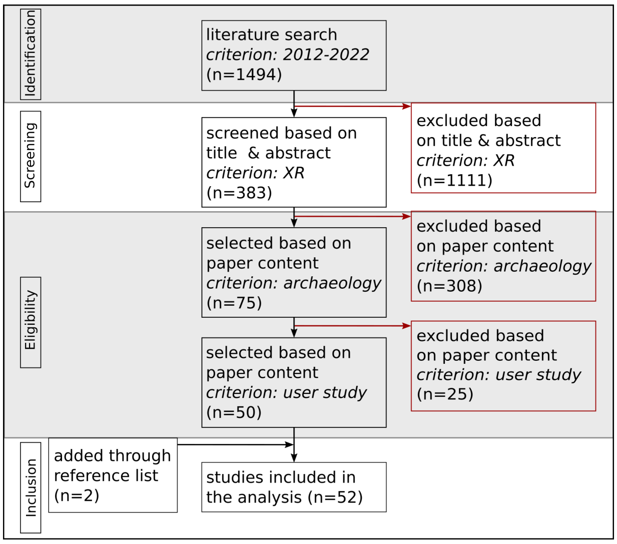

Figure 1: Flow diagram of the selection and screening process.

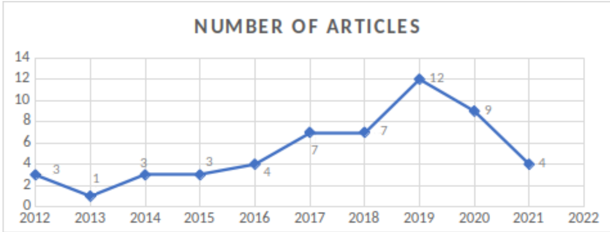

Figure 2: Number of articles published each year on the application of XR in archaeology with complete user study.

Cultural Heritage (CH). With the numerous number of works published in this domain, the urgent need to have a clear classification has arisen (cf. [46]). In this context, there are previous literature reviews [49, 70, 71]. Specifically, [49] highlights the advantages of having multisensorial stimuli in VR/AR experiences, but emphasises significant limitations in the standardisation of system evaluation. [70] concludes that most works focus on system design and development. Besides, [71] points out the benefits of the use of AR to improve students' motivation to learn in CH and underlines some challenges such as usability, content creation, information overload, and improper use of technology. However, how the XR domain has contributed to archaeology, especially when the interaction and user experiment reside at the heart of such systems, rests largely unexplored.

Even for our current study, most of the investigated articles present a strong connection with CH. Indeed, archaeology and CH are strongly intertwined, especially regarding the study and the preservation of the material remains from a close or remote past [15]. Therefore, a large extent of what archaeology studies can be seen as a form of tangible cultural heritage [72]. But the stratigraphic method and excavation techniques are proper only to archaeology, in the same manner that CH includes also the "intangible expression of human culture (cosmology, folklore, and oral histories)" [72], which are only indirectly related to archaeology. Consequently, in this review, we discern their complex and intertwining relationship while trying to bring forth the particular features of application of XR in archaeology.

In order to situate our review in the current context of this intersection of the two domains (i.e., XR and archaeology), a first literature search was performed in December 2020, followed by an updated one in April 2022. The papers to be included had to be published between 2012-2022 in English. Several digital libraries, search engines and databases were consulted: ACM Digital Library, IEEE Xplore Digital Library, Science Direct, Proquest, Springer Link, Eurographics Digital Library, MIT Press Direct, and Google Scholar. The search keywords were: ("augmented reality" OR "mixed reality" OR "virtual reality" OR "extended reality") AND ("archaeology" OR "virtual archaeology" OR "cultural presence" OR "virtual heritage" OR "cultural heritage"). The reason for the latter search terms was to avoid omitting the papers which put forward CH more than archaeology even though they deal with both of them. All the 1494 papers collected went through the process described in Fig. 1.

In this paper, we narrow down the scope of the systems to the works involve the use of dedicated XR devices such as head-mounted displays, projection-based systems or AR-based interfaces. For this reason, the articles talking about 3D content or its interaction on a desktop interface were not included. Therefore, in the screening phase, we selected the works which employed some devices or developed interaction techniques adapted for an XR experience based on their title and abstract. 383 papers in total passed this screening process.

In the next phase of eligibility, the whole content of each article was first screened to verify whether it is directly related to archaeology. We excluded thus 308 papers based on this criterion. We repeated the same process to check whether the article presents the results of a user study. Accordingly, all the articles with an uncompleted user study or no results mentioned were excluded, which left only 50 eligible articles. Next in the inclusion phase, we filtered the list of references from each paper to collect those that meet the same criteria as mentioned above. We included two more relevant articles to the list. At the end of the whole process, 52 articles were targeted for our final analysis.

All the databases and libraries were thoroughly researched to have the most exhaustive bibliography. However, there is a chance that some articles may have been skipped if they had different keywords than the ones used in the search engines or if they did not explicitly mentioned archaeology.

3 Results

To be able to answer the research questions, we analysed the 52 selected papers based on different factors. In the following section, we present the results of the classification of these articles into different axes in terms of devices, location, targeted users, interaction, and collaboration. For each category, a high-level overview will be presented followed by some highlighting of the research papers that are representative or make an impact in the category.

The process of classification was derived from the analysed papers to have the most fitting categories that can potentially be applied even to future works. The separation based on device criterion was functional to the article selection process because it hints at the presence of XR in such systems. The location and targeted-user categories answer the questions of who and where. Meanwhile, the interaction and collaboration categories give more information about how the existing technologies were employed. With these categories we hope to give a broad view of the field. Besides, Section 3.8 analyses in depth the user studies conducted in the selected papers. The number in each category may not add up to the total number of the articles because they have been counted in a cumulative way. For instance, if an article mentions two interaction techniques, it would fall into two separate categories and be counted twice.

3.1 Number of Publications

As it can be seen in Fig. 2, there has been a slow increase in the number of published articles for the topic in the course of the last decades. It shows a growing tendency of research interest in using XR for archaeology which peaked in 2019. There was a setback in 2020 and 2021, which is most likely due to the Covid-19 pandemic, making user studies almost impossible to be conducted in person. There is no data on the 2022 publication as it is still early to have any published articles on this topic.

3.2 Devices

In order to get insights into the types of display devices used in different user studies, we categorised them into *user–centric* and room-scale real-world environment (or *room–centric*) group [11] (see Fig. 3). The term *user–centric* implies the XR devices that put the focus on the user such as VR/AR headsets (e.g., HTC

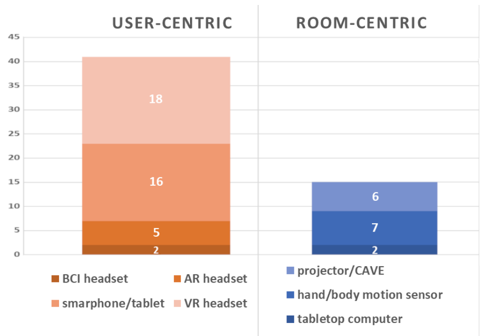

Figure 3: Number of articles employed room–centric and user–centric devices and platforms.

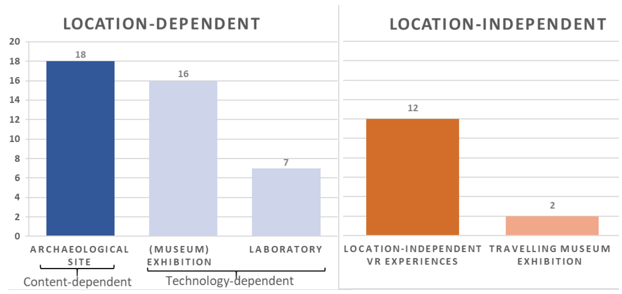

Figure 4: Number of articles categorised into location–dependent and –independent groups.

VIVE, Oculus Rift, Microsoft HoloLens, Google Cardboard, Samsung Gear VR) and mobile devices used for AR. Indeed, for the latter, mobile phones and tablets have been largely used in many applications for the AR exploration of archaeological sites. Brain-computer technologies are also considered, although their diffusion is not as wide as XR headsets. The *user–centric* devices are a popular choice in the selection of the XR platforms to use, making 73% of the whole published works with XR headsets and mobile devices taking the lead. On the other hand, *room–centric* setup includes room–scale devices that are embedded in the environment surrounding the user. This is the case of devices such as stereoscopic projectors or CAVE-like systems [22], hand/body sensing devices, and tabletop computers. There are only 27% of papers using or developing interaction techniques on *room–centric* platforms.

User–centric The user–centric devices assure a better portability. That may be a reason why the number of studies employing them is higher than room–centric ones. Mobile devices are often strictly related to onsite visits. Some examples include the historical site of the Viking fortress of Aggersborg [38], the prehistoric rock art paintings in Cova dels Cavalls [8], and the reconstruction of the Cisneros Marketplace in Medellin [35]. When developing AR-based mobile application for outdoor settings, there are several issues that system designers and researchers have to consider such as the problem of brightness, limited battery usage, and overheating of devices when used during summer time [47].

AR headsets (Microsoft HoloLens to be specific) have been used to enhance museum exhibitions, e.g., MuseumEye in the Egyptian Museum in Cairo [32, 33], and in the Museo archeologico de la Almoina [52]. Many existing AR interfaces using headsets have been developed to provide archaeologists with measurement and annotation tools [3, 31].

VR headsets are generally adopted for the reconstruction and exploration of ancient sites as they were in the past or for the documentation of their present state. Amongst others, there are 800-year-old Yuan Dynasty site [20], Neolithic site of Çatalhöyük [65], Choirokoitia [19], Hera II Temple of Paestum [12] and Itapeva Rocky Shelter [9]. In addition, there are systems which allow archaeologists to become accustomed to excavation techniques even on sites that have already been excavated and gain field experience remotely [26, 69]. VR technology can be used for the visit to hard-to-reach places [13, 16, 53].

Another innovative interaction approach in user–centric platform is non-invasive Brain-Computer Interfaces (BCI). This type of interfaces is based on the measurement of brain waves through electroencephalogram (EEG) [48]. A BCI headset has been used to pilot the tour of the 3D reconstruction of the city of Rome in 340 A.D. [73], or to measure participants' response through EEG while visiting the reconstruction of the underwater site of Baiae [76].

Always in the field of underwater archaeology, there is one example of marine AR experience to visit the site of submerged ancient roman Villa con ingresso a protiro, in Baiae [75]. It is one of the first experiments that evaluates user experience using a tablet-based interface for AR undersea.

Room–centric Room–centric systems can provide alternative ways of interaction with the digital artefacts in archaeology. Amongst them, projection-based systems are often used in experiments to compare different approaches of interaction in XR environments in order to find the most intuitive or natural one (for instance, compared to VR headset [42, 46]). They are occasionally employed in conjunction with 3D-printed reproduction of ancient artefacts [30, 67]. Besides, CAVE environments have been used to support marine archaeology hypotheses (e.g., [41]).

Body motion sensors are sometimes valued as the right solution for an exhibition because the showroom can provide all the necessary space. For example, in the Etruscanning project, the user can use body gestures to move in the virtual reconstruction of Etruscan tombs [62]. This type of platform has been used to test navigation techniques in the *domus olearia* "a Roman country house and olive farm dedicated to oil production" [5, 6]. Similarly, in the MayaArch3D project [66], a virtual reconstruction of the ancient Maya city of Copan, can be explored. Interestingly, the work in [37] uses breath recognition through the body movement as a way to control the interaction.

Tabletop interface has been used for Valcamonica analysis of petroglyphs, combing projector and VR together [42]. This platform was also employed in the analysis of crowd behaviour to study the positioning of burial sites around Stonehenge [18]. In general, these two works are the only ones applying the tabletop-oriented interaction amongst all the selected papers.

3.3 Location Dependency

By analysing the designated location for the experience to take place, we can acquire more information on the purpose of the experience itself. As seen in Fig. 4, a first distinction can be made between location–dependent and –independent systems. The location–independent systems usually provide VR experiences that can be installed and run almost everywhere. For instance, there exist VR systems accessible online for everyone equipped with a VR headset, or travelling museum exhibitions that can take place in different cities over the years. Some examples include the VR tour of Choirokoitia allowing visitors to virtually navigate through the archaeological site and acquire historical information from various important points of interest [19], and a VR diving simulation inside underwater city remains in Baiae [68]. Travelling museum exhibition are also not bound to a singular location. There are two cases of the exhibition "Etruscans" held in Italy and in the Netherlands in 2011 [61], and the reconstruction of a Dutch flute shipwreck whose exhibit was held in Iceland and Australia in 2019 [53]. In total, there are only 14 papers (25%) whose systems are independent from the location.

Conversely, the location dependency relies on two main factors: content and technology. For the content–dependent systems (33% or 18 articles), usually they are bound to an archaeological site, such as the underwater site of Baiae [76] or the ruins of Conimbriga [50]. They may be developed for a city area with a relevant archaeological past, such as Malolos City [24] or Viking fortress of Aggersborg [38]. These types of systems usually use mobile user–centric devices. The technology–dependent systems (42% or 23 articles) are those that need a particular device or platform which is not available anywhere but in a laboratory or in a dedicated room. In this case, the exact geographical location

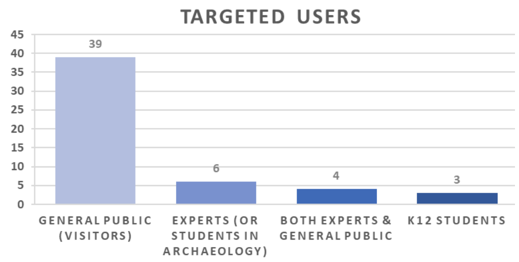

Figure 5: Number of articles grouped according to the targeted users.

is not important, but a specific equipment must be present, in order for the experience to take place. This could allow the experience to take place even in more than a single location. Several examples involve BCI [73, 76], tabletop interfaces [18, 42], projector screen [30, 41, 42, 46, 67]. The other cases are permanent exhibitions that are exclusive to museums (e.g., [23, 32, 33]).

3.4 Targeted Users

Moving the focus on the targeted users, Fig. 5 shows a prominence of systems (75% or 39 articles) developed explicitly for visitors. In this case, the user is part of the general public and does not necessarily have previous knowledge on the archaeological subject at hand (e.g., ancient Greek art [23] or rock art [8, 74]). As for the case of learning platforms, the targeted users of such systems are the students. They can be students of elementary to high schools who would take part in a historic archaeological virtual tour [2, 14, 63].

In other cases, XR systems have been conceived to help archaeologists and archaeology students learn and test the techniques that can be useful for their profession: measurement estimation for micro-excavation [31] or on an excavation site [69], or photogrammetry in a submarine site [26]. For this particular scenario, VR proves itself to be helpful. In the case of submarine sites, the excavation time is very limited and the training in a simulated environment is a considerable advantage [26]. There is also the opportunity to give archaeologists a set of virtual tools to enhance their work, such as systems to add annotations and virtually manipulate artefacts [3, 9, 41].

There exists also a growing interest in targeting both general public and experts in some systems. A possibility could be to address to the expert or the public in a different way, with a level of detail adapted to their requirements.

For example, Ridel et al. [67] tested an interaction technique called "revealing flashlight" whose different variants were chosen according to the environment (a laboratory or a public exhibition) and to the kind of audience. In ArkaeVision [12], the user' profile (according to age and culture) is used to calibrate the experience and determine which content will be shown to them. Another possible approach is to put in communication the experts and the public from two different standpoints [14]. One of the purposes in combining both types of users is to compare their reaction to a system design or an interaction technique. For instance, the user study in [39] compares the reaction of two types of users who were exposed to a real 3D printed copy of an ancient Chinese porcelain bowl and its (visual and haptic) digital representation.

3.5 Comparison between targeted users and devices

Fig. 6 shows the relations between the types of users, devices and targeted places. Some details can be noticed. The systems developed for students are based exclusively on user-centric devices. Moreover, when targeting both the general public and experts, projectors or VR headsets are preferred. Also, when experts are involved (both alone or together with the public), the mobile devices do not seem to be a choice and VR/AR headsets are the most common solution. In general, VR headsets are flexible and used for a wide typology span of users, while mobile devices targeted chiefly general public and students. Location-independent VR headsets are the most common alternative regardless of which type of user is targeted. The number of user studies for systems

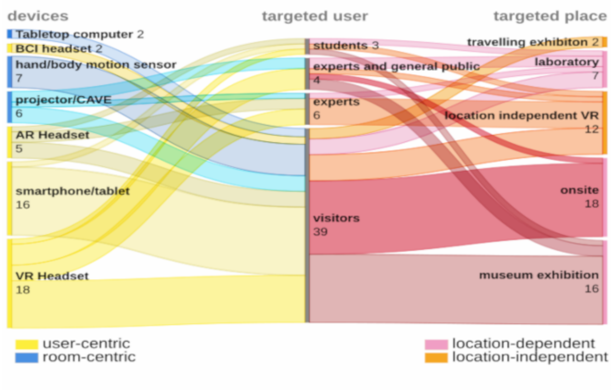

Figure 6: Overview of devices, targeted users and targeted places.

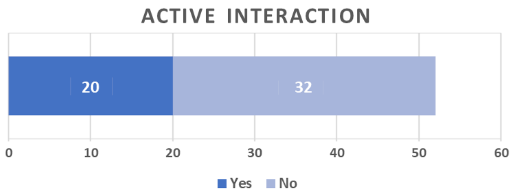

Figure 7: Number of articles with or without active interaction

that require a laboratory setting is slightly inferior. Surprisingly, there are no dedicated user studies evaluating systems dedicated for experts working onsite or visiting/working on exhibitions.

3.6 Interaction

A distinction can be made between the selected articles is that whether they design, develop, or integrate some form of active interaction. In general, the number of the systems without active interaction is slightly higher (32 vs. 20), as shown in Fig. 7.

For active interaction we intended the capability of systems which allows the user to change the state of the environment. This includes modifying the scene [31], taking notes [3], creating objects [9], manipulating and analysing objects [3, 14, 30, 33, 41, 43, 46, 63, 64, 66], breaking objects in the scene [20], and changing illumination [23, 34, 67] or timeline [28, 34]. In contrast, passive interaction involves the systems in which the user's actions are limited to exploring the scene [5, 6, 12, 13], looking around [35, 38, 62, 65, 75], watching or pausing the scene [33, 53], and getting information from points of interests [19, 76] or from non-player characters [73]. The interaction within these systems is not always the centre of the user study: for example, in the case of storytelling [47, 68] or when the visual quality is the focus [64].

3.7 Collaboration

In the majority of the articles, the collaborative aspect is not taken into consideration. The systems are designed to be used by a single person and the presence of other users (in the case) does not influence what the user sees or perceives. Only four papers have considered the multiple-user aspect.

An exhibition is created to be experienced by three users at a time and the action done by one influences what is seen by the others, following the idea that "accessing different chambers and finding artefacts was structured as a group process" [28]. Kulik et al. [42] present a system which allows the synchronous analysis from different perspectives of petroglyphs from Valcamonica in a collaborative environment, which uses a multi-stereoscopic projection technology. The multi-user aspect is important in this process because such complex tasks can be split between the users. The other advantage that multiple-user systems provide is for collaborative interpretation and analysis of archaeological data remotely, especially for the sites that cannot support a large number of simultaneous visitors such as Pleito Cave [16]. The article [33] acknowledges the issues and the possibilities of multiple users present at the same time in their system but did not really present a solution.

3.8 User Studies

In this section, we consider user studies in more details: how they were conducted, with what kind of participants and how many of them, what was measured and in what stage of the deployment process. These steps have been schematised in Fig. 8.

3.8.1 User tasks

Typically in experiments or user studies, users have to perform some task within a system to get acquainted with it and/or to evaluate it. In many experiments, however, they are not required to execute a specific task but just to use the system in a free, spontaneous way and to discover for themselves what its capabilities are. In contrast, a specific activity can be assigned to the user in some cases. The most common one is to navigate and explore the scene in a virtual tour or visit. It can be conducted in a free roaming (such as navigating inside the virtual city and interact with the agents via an avatar [73]), or to reach some points of interest to acquire some information by watching simulation scenes [53] or battles [32]. Another approach in interacting with a system is for the user to inspect, manipulate (move/rotate) artefacts or explore their properties (e.g., using haptic [39]). Specific to archaeological applications, some user studies require the user, by using a peculiar set of tools, to dig, annotate, measure distances, create, and delete points of archaeological ruins [3, 9, 69].

There is a distinction between tools designed for exhibitions with general public and tools made for experts, and how they are evaluated in the experimental task. In many cases for general public, it is simply about using the systems in a virtual tour. By contrast, in applications for experts, it is rather about some way of manipulating the scene or the objects inside of it. The work in [9] describes several available actions: creating / editing / deleting points of interest; paintings; text annotations; or accessing transition visualisation (comparing digital scenario to 360° photography). There also exist several more specifically designed archaeological tasks including: deciphering inscriptions of an Egyptian stele [67]; digging, teleporting, measuring [69]; or putting markers on the site, measuring distances between them, and then taking photos of the site [26].

The interaction with the artefacts is also a very important aspect to evaluate in the user studies. For instance, there are tasks designed to study the placement of amphorae in a virtual underwater environment [41], to find and estimate the position of an object within another [31], or to recognize the handwriting on an archaeological object [39]. A completely different task was the selection of the most suitable place for a burial site which aimed at analysing crowd behaviour [18].

3.8.2 Evaluation Measurements

During the evaluation or after having tested the system, different methods can be employed to gather information from the participants. The most common technique is subjective questionnaires, used in 38 of 52 (73%) user studies, that the user has to fill in before and/or after the experiment. In some cases, a specific questionnaire is designed appositely (e.g., cultural presence questionnaire [65] for the evaluation of the system, or vice versa, a system designed to test an evaluation method [73]). But an already purposefully constructed questionnaire is customarily preferred when evaluating usability such as SUS [3, 9, 10, 42, 52] or NASA TLX (Task Load Index) [75, 76].

In few cases (e.g., [28, 41, 62, 69]), the participants are also directly interviewed in a subjective manner to collect their impression at the moment of the user study without the rigidity of a pre-structured survey in order to "capture ground-truth data" [62].

While using the system, think-aloud feedback [3, 5, 38, 52, 53, 67, 73, 75] or direct observation [38, 62] are often used to acquire qualitative data. This type of feedback appears to be the preferred technique in user studies when there are usability experts evaluating a system or if a system is designed for an expert in the field. This is a way to tailor the system adherently to the individual user's needs. The feedback is strictly related to the user, and in this case, only a few users take part in the study.

Other less common employed techniques include: beta testing [66]; crowd behavior mining [18]; task completion performance [5, 26]; knowledge test (the user has to choose the right answer) [14, 34]; and EEG recording (for presence and engagement) [76].

3.8.3 Pre- and Post-experimental Evaluation

Except for think-aloud feedback and direct observation mentioned in Section 3.8.2, the participants usually complete the evaluation part after the experiment (post-test). Before the experiment, a pre-test can also be conducted to collect additional information. For the purpose of this study, we did not consider inquiries about the user's previous experience and demographic surveys as a pre-test, as well as any information that was not directly relevant to the data collected in the post-test. In this context, the pre-test is often present in the user studies about learning in order to test knowledge on the focused topic before and after the experience [8, 19, 24]. It is also used to evaluate the attitude and interest towards CH and the archaeology of a certain place, for example in Cyprus [19, 44]. The pre- and post-test can be conveniently conducted inside VR [26]. Usually the post-test happens immediately after the XR experience, but in some case it can be delayed in time and administered via e-mail [16].

3.8.4 Size of Participant Sample

The size of participant pool is an important factor to be considered in user studies. It is usually small (we intended less than 20 participants for small scale) in laboratory settings [4, 8, 9, 13, 31, 41], or for exhibition systems in a prototypical state which needs to be further expanded before a test at a bigger scale [38]. Small-scale user studies follow the principle that running multiple tests with fewer participants is better than a single test with a larger number of them, at least in the case of usability evaluation [57,58]. In addition, a small number of participants is common for systems evaluated with user feedback from experts [9, 53].

The medium-sample participant size (between 20 and 100 participants) is usually gathered for museum-exhibition– or laboratory–based system evaluation that needs to reach a certain threshold of observations to have significant evidence in their measurements based on statistical power analysis [21]. Besides, the installation and conduct of system evaluation in a great exhibition or a big museum hall facilitate the collection of large number of participants through visitors (more than 100) [14, 18, 32, 37, 61, 64]. For some user studies, the number of participants was not available [28, 47, 66].

3.8.5 Formative and Summative Evaluation

An application, a system or an interaction can be evaluated in different phases of its life cycle: at the beginning stage to improve its design while testing (*formative*) up until the final step to evaluate its final version (*summative*) [7, 25]. This is the most common and affirmed approach that was judged purposeful for the current review, although other supplementary categorisations have been proposed (e.g., [17, 59]).

There are 30 studies presenting a summative evaluation, 20 a formative one, and two evaluations using both approaches. All the systems built for archaeological experts [3,9,26,31,41,42,75] utilise a formative evaluation method. There is a tendency of having a summative evaluation on systems built for museum exhibitions (10 summative vs. 6 formative). In one of the user studies conducting both formative and summative evaluation, Ridel et al. [67] firstly conducted a preliminary study on the interaction technique they proposed (i.e., formative), then later run another one during exhibitions (i.e., summative). In this case, the targeted users were switched from experts to general public. Flynn in [28] states having conducted a "formative and summative qualitative analysis" but does not give many details.

3.8.6 Qualitative and Quantitative Evaluation

According to what is measured, the acquired data can be numerically quantifiable (*quantitative*) or have a more descriptive nature (*qualitative*). Most of user studies (60% or 31 articles) collect only quantitative data through Likert-scale structured answers to

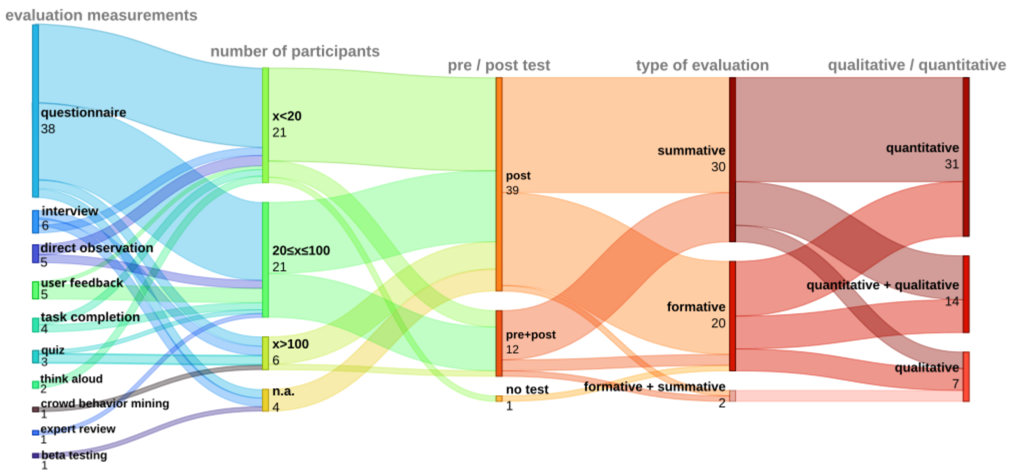

Figure 8: Overview of different steps of how user studies are conducted in XR systems for archaeology applications.

questionnaires. The other quantitative data include EEG (electroencephalogram) data [76] and data acquired while completing some task, such as the time spent for creating and interacting with archaeological artefacts and notes [9], or the time spent and the collision data while visiting the VR reconstruction of a Roman *Domus* [5]. Doležal et al. in [26] calculate the difference between the measured area and the actual area of a VR shipwreck site.

The studies collecting only qualitative data are fewer (13% or 7 articles) than the previously mentioned quantitative ones. The qualitative evaluation often provides a description of different values, such as understanding how intuitive a system is [38], having a general idea about the usability [67], or gathering data about users' immersion and learning experience [28], evaluating perception [31, 33], playability & enjoyability [73], and engagement [66]. Usually, when the evaluation has a formative intent, the data is of qualitative type (cf. [45, p. 273]). In our review, out of the seven studies that collect only qualitative data, four are sole formative evaluation [38, 47, 66, 73]. However, Hammady et al. in [33] used instead a qualitative method (semi-structured interview) for a rather summative purpose: the validity of a virtual guide in a museum over a human one.

Since at times it is complicated to interpret only numerical results, some studies collect both qualitative and quantitative data (27% or 14 articles). This is often the case for evaluating presence [9, 20, 52, 63] or usability [3, 5, 9, 10, 23, 41, 52].

3.8.7 Cognitive Aspects

There have been previous attempts at proposing a classification for the evaluation of user studies in XR (e.g., [27, 54]). In this review, we followed the model proposed in [54] to give a panorama of different valued measurements in XR within the archaeological context. It uses an inferential method to derive different categories, starting from what is measured in each user study. We adapted their categories to the articles collected for this review. The final classification is shown in the Table 1. The resulting categories, with the respective selection criteria, are as follows:

- **Usability & Learnability:** the canonical definition of usability includes effectiveness, efficiency, satisfaction, and ease of learning how to use a system. The focus is on the system performance;
- **Emotion:** this factor is used to capture the mental state (feelings, behaviours) of the user;
- **Learning & Education:** this measure describes the capability of system in supporting the acquiring and retaining of information in the most effective (and less boring) way;
- **Perception:** it is to measure the user's understanding and perception of the environment through the interpretation of sensory information (e.g., visual, auditory, or haptic);
- **Presence:** this aspect tries to capture the feeling of no barriers between oneself and the virtual environment;
- **Cognitive Load:** it is about the quantity of mental effort that has to be sustained by the user; and
- **Engagement & Attention:** they describe the user's ability to selectively concentrate and focus on one single aspect.

The aspect of accessibility was not considered in most of the papers, except in [61, 63].

4 Discussion

In this section, on the basis of the different categorisations provided in the previous section, we aim to answer one by one the research questions raised in the Introduction.

RQ1: What are the related fields of study or other subjects to be aware of, in the design and development of an XR-based system for archaeology?

As anticipated in Section 2, there is a close connection between Archaeology and Cultural Heritage (CH), so that an application built within the context of CH may contain the same elements or have the same structures and interaction techniques as one built for archaeology and vice versa. Indeed, they overlap in many aspects, even for excavation process support (cf. [16, 31]), and not just in visualisation and outreach of CH.

Another related field mentioned in many of the studies is education (e.g., experiential learning [12, 19], blended learning [64], teaching [14]). There are two main reasons why this is an important field linked to XR and archaeology: it brings forth the interest of students on archaeology in educational outreach; it also allows archaeological students to learn and practise field activities [69], even for locations that are hard to reach [9] or accessible only for a limited period of time [26].

A popular combination that has always been considered, especially if the XR system is addressed to a non-expert public, is storytelling and serious games. These domains are crucial in determining the level of immersion and presence, as well as providing some engagement and edutainment to the user [19, 33, 51, 62, 68, 76]. This is especially the case for the presentation of archaeological artefacts or the visit of a site. It is due to the fact that it is more engaging by making an object or place more vivid and memorable through reviving its story or game play. With the same purpose, but by the opposite method, the temporal distance can also be emphasised. Indeed, the concept of the flow of time is very central in archaeology [36]. To highlight the change of time, there exist metaphors deriving from philosophical concepts [37] that can be referred to, or body sensations [28] that focus on breath. In these last cases, the human-centric part of human-computer interaction gains more importance.

Although not so much present in the articles (e.g., [65]), studies in the psychology and neuroscience field are also to be advocated as they lay down the foundation in helping us to understand how people can relate to archaeology using XR technology. Another study field that takes the same approach but puts more focus

Cognitive Aspects						
Usability & Learnability	**Emotion**	**Learning & Education**	**Perception**	**Presence**	**Cognitive Load**	**Engagement & Attention**
learnability [4] [13] [62]	acceptance [20] [43]	learning outcomes [63]	realism [64]	immersion [9] [68]	perceived task overload [63]	engagement [16] [64] [66] [76]
ease of use [23] [68]	affective impact [23]	educational usefulness [2]	perception [30]	object presence [16] [28] [41] [43] [76]	task load [75]	attention [4] [6]
efficiency [62]	attitude towards the experience [44] [46]	edutainment [68]	similarity between the real artifact and its digital representation [39]	presence [9] [19] [34] [41] [46] [50] [52] [76]		
intuitivity [6] [38]	behaviour [18]	learning performance [19]	spatial context [34]	spacial presence [63]		
navigation performance [6]	enjoyability [73]	learning rate [14] [24]	user perception [33]	cultural presence [65]		
user-friendliness [24]	enjoyment [4] [13] [32] [50] [63]	learning [64]	spatial perception [31]	sense of presence [20] [44]		
system efficiency [4]	experience feedback [53]	learning experience [28]				
system performance [10] [35]	technology acceptance model [32]	knowledge [50]				
usability [2] [3] [4] [5] [8] [9] [10] [12] [23] [34] [35] [41] [43] [51] [52] [61] [67] [74]	user acceptance [24]					
usefulness [32] [42]	appreciation [37]					
efficacy [13]	playability [73]					
device efficacy [4]	stimulation [62]					
user experience [40] [44] [47] [61] [75]	attraction [37]					
	motivation [69]					
	emotional response [76]					

Table 1: Classification of all the selected articles into different categories of cognitive aspects.

on interaction is brain-computer interfaces using objective input from the human brain in real time. These interfaces can provide a layer of connection closer to what a person perceives but it has not acquired so much notoriety yet (only [76] in our review). Last but not least, one must not forget that the rise of XR technology in archaeology is all thanks to the advances of 3D reconstruction, data visualisation, real-time rendering, amongst others in the field of computer graphics. For instance, to be able to present archaeological artefacts not only in the most pleasant and attractive way, but also to make the interaction with them easy and user-friendly, data visualisation techniques (in particular 3D data visualisation) can prove advantageous as shown in [3, 31].

RQ2: What is the best way or the best recommendable methodology to evaluate an XR archaeology system? Could a standardised model of analysis be proposed for evaluation of XR systems for archaeology?

At the end of our review, the answer to this question proved to be complex due to many cognitive aspects measured in a user study (Table 1). The two most considered aspects are the usability & learnability as well as emotion. This reflects the ubiquitous distinction between archaeology as a scientific procedure to experts or just as preservation of the past to be presented to general public. This difference in purpose distinctively decides the focus of the design and development of XR platforms for archaeology. It can be either a system that requires a certain amount of considerations on the interaction for experts to exercise their expertise, or just to have an impact on the user experience. In the latter situation, it is important to measure the emotional aspect and the presence and to evaluate how the system reacts to the user [45]. Besides, the cognitive aspect whose significance does not vary so much depending on the context is that of perception as it is providing understandable information about the environment. However, not so many papers focus on this component as well as on the cognitive load, an important factor when performing technical tasks in XR. The cognitive load can also be split between multiple users in collaborative environments [42].

Evaluating the learning aspect is not fundamental for archaeological systems in XR but it is ancillary to bringing forward archaeological topics to school students or non-specialists in an education outreach. The approach of making this in the most pleasing way is covered by the engagement and attention aspect. To do so, serious games and storytelling are often the recipe of success to be embedded in the design of XR systems.

Regarding the size of participant pool, a prototype or a system that targets experts will have a smaller number of subjects. Its type of evaluation will most likely be formative to be able to adapt to personalised requirements and flexibly improve at each iteration. On a side note, hierarchical task analysis (HTA) approach [1] is recommended to help unpack the complex cognitively loaded tasks of the archaeological experts in immersive environments. We noticed a remarkable difference in the quality of the analysis of the data. In some articles, in-depth analyses serve to gain strong evidence to support the hypotheses [5, 6, 44, 46, 50, 51]. Other papers just report the general impression, showing the percentage or the mean and standard deviation of answers to the questionnaires.

To conclude, there is not a generalised standard method to evaluate XR applications for archaeology, but some guidelines can still be traced. No matter the purpose of the application, a usability test is almost always inescapable and it appears to be the most common and accepted practice. Evaluating emotional aspects is more important in the case when the impression on the user is the focal point. It has been referred to as the "wow factor" of virtual archaeology [29] in exhibition or virtual tours.

RQ3: Are there interaction techniques or platforms that can target different types of users?

From the selected articles, we can find two main types of systems conceived for general public and experts. The experts are archaeologists, students of archaeology, and curators of museums. Regarding XR devices, there is not a substantial distinction in choice between those designed for different types of users.

The distinction lies between room-centric and user-centric devices that align themselves around the indoor and outdoor usage axis (cf. Section 3.2). VR devices are a case apart because they are location-independent for the most part. So, they have the advantage of portability and repeatability in different places, which is not often possible for room-centric devices. This is not so restricting in case of a system that has to be used in a fixed and not so extended location. Mobility is also a considerable advantage for AR, although there are also limits in these technologies such as overheating during exhibition, visibility outdoor for AR headsets, and low computational performance. Studies that involve a room-centric device seem to require a greater effort in terms of setup, which again can be fine for a presentation to a large public. Besides, it is not reducible to smaller studies all the times, and in fact, studies that involve experts with room-centric devices are usually part of a larger project [41, 42, 62].

One of the major differences observed in evaluating experts and general public is the number of participants. It is tied to the availability of a greater number of subjects for the second group. It translates, therefore, in a larger number of studies closer to the scope of CH exhibitions and visits than to the scientific and technical part of archaeology (cf. [31]). In user studies with archaeologists, it is preferable to use qualitative and formative evaluation. This had to be expected as a standard technique of interaction in XR is yet to be found and is still a relevant topic of research. But the interaction methods that were tested seem to perform well when addressing multiple target users and general public [12, 34, 67], with a distinction between the quantity of information to visualise according to the targets. Based on this, we can conclude that the experimental techniques are not yet established, which can be used both for experts and general public even though there are some remarkable works such as [23, 42, 67].

On the basis of the analysed papers, we evidenced some general tendencies that can be helpful to design an XR system for archaeology. First, it is important to consider how much stress has to be put on graphical aspect and on realism because they would significantly impact the perception and experience of the users. Second, when a system is developed to target archaeological experts, mobile devices usually do not take the lead in the list of options to go to. Otherwise, there is no specific device that stands out for a specific target of users. Considering usability and emotional aspects, the latter has more emphasis in a system intended for visitors of an exhibition than for experts. Another question that a designer of an XR system has to ponder is whether to embed storytelling or not to improve the flow of the experience of the user. In addition, the type of evaluation technique varies strongly according to the objectives of the user study required and the depth of statistical analysis as mentioned above.

5 Conclusion

In this paper we conducted a literature review about user studies for archaeology in eXtended Reality (XR). At the end of the process, 52 articles were selected for further analysis. On these grounds, we categorised them according to different axes: devices, location, targeted users, interaction, and collaboration. The user studies present in each article were considered with a further level of detail with the purpose of deducing some useful guidelines for future work on XR systems for archaeology. We also identified what are the related fields to be aware of, while working at this junction of two main domains. The heterogeneity of the evaluated measurements in the studies proved to be a limit to the generalisation of the procedure. However, no substantial difference was found between the interaction methods of systems that have a different user target (namely, general public and archaeological experts). In the review of the articles, a lack of many user studies focusing on multiple-user interfaces and collaborative aspects was also evidenced. The review also denoted a growing interest in marine and underwater archaeology, which can lay down the groundwork for the potential benefits of XR. While XR has an important presence in museums and exhibition halls, fewer studies have been conducted for XR systems helping archaeologist in their daily work. Therefore, an increasing amount of future research is to be expected.

ACKNOWLEDGEMENTS

This work was supported by French government funding managed by the National Research Agency under the Investments for the Future program (PIA) with the grant ANR-21-ESRE-0030 (CONTINUUM project), and for the first author, by PhD grants from Région Ile-de-France and Noovae company.

REFERENCES

[1] J. Annett. Hierarchical task analysis. In E. Hollnagel, editor, *Handbook of Cognitive Task Design*, pages 17–36. Lawrence Erlbaum Associates, 2003.

[2] P. Arias-Espinoza, A. Medina-Carrión, V. Robles-Bykbaev, Y. Robles-Bykbaev, F. Pesántez-Avilés, J. Ortega, D. Matute, and V. Roldán-Monsalve. e-pumapunku: An interactive app to teach children the cañari and inca indigenous cultures during guided museum visits. In *2018 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)*, pages 1–5, 2018.

[3] J. Barbier, P. Kenny, J. Young, J.-M. Normand, M. T. Keane, M. O'Sullivan, and A. Ventresque. Maap annotate: When archaeology meets augmented reality for annotation of megalithic art. In *2017 23rd International Conference on Virtual System Multimedia (VSMM)*, pages 1–8, 2017.

[4] L. Barbieri, F. Bruno, and M. Muzzupappa. Virtual museum system evaluation through user studies. *Journal of Cultural Heritage*, 26:101–108, 2017.

[5] V. Barneche-Naya and L. A. Hernández-Ibáñez. Ux aspects of kinect-based movement schemes inside virtual environments for museum installations. In P. Zaphiris and A. Ioannou, editors, *Learning and Collaboration Technologies. Ubiquitous and Virtual Environments for Learning and Collaboration*, pages 133–150, Cham, 2019. Springer International Publishing.

[6] V. Barneche-Naya and L. Hernandez Ibañez. A comparative study on user gestural inputs for navigation in nui-based 3d virtual environments. *Universal Access in the Information Society*, 20:1–17, 08 2021.

[7] C. Biron and M. Karanika-Murray. Process evaluation for organizational stress and well-being interventions: Implications for theory, method, and practice. *International Journal of Stress Management*, 21, 01 2013.

[8] S. Blanco-Pons, B. Carrión-Ruiz, J. Luis Lerma, and V. Villaverde. Design and implementation of an augmented reality application for rock art visualization in cova dels cavalls (spain). *Journal of Cultural Heritage*, 39:177–185, 2019.

[9] E. Borba, A. G. Corrêa, R. Lopes, and M. Zuffo. Usability in virtual reality: evaluating user experience with interactive archaeometry tools in digital simulations. *Multimedia Tools and Applications*, 79, 02 2020.

[10] F. Bousbahi and B. Boreggah. Mobile augmented reality adaptation through smartphone device based hybrid tracking to support cultural heritage experience. In *Proceedings of the 2nd International Conference on Smart Digital Environment*, ICSDE'18, page 48–55, New York, NY, USA, 2018. Association for Computing Machinery.

[11] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., USA, 2004.

[12] G. Bozzelli, A. Raia, S. Ricciardi, M. De Nino, N. Barile, M. Perrella, M. Tramontano, A. Pagano, and A. Palombini. An integrated vr/ar framework for user-centric interactive experience of cultural heritage: The arkaevision project. *Digital Applications in Archaeology and Cultural Heritage*, 15:e00124, 2019.

[13] F. Bruno, L. Barbieri, A. Lagudi, M. Cozza, A. Cozza, R. Peluso, and M. Muzzupappa. Virtual dives into the underwater archaeological treasures of south italy. *Virtual Reality*, 22:1–12, 06 2018.

[14] A. Bustillo, M. Alaguero, I. Miguel, J. M. Saiz, and L. S. Iglesias. A flexible platform for the creation of 3d semi-immersive environments to teach cultural heritage. *Digital Applications in Archaeology and Cultural Heritage*, 2(4):248–259, 2015.

[15] J. Carman. *Archaeology and heritage: An introduction*. Continuum, London; New York, 2002.

[16] B. Cassidy, D. Robinson, G. Sim, and D. Gandy. A virtual reality platform for analyzing remote archaeological sites. pages 1–5, 07 2018.

[17] H.-T. Chen. Epilogue: Synthesizing formative/summative evaluation issues and beyond. *Evaluation Practice*, 17(2):163–167, 1996.

[18] E. Ch'ng, V. Gaffney, P. Garwood, H. Chapman, R. Bates, and W. Neubauer. Merging the real with the virtual: Crowd behaviour mining with virtual environments. In *2016 22nd International Conference on Virtual System Multimedia (VSMM)*, pages 1–9, 2016.

[19] M. Christofi, C. Kyrlitsias, D. Michael-Grigoriou, Z. Anastasiadou, M. Michaelidou, I. Papamichael, and K. Pieri. *A Tour in the Archaeological Site of Choirokoitia Using Virtual Reality: A Learning Performance and Interest Generation Assessment*, pages 208–217. 01 2018.

[20] E. Ch'ng, Y. Li, S. Cai, and F.-T. Leow. The effects of vr environments on the acceptance, experience, and expectations of cultural heritage learning. *J. Comput. Cult. Herit.*, 13(1), feb 2020.

[21] J. Cohen. Statistical power analysis. *Current Directions in Psychological Science*, 1(3):98–101, 1992.

[22] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The cave: Audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–72, 1992.

[23] A. Damala, E. Hornecker, M. J. van der Vaart, D. Van Dijk, and I. Ruthven. The Loupe: Tangible Augmented Reality for Learning to Look at Ancient Greek Art. *Mediterranean Archaeology and Archaeometry*, 16(5):73 – 85, 2016.

[24] D. R. dela Cruz, J. S. Sevilla, J. W. D. San Gabriel, A. J. P. Dela Cruz, and E. J. S. Caselis. Design and development of augmented reality (ar) mobile application for malolos' kameztizuhan (malolos heritage town, philippines). In *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pages 1–9, 2018.

[25] W. Dick, L. Carey, and J. O. Carey. *The Systematic Design of Instruction*. Pearson, 1996.

[26] M. Doležal, M. Vlachos, M. Secci, S. Demesticha, D. Skarlatos, and F. Liarokapis. Understanding underwater photogrammetry for maritime archaeology through immersive virtual reality. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W10:85–91, 2019.

[27] A. Duenser, R. Grasset, and M. Billinghurst. A survey of evaluation techniques used in augmented reality studies. *ACM SIGGRAPH ASIA 2008 ...*, 01 2008.

[28] B. Flynn. v-embodiment for cultural heritage. In *2013 Digital Heritage International Congress (DigitalHeritage)*, volume 1, pages 347–354, 2013.

[29] M. Forte. Introduction to cyber-archaeology. *Cyber-Archaeology, BAR International Series*, 2177:9–13, 01 2010.

[30] P. Franco, C. Camporesi, F. Galeazzi, and M. Kallmann. 3d printing and immersive visualization for improved perception of ancient artifacts. *Presence Teleoperators Virtual Environments*, 24, 06 2015.

[31] R. Gaugne, Q. Petit, M. Otsuki, V. Gouranton, and T. Nicolas. Evaluation of a Mixed Reality based Method for Archaeological Excavation Support. In Y. Kakehi and A. Hiyama, editors, *ICAT-EGVE 2019 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2019.

[32] R. Hammady, M. Ma, and C. Strathearn. Ambient information visualisation and visitors' technology acceptance of mixed reality in museums. *J. Comput. Cult. Herit.*, 13(2), jun 2020.

[33] R. Hammady, M. Ma, C. Strathearn, and M. Mohamad. Design and development of a spatial mixed reality touring guide to the egyptian museum. *Multimedia Tools and Applications*, 79, 02 2020.

[34] P.-H. Han, Y.-S. Chen, I.-S. Liu, Y.-P. Jang, L. Tsai, A. Chang, and Y.-P. Hung. A compelling virtual tour of the dunhuang cave with an immersive head-mounted display. *IEEE Computer Graphics and Applications*, 40(1):40–55, 2020.

[35] M. Hincapié Montoya, C. Diaz, M. I. Zapata Cárdenas, and C. Mesias. Methodological framework for the design and development of applications for reactivation of cultural heritage: Case study cisneros marketplace at medellin, colombia. *Journal on Computing and Cultural Heritage*, 9:1–24, 01 2016.

[36] S. Holdaway and L. Wandsnider. Time in archaeology: An introduction. *Time in Archaeology: Time Perspectivism Revisited*, 01 2008.

[37] C.-K. Hsieh, W.-C. Liao, M.-C. Yu, and Y.-P. Hung. Interacting with the past: Creating a time perception journey experience using kinect-based breath detection and deterioration and recovery simulation technologies. *J. Comput. Cult. Herit.*, 7(1), feb 2014.

[38] C. Jakobsen, J. Larsen, M. Nørlem, and M. Kraus. Reviving aggersborg: Conveying lost heritage sites through indirect augmented

reality. pages 1–4, 03 2017.

[39] M. H. Jamil, P. S. Annor, J. Sharfman, R. Parthesius, I. Garachon, and M. Eid. The role of haptics in digital archaeology and heritage recording processes. In *2018 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*, pages 1–6, 2018.

[40] A. N. Jayawardena and I. Perera. A framework for mixed reality application development: A case study on yapahuwa archaeological site. In *2016 Sixteenth International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 186–192, 2016.

[41] I. Katsouri, A. Tzanavari, K. Herakleous, and C. Poullis. Visualizing and assessing hypotheses for marine archaeology in a vr cave environment. *J. Comput. Cult. Herit.*, 8(2), mar 2015.

[42] A. Kulik, A. Kunert, S. Beck, C.-F. Matthes, A. Schollmeyer, A. Kreskowski, B. Fröhlich, S. Cobb, and M. D'Cruz. Virtual Valcamonica: Collaborative Exploration of Prehistoric Petroglyphs and Their Surrounding Environment in Multi-User Virtual Reality. *Presence: Teleoperators and Virtual Environments*, 26(3):297–321, 08 2017.

[43] P. Kyriakou and S. Hermon. Can i touch this? using natural interaction in a museum augmented reality system. *Digital Applications in Archaeology and Cultural Heritage*, 12:e00088, 12 2018.

[44] C. Kyrlitsias, M. Christofi, D. Michael-Grigoriou, D. Banakou, and A. Ioannou. A virtual tour of a hardly accessible archaeological site: The effect of immersive virtual reality on user experience, learning and attitude change. *Frontiers in Computer Science*, 2, 2020.

[45] J. Lazar, J. H. Feng, and H. Hochheiser. *Research Methods in Human-Computer Interaction*. Morgan Kaufmann, Cambridge, MA, 2 edition, 2017.

[46] A. Leopardi, S. Ceccacci, M. Mengoni, S. Naspetti, D. Gambelli, E. Ozturk, and R. Zanoli. X-reality technologies for museums: a comparative evaluation based on presence and visitors experience through user studies. *Journal of Cultural Heritage*, 47:188–198, 2021.

[47] G. Liestøl, M. Bendon, and E. Hadjidaki-Marder. Augmented reality storytelling submerged. dry diving to a world war ii wreck at ancient phalasarna, crete. *Heritage*, 4(4):4647–4664, 2021.

[48] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 4(2):1–13, jan 2007.

[49] A. Marto, A. Gonçalves, M. Melo, and M. Bessa. A survey of multisensory vr and ar applications for cultural heritage. *Computers Graphics*, 102:426–440, 2022.

[50] A. Marto, M. Melo, A. Gonçalves, and M. Bessa. Multisensory augmented reality in cultural heritage: Impact of different stimuli on presence, enjoyment, knowledge and value of the experience. *IEEE Access*, 8:193744–193756, 2020.

[51] A. Marto, M. Melo, A. Gonçalves, and M. Bessa. Development and evaluation of an outdoor multisensory ar system for cultural heritage. *IEEE Access*, 9:16419–16434, 2021.

[52] A. Martí-Testón, A. Muñoz, J. E. Solanes, L. Gracia, and J. Tornero. A methodology to produce augmented-reality guided tours in museums for mixed-reality headsets. *Electronics*, 10(23), 2021.

[53] J. McCarthy and K. Martin. Virtual reality for maritime archaeology in 2.5d: A virtual dive on a flute wreck of 1659 in iceland. In *2019 23rd International Conference in Information Visualization – Part II*, pages 104–109, 2019.

[54] L. Merino, M. Schwarzl, M. Kraus, M. Sedlmair, D. Schmalstieg, and D. Weiskopf. Evaluating mixed and augmented reality: A systematic literature review (2009-2019), 2020.

[55] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77:1321–1329, 1994.

[56] M. Nawaf, P. Drap, M. Ben-Ellefi, E. Nocerino, B. Chemisky, T. Chassaing, A. Colpani, V. Noumossie, K. Hyttinen, J. Wood, T. Gambin, and J. C. Sourisseau. Using virtual or augmented reality for the time-based study of complex underwater archaeological excavations. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, VIII-M-1-2021:117–124, 2021.

[57] J. Nielsen. Why you only need to test with 5 users. 01 2000.

[58] J. Nielsen and T. K. L. A mathematical model of the finding of usability problems. In *Proceedings of ACM INTERCHI'93 Conference*, pages 24–29, 1993.

[59] D. Nutbeam and A. Bauman. *Evaluation in a Nutshell: A Practical Guide to the Evaluation of Health Promotion Programs*. In a Nutshell. McGraw-Hill, 2006.

[60] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, 2021.

[61] S. Pescarin, M. Wallergård, W. Hupperetz, A. Pagano, and C. Ray. Archeovirtual 2011: An evaluation approach to virtual museums. pages 25–32, 09 2012.

[62] E. Pietroni and A. Adami. Interacting with virtual reconstructions in museums: The etruscanning project. *J. Comput. Cult. Herit.*, 7(2), jun 2014.

[63] C. Pollalis, E. J. Minor, L. Westendorf, W. Fahnbulleh, I. Virgilio, A. L. Kun, and O. Shaer. Evaluating learning with tangible and virtual representations of archaeological artifacts. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '18, page 626–637, New York, NY, USA, 2018. Association for Computing Machinery.

[64] A. Puig, I. Rodríguez, J. L. Arcos, J. A. Rodríguez-Aguilar, S. Cebrián, A. Bogdanovych, N. Morera, A. Palomo, and R. Piqué. Lessons learned from supplementing archaeological museum exhibitions with virtual reality. *Virtual Reality*, 24:343–358, 2019.

[65] L. Pujol. Cultural presence in virtual archaeology: An exploratory analysis of factors. *Presence: Teleoperators and Virtual Environments*, 26:247–263, 05 2018.

[66] H. Richards-Rissetto, J. Robertsson, J. von Schwerin, G. Agugiaro, F. Remondino, and G. Girardi. *Geospatial Virtual Heritage: A Gesture-Based 3D GIS to Engage the Public with Ancient Maya Archaeology*, pages 118–130. 02 2014.

[67] B. Ridel, P. Reuter, J. Laviole, N. Mellado, N. Couture, and X. Granier. The Revealing Flashlight: Interactive spatial augmented reality for detail exploration of cultural heritage artifacts. *Journal on Computing and Cultural Heritage*, 7(2):6:1–6:18, May 2014.

[68] S. Rizvic, D. Boskovic, F. Bruno, B. D. Petriaggi, S. Sljivo, and M. Cozza. Actors in VR storytelling. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. IEEE, sep 2019.

[69] L. Shackelford, W. D. Huang, A. Craig, C. Merrill, and D. Chen. Relationships between motivational support and game features in a game-based virtual reality learning environment for teaching introductory archaeology. *Educational Media International*, 56(3):183–200, 2019.

[70] C. Teeng, C. K. Lim, A. Rafi, K. Tan, and M. Mokhtar. Comprehensive systematic review on virtual reality for cultural heritage practices: coherent taxonomy and motivations. *Multimedia Systems*, 28, 06 2022.

[71] G. Vargas, R. Fabregat, A. Ramos, and T. Jové. Survey: Using augmented reality to improve learning motivation in cultural heritage studies. *Applied Sciences*, 10:897, 01 2020.

[72] M. Vecco. A definition of cultural heritage: From the tangible to the intangible. *Journal of Cultural Heritage*, 11(3):321–324, 2010.

[73] A. Vourvopoulos, F. Liarokapis, and P. Petridis. Brain-controlled serious games for cultural heritage. In *2012 18th International Conference on Virtual Systems and Multimedia*, pages 291–298, 2012.

[74] Q. Zhang, X. Zhu, H. Yu, and Y. Jiang. Enhancing rock painting tour experience with outdoor augmented reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 118–121, 2019.

[75] J. Čejka, M. Mangeruga, F. Bruno, D. Skarlatos, and F. Liarokapis. Evaluating the potential of augmented reality interfaces for exploring underwater historical sites. *IEEE Access*, 9:45017–45031, 2021.

[76] F. Škola, S. Rizvić, M. Cozza, L. Barbieri, F. Bruno, D. Skarlatos, and F. Liarokapis. Virtual reality with 360-video storytelling in cultural heritage: Study of presence, engagement, and immersion. *Sensors*, 20:5851, 10 2020.