CHAPTER 9

Evaluation in Virtual Heritage

Panayiotis Koutsabasis
University of the Aegean, Department of Product and Systems Design Engineering, Interactive Systems Design Laboratory, Greece
kgp@aegean.gr

Abstract

Evaluation in virtual heritage is concerned with learning about and assessing the extent to which an interactive system offers a satisfactory user experience (UX) and meets user goals and expectations. Evaluation in virtual heritage is an empirical process of research, which reaches for conclusions about the quality of a system by observing, measuring (aspects of), and interpreting the UX. It is inherently a complex activity that requires careful planning and selection of methods. It does not rely on underlying technology; however, adaptations of process and methods must be made to allow for results and feedback in context. Therefore, it must be designed so that it is useful, reliable, valid, and productive. Evaluation methods and processes are of interest to both cultural heritage (CH) professionals and technology designers, who aim to provide systems that address the widest range of potential users. This chapter discusses basic concepts, processes, and empirical evaluation methods in virtual heritage, with examples.

How to cite this book chapter:
Koutsabasis, P. 2021. Evaluation in Virtual Heritage. In: Champion, E. M. (ed.) *Virtual Heritage: A Guide*. Pp. 115–127. London: Ubiquity Press. DOI: https://doi.org/10.5334/bck.k.

Introduction

Virtual Heritage Technology

An increasing number of interactive systems aim to enhance the user experience (UX) of visitors at CH places and sites including museums, exhibitions, archaeological places, historic cities, or settlements. These systems have various goals, such as information presentation, learning, visitor engagement (when users assume roles and pursue learning goals), digital representation, preservation, or reconstruction of monuments, sites, or the like, combined with gamification elements or developed into a game. They are developed with contemporary and emerging technologies such as interactive walls, tables and surfaces, virtual/augmented/mixed reality (VR/AR/MR) systems, 3D virtual worlds, mobile location-based services and games, and so on. In a recent review article, Nikolakopoulou and Koutsabasis (2020) reported on the most common interactive technologies and interaction styles of virtual heritage: 3D game engines, mobile technologies, kinaesthetic interaction, physical computing, VR, and AR.

Importance of Evaluation of Virtual Heritage

From a design and a user-centred perspective, evaluation in virtual heritage is concerned with learning about and assessing the extent to which an interactive system offers a satisfactory UX meeting user goals and expectations, adapting ideas and methods from fields such as human-computer interaction (HCI) and interaction design.

The evaluation of virtual heritage leads to findings and recommendations about user acceptance, which interest both CH professionals and technology designers. Evaluation in virtual heritage is an empirical process of research: it reaches conclusions about the quality of a system by observing, measuring, and interpreting (aspects of) the UX. An indicative list of empirical evaluation methods includes observation, interview, user testing, field testing, field studies, questionnaires, surveys, and diary studies. Evaluation concepts, methods, processes, and tools are generic: they are independent of interactive technology; however, some adaptations or specifications may have to be made. Furthermore, to achieve useful recommendations for (re)design, the affordances of interactive technologies must be considered.

The purpose of this chapter is to highlight important concepts and present practical guidelines for the evaluation of virtual heritage. Here, the term *virtual heritage* denotes any type of computer-based interactive systems that promote CH. A visual model of the main concepts discussed is depicted in Figure 21.

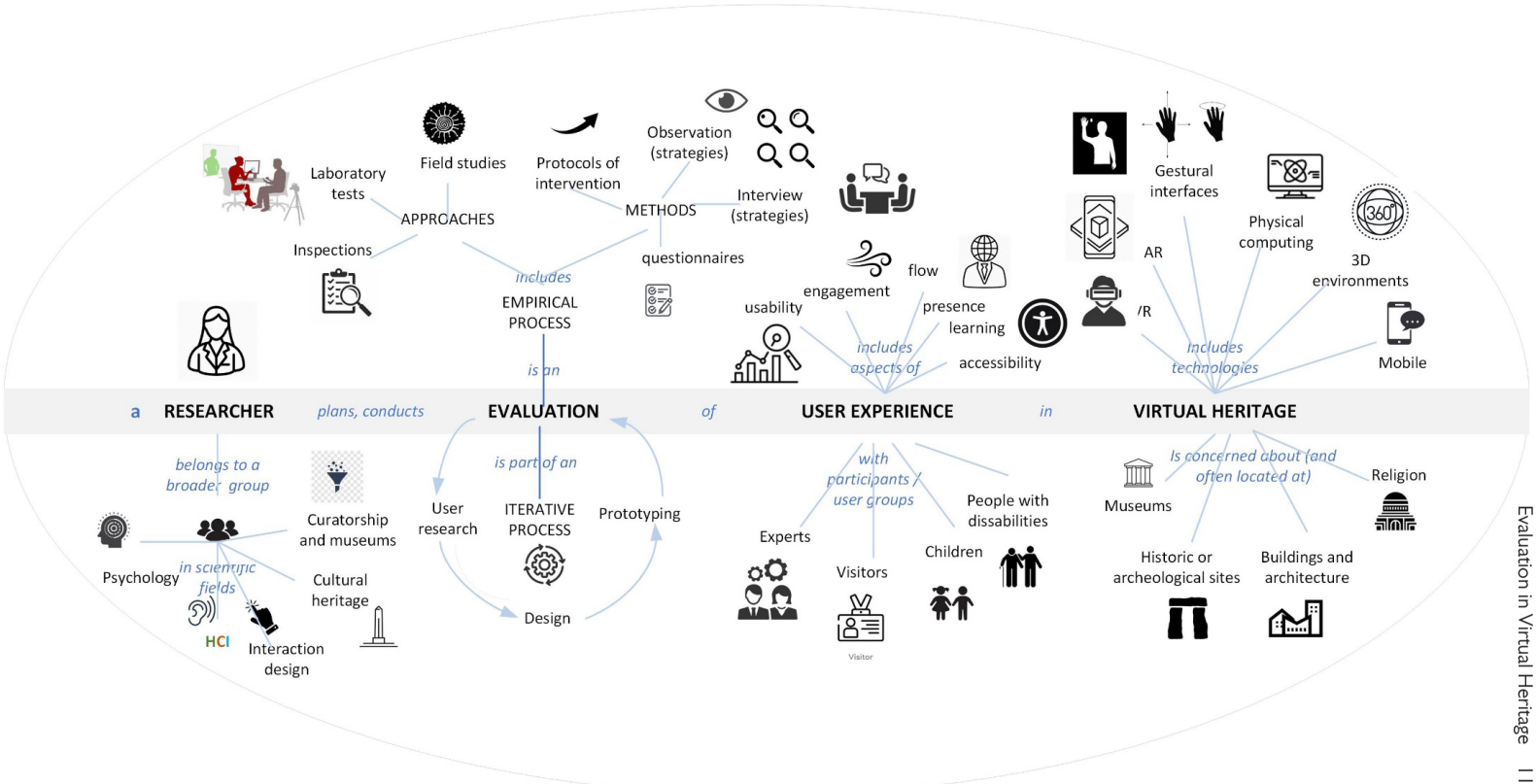

Figure 21: Visual model of the concepts about evaluation in virtual heritage[1].

Evaluation Process: Empirical and Iterative

Evaluation in virtual heritage is an empirical process of research; it reaches conclusions about the quality of a system by observing, measuring (aspects of), and interpreting the UX, which consists of various elements that arise during interactions such as usability, accessibility, engagement, sensitization, findability, learning effect, and so on. Empirical evidence, that is, the record of one's experiences, can then be analyzed quantitatively or qualitatively. Users are the most important factor in an evaluation process: they do not have to be many, but good representatives of real users representing a wide range of the intended audience or experts (in the case of system inspections).

Evaluation in virtual heritage essentially happens iteratively during system development, to feed through the design, and operation, to help reflect on experience and impact. At design time, evaluation is exploratory, such as seeking required features. At the prototyping stage it is formative, generating issues for redesign. At the piloting or operation stage it can be summative, reaching to conclusions about the outcome. Each iteration differs in terms of the technical maturity of the system, the process and intended outcomes, the intended participants' (evaluators') profiles, and data collected and processed.

Dimensions of Evaluation

The evaluation of virtual heritage usually emphasises one or more dimensions that arise during interaction with technology. In the review of 83 evaluation studies, Nikolakopoulou and Koutsabasis (2020) identify the main dimensions of evaluations in:

- User experience (19.9% of studies).
- Usability (19.1%).
- Perceived usability (8.5%).
- Engagement (7.1%).
- Learning (6.4%).

In this section, we briefly present some of these dimensions, for which there is a large corpus of background work in many fields related to humanities and human sciences, culture, design, and technology. These dimensions are often not seen in isolation in evaluation of virtual heritage but in combination with each other and others.

User Experience

UX is a general, fluid, changing, personal, and subjective concept, with many definitions. According to ISO 9241-210 (Ergonomics of Human-System

Interaction), UX is 'a person's perceptions and responses that result from the use or anticipated use of a product, system or service'. According to Norman and Nielsen Group, UX 'encompasses all aspects of the end-user's interaction with the company, its services, and its products'. Many models of UX have been proposed, such as those of Forlizzi and Battarbee (2004) and Karapanos et al. (2009). It is now widely agreed that UX incorporates pragmatic and hedonic product qualities: the former refers to the utility and usability aspects, while the latter consider the aspects of pleasure and emphasize stimulation, fun, identification generated by the use of a product or system.

Usability

According to ISO 9241, usability is 'the degree to which a product can be used by specified users, to achieve specified goals, with effectiveness, efficiency, and personal satisfaction, in a specified context of use'. Usability professionals have developed a whole corpus of evaluation methods and techniques (e.g., Nielsen 1994; Cairns & Cox 2008), and they all share the following (Lewis 2014):

a) a careful plan of study, including initial instructions and debriefing protocols;
b) participants who are members of the population of interest; and
c) appropriate tasks and environments.

By definition, usability evaluation refers to performance and preference; the former is measured by metrics such as task success, task time, and errors, while the latter is obtained directly or indirectly via interviews, observation, questionnaires, and so on. UX evaluation builds on usability methods or includes them as an essential part of the process (Albert & Tullis 2013).

Engagement and Flow

There are many definitions of engagement in HCI. According to O'Brien and Toms (2008), 'engagement is a quality of user experiences with technology that is characterized by challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, interest, and affect'. According to Doherty and Doherty (2018), engagement is often related to flow theory, which posits the existence of a state of optimal and enjoyable experience characterized by challenge, immersion, control, freedom, clarity, immediate feedback, temporal insensitivity, and changes in one's sense of identity (Cowley et al. 2008). In virtual heritage evaluation, flow has been explored in CH projects of mobile museum narratives (Roussou & Katifori 2018). Engagement has been measured (Doherty & Doherty 2018) with various techniques and tools, including questionnaire, behavior logging, observation, task outcomes, interviews, eye tracking, and the like.

Presence

The concept of presence originates in telepresence research. Minsky (1980) broadly described it as the feeling of being present (by appearance) or having an effect at a place other than their true location via technology mediation. Since then the concept of presence has been extensively discussed in computer-mediated communication (CMC) and VR research, where it has been defined (Lombard & Ditton 1997) as 'the feeling of being there' or 'an illusion that a mediated experience is not mediated.' In virtual heritage, the concept of cultural presence has been proposed (Pujol & Champion 2012), as the feeling of 'being – not only physically but also socially, culturally – there and then'.

Learning

There are many definitions of learning in general as well as in relation to technology as a mediator or to CH. Relevant to virtual heritage is non-formal and informal learning that is pervasive, self-regulatory, active, and participatory,[2] in contrast to formal learning processes involving tutors, learning goals, and a classroom setting. Technology-mediated learning is typically assessed in comparative evaluations in a control versus test group of learners.

Accessibility

According to ISO 9241, accessibility refers to the usability of a product, system, service, or facility to people with the greatest extent of abilities. Thus, it can be evaluated with the same methods as those of usability, provided that people with disabilities are included. In addition, the accessibility of mainstream technologies has been specified into detailed guidelines that translate to technical features to which an accessible system must comply with. This is particularly relevant for the Web, where there are many open standards which promote Web Accessibility.[3]

Evaluation Approaches

Main evaluation approaches can be broadly classified into inspections, laboratory tests, and field studies.

Inspections

An inspection of virtual heritage typically takes place during the design process, by experts who experience the system, when this is not yet fully functional

or operable. Inspections are productive processes, but unless a systematic procedure is followed, it can easily get out of hand.

Inspection methods have been proposed in HCI several years ago; two of the most common are the cognitive walkthrough (Mahadoty 2010) and heuristic evaluation (Nielsen 1994). The latter includes:

- A set of guidelines or heuristics to which the system must comply with, such as system visibility, consistency, persistence, and recovery from errors.
- A systematic common process followed by evaluators, which:
  - firstly, includes atomic inspection and identification of findings matched to the heuristics list, and
  - secondly, requires from evaluators to meet and combine their findings into a common list with severity marks and priorities.

In virtual heritage, we neither have well-established sets of guidelines or heuristics nor can we easily identify a small number of experts that possess all required knowledge and skills to perform an inspection; thus, inspections are uncommon in scientific literature. Until we have these resources, inspections in virtual heritage are typically performed in practice by expert evaluators about aspects of the UX (e.g., usability, accessibility, presence), content (e.g., museum curators and other CH professionals), and interaction technology. To put some method in the process, these experts should work in a systematic manner, perhaps with a set of goals and guidelines that are produced in-context, to produce a common set of design recommendations that the system under evaluation must conform to.

Laboratory Tests

Laboratory testing takes place at the final stages of the design, or a design iteration. It involves test users who provide direct and indirect findings and feedback. The process has been detailed in many HCI textbooks (e.g., Albert & Tullis 2013; Cairns & Cox 2008; Nielsen 1994). It is generic and can be adapted considerably depending on target technology, availability of users, time, and other resources; it also requires considerable preparation.

It essentially involves the following steps:

(a) defining the place, time and prototype for evaluation,
(b) recruiting representative users,
(c) defining user tasks for testing,
(d) defining measures and data analysis,
(e) conducting the test – for each user the process is identical and data collection takes place,
(f) summarizing the test,
(g) data analysis and reporting.

It is the most common method for general evaluation of interactive systems, where the evaluation may be repeated in an iterative development process. In literature about virtual heritage evaluation, it seems that most approaches are laboratory tests of various configurations (Nikolakopoulou & Koutsabasis 2020). Testing is often comparative, between/among alternate systems, such as reported by Jylhä et al. (2015). They evaluated their wearable interface for exploring urban POIs by assigning 12 users in two groups: a baseline group that used a mobile phone app and the test group that used the wearable app. Comparative testing can also be performed between/among user groups, for example, expert/novice users (e.g., Panayiotou and Lanitis 2016, on 3D animated paintings) or between adults/children (e.g., use of a gesture-based app of Koutsabasis and Vosinakis 2016) or within system configurations. It may also be formative, that is, emphasizing on qualitative analysis and generating design recommendations, or summative, that is, emphasizing quantitative analysis and statistical testing.

It might also happen in online platforms, if the system under evaluation permits, although the researcher loses contact with users; or it might occur in the field, which is useful but must be well planned so that the process is controlled for all users.

Field Studies

A field study is a general process of observing actual users interact with technology located in the real place, time, and context, gathering data and reaching conclusions about aspects of user interactions with technology. Field studies are invaluable for assessing the quality and impart of virtual heritage. They may be employed not only to assess dimensions of user interaction with a recently introduced system as well as to evaluate existing interactive technologies in cultural places and sites. Essentially, a field study boils down to sophisticated observation and technology-mediated recording of user activity.

There are several field studies in virtual heritage literature, such as, for example, the work of Rubino et al. (2015); they integrated a location-based mobile game in the museum visit and evaluated visitors' behaviour and learning by inviting them to play the game during their visit. Or consider the work of Caggianese et al. (2018), who installed a gesture-based interactive holographic projection in the museum and evaluated visitors' engagement and the attractiveness of the system. While some aspects of the study can be controlled, especially the tasks that users or visitors are asked to perform, it is still a form of field testing. Another example of a field test is the work of Koutsabasis and Vosinakis (2018), who invited teenage museum visitors to digitally sculpt Cycladic figurines in an interactive kinaesthetic game.

Evaluation Methods and Techniques

Various methods and techniques for the evaluation of virtual heritage have been employed. In the review by Konstantakis and Karidakis (2020), they examine a long list of evaluation methods. According to the review of 83 cases of virtual heritage evaluations (Nikolakopoulou & Koutsabasis 2020), most make use of questionnaires (39.9%), observation (19%), and interviews (16.1%).

Questionnaires

Questionnaires may be either standardized or developed by evaluators. The former has gone through the process of psychometric validation for several dimensions of evaluation, like UX (e.g., User experience Questionnaire, Strepp 2017), usability (e.g., System Usability Scale, Brooke 2013), presence (Witmer & Singer 1998), and so on.

Observation

Observation can be organized into many metaphors (Shafer 2009), such as: 'fly on the wall' (observe unnoticed), 'shadowing' (discretely follow a user), and 'secret agent' (play the role of a user).

Interviews

They can take many forms, such as structured, semi-structured, directed storytelling, group interviews, site walkthroughs, contextual inquiries (Bayer & Holtzblatt, 1995), and so on, and can happen in the field (preferably), in the office, online, or by phone.

Intervention Protocols

These protocols are mainly employed in user testing, besides post-hoc interviews, and they usually include (Van Den Haak 2003):

a) concurrent or retrospective think-aloud, that is, when users are encouraged to speak their thoughts about their interactive experiences, as in Correia et al. (2014), who encouraged users to think aloud during the use of their interactive installation, which enabled them to reconstruct medieval illuminations of old books, and
b) constructive interaction, that is, testing in pairs of users who interact with each other, which is a very productive method for formative assessments (Koutsabasis et al. 2007).

These are all empirical methods of research and therefore

a) they all yield knowledge and results, with a limited scope,
b) they can be all useful, but their application can yield errors,
c) errors of one method can be corrected by another, and
d) different methods must be combined in a comprehensive evaluation.

Other General Issues

Important issues to consider in the evaluation of virtual heritage include:

User representativeness

The selection of representative users is the most important aspect for a successful and valid evaluation. Preferences, knowledge, and skills vary among people; therefore, we often observe significant variability in performance and preference. Thus, it is important to recruit representative users, if it is not possible to recruit people who will actually use the system.

Ethical issues and privacy

Any empirical evaluation requires the participation of users, who must always participate freely and willingly. Cairns and Cox (2008) use the acronym VIP (vulnerable participants, informed consent, privacy) to denote the three major ethical issues that must be addressed at any evaluation process. Another ethical issue, from a scientific perspective, is about evaluation data, which must be either open or readily available to potential reviewers or colleagues.

Evaluator knowledge and skills

Currently, evaluation of virtual heritage is performed by researchers from fields including HCI, cultural heritage, design, psychology, learning, and the like. We do not foresee that this will change in the near future because evaluation is a holistic process that requires multifarious expertise taking into account technology affordances and digital curation, and can provide insights about digital content, user interface, and interaction design, software engineering, usability, and user experience.

Planning the evaluation and managing trade-offs

An evaluation of virtual heritage requires planning in terms of goals, approach, participants, measures, place, and time. As with any practice-oriented activity,

planning must consider practical constraints and trade-offs regarding availability of main resources such as time, users, and technology.

Conclusion

Virtual heritage is a highly suitable domain for contemporary interactive technology development. Cultural heritage organizations are addressing the widest possible range of potential visitors, with an emphasis on younger people and children, who are attracted by interactive technology. Visitors of cultural heritage sites are interested in maximizing their experience in terms of sensitization and learning, mediated by technology. Technology developers pursue novel designs in virtual heritage, which provides a challenging context welcoming novel interaction with technology for a wide range of user requirements. Evaluation of virtual heritage can be performed with various approaches and methods to reflect several dimensions of UX. If it is conducted with care and rigour, evaluation can ensure that the requirements of interested parties are incorporated into the technology solution.

Notes

[1] Free icons obtained from https://www.flaticon.com/

[2] https://museum-id.com/informal-learning-museums-opportunities-risks-gina-koutsika/

[3] https://www.w3.org/WAI/

References

Albert, W and **Tullis, T** 2013 Measuring the user experience: Collecting, analyzing, and presenting usability metrics. *Newnes*.

Beyer, H R and **Holtzblatt, K** 1995 Apprenticing with the customer. *Communications of the ACM*, 38(5): 45–52.

Brooke, J 2013 SUS: A retrospective. *Journal of Usability Studies*, 8(2): 29–40.

Caggianese, G, Gallo, L, and **Neroni, P** 2018 Evaluation of spatial interaction techniques for virtual heritage applications: A case study of an interactive holographic projection. *Future Generation Computer Systems*, 81: 516–527.

Cairns, P and **Cox, A L** 2008 *Research Methods for Human-Computer Interaction*. Cambridge University Press.

Cowley, B, Charles, D, Black, M, and **Hickey, R** 2008 Toward an understanding of flow in video games. *Computers and Entertainment*, 6, 2(2008): 1.

Doherty, K and **Doherty, G** 2018 Engagement in HCI: Conception, theory and measurement. *ACM Computing Surveys (CSUR)*, 51(5): 1–39.

Forlizzi, J and **Battarbee, K** 2004 Understanding experience in interactive systems. In: Proceedings of the 5th conference on Designing interactive systems: Processes, practices, methods, and techniques (pp. 261–268).

ISO 9241. Ergonomics of Human-System Interaction. International Organization for Standardization.

Jylhä, A, Hsieh, Y T, Orso, V, Andolina, S, Gamberini, L, and **Jacucci, G** 2015 A wearable multimodal interface for exploring urban points of interest. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 175–182). ACM.

Karapanos, E, Zimmerman, J, Forlizzi, J, and **Martens, J B** 2009 User experience over time: An initial framework. In: Proceedings of the SIGCHI conference on human factors in computing systems (pp. 729–738).

Konstantakis, M and **Caridakis, G** 2020 Adding culture to UX: UX research methodologies and applications in cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH)*, 13(1): 1–17.

Koutsabasis, P and **Vosinakis, S** 2018 Kinesthetic interactions in museums: Conveying cultural heritage by making use of ancient tools and (re-) constructing artworks. *Virtual Reality*, 22: 103–118, Springer.

Koutsabasis, P and **Vosinakis, S** 2016 Adult and children user experience with leap motion in digital heritage: The Cycladic sculpture application. In: Euro-Mediterranean Conference (pp. 350–361). Springer International Publishing.

Koutsabasis, P, Spyrou, T, and **Darzentas, J** 2007 Evaluating usability evaluation methods: Criteria, method and a case study. In: International Conference on Human-Computer Interaction (pp. 569–578). Springer, Berlin, Heidelberg.

Lewis, J R 2014 Usability: Lessons learned... and yet to be learned. *International Journal of Human-Computer Interaction*, 30(9): 663–684.

Lombard, M and **Ditton, T** 1997 At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2).

Mahatody, T, Sagar, M, and **Kolski, C** 2010 State of the art on the cognitive walkthrough method, its variants and evolutions. *International Journal of Human-Computer Interaction*, 26(8): 741–785.

Minsky, M 1980 Telepresence. *Omni*, 45–51.**Nielsen J** 1994 *Usability Engineering*, Elsevier.

Nikolakopoulou, V and **Koutsabasis, P** 2020 Methods and practices for assessing the user experience of interactive systems for cultural heritage. In: *Applying Innovative Technologies in Heritage Science* (pp. 171–208). IGI Global.

O'Brien, H L and **Toms, E G** 2008 What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the*

American Society of Information Science and Technology, 59, (6, Apr. 2008): 938–955.

Panayiotou, S and **Lanitis, A** 2016 Paintings alive: A virtual reality-based approach for enhancing the user experience of art gallery visitors. In: Euro-Mediterranean Conference (pp. 240–247). Springer International Publishing.

Roussou, M and **Katifori, A** 2018 Flow, staging, wayfinding, personalization: Evaluating user experience with mobile museum narratives. *Multimodal Technologies and Interaction*, 2(2): 32.

Rubino, I, Barberis, C, Xhembulla, J, and **Malnati, G** 2015 Integrating a location-based mobile game in the museum visit: Evaluating visitors' behaviour and learning. *Journal on Computing and Cultural Heritage (JOCCH)*, 8(3): 15.

Pujol, L and **Champion, E** 2012 Evaluating presence in cultural heritage projects. *International Journal of Heritage Studies*, 18(1): 83–102.

Schrepp, M, Hinderks, A, and **Thomaschewski, J** 2017 Construction of a benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4: 40.

Shaffer, D 2009 *Designing for Interaction*. New Riders.

Van Den Haak, M, De Jong, M, and **Jan Schellens, P** 2003 Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*, 22(5): 339–351.

Witmer, B G and **Singer, M J** 1998 Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3): 225–240.