Reliving the Experience of Visiting a Gallery: Methods for Evaluating Informal Learning in Games for Cultural Heritage

Kalliopi Kontiza
University College London
London, United Kingdom
k.kontiza.12@ucl.ac.uk

Antonios Liapis
University of Malta
Msida, Malta
antonios.liapis@um.edu.mt

Catherine Emma Jones
University of Luxembourg
Luxembourg, Luxembourg
catherine.jones@uni.lu

ABSTRACT

When evaluating the effectiveness of gamified app experiences in cultural heritage venues in terms of informal learning outcomes, a core challenge is the complexity involved in assessing intangible measures such as visitors' appraisal of artwork. A comprehensive summary of the literature for conducting museum visitor evaluations is needed in order to understand how to measure the impact of gamification on user engagement, and the enhancement of the cultural heritage experience on learning. This paper first reviews related literature regarding the application of intrusive versus non-intrusive user evaluation methods, focusing on the REMIND protocol for conducting experiments with museum visitors. We relay our findings when applying the REMIND protocol in four gamified cultural heritage applications in the CrossCult project. Focusing on the assessment of informal learning in an application specifically designed for the visitors of the National Gallery of London, the paper concludes with recommendations, challenges, and future steps in evaluating games for cultural heritage.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; • **Applied computing** → *Interactive learning environments.*

KEYWORDS

Evaluation protocols, reflection, cultural heritage venues, Think Aloud, Remind.

1 INTRODUCTION

According to the UNESCO foundation, Cultural Heritage (CH) is "the legacy of physical artefacts and intangible attributes of a group or society that are inherited from past generations, maintained in

the present and bestowed for the benefit of future generations". Digital applications have serviced CH through visualisation (3D modeling, animation, virtual worlds) [30], through mobile tour guides [28], as well as through games [11]. Games for CH have the power to motivate prolonged and deep interactions with CH artefacts, narratives and concepts, immerse players in a virtual heritage site, prompt collaboration with fellow players and facilitate learning [2]. On the other hand, designing games for CH is a very challenging endeavour, especially when the games are expected to be played in-situ (at the CH venue). Not only does such a game have to balance entertainment and learning goals but it must also respect the rules and constraints of the venue, such as keeping one's voice down, navigating throngs of visitors, low-light conditions and rules on no flash photography, limited internet connectivity etc. If designing games and applications for CH is challenging, evaluating them is even more so. While usability or technical quality can be measured via established questionnaires and methods, the informal learning potential or the thoughts and reflections triggered by the CH application are far more difficult to assess. Efforts have been made to provide evaluation frameworks for CH games [2, 11, 20], but little attention is given to high-order thinking such as creativity, re-framing [46] or how visitors integrate new information into their existing beliefs and knowledge [65].

Museum-related research tends to focus on visitors' behaviour, which results in external observations. Analytical methods may include direct observation, where "the observer stands next to the visitor and notes his or her reaction while viewing an artistic work" [57]. This external approach largely disregards visitors' personalities or the deep cognitive and affective processes that occur when a person seriously considers and reflects on a topic or an experience. This paper discusses solutions to this conceptual challenge by comparing two thinking 'out loud' protocols. Specifically, our comparison explores how a temporal gap between the experience of the visitor and the verbalisation of this experience affects evaluation. Driven by the goals of the large-scale project CrossCult to identify whether games and gamification enhance the CH experience and its impact on learning, we investigate the application of these qualitative evaluation protocols on gamified CH apps and CH games for four distinct use cases. Throughout this paper, we use the broad definition of gamification by Deterding et al. [24] when referring to the CH apps developed under CrossCult, as they integrate—each app in its own way—a variety of game interfaces (e.g. badges), game models (e.g. curiosity), and game design methods such as prototyping and playtesting [37].

The CrossCult project was funded by the European Commission's Horizon 2020 programme, and ran from March 2016 until

February 2019. The project engaged 11 partner institutions, including computer scientists, social sciences and humanity researchers, historians and private companies. CrossCult implemented four use cases involving a total of 8 venues across Europe, which will be shortly presented in this paper. CrossCult applications used cutting-edge technology such as semantic reasoning and recommender systems to connect existing digital cultural assets and to combine them with interactive experiences. Its ultimate goal was to increase retention, stimulate reflection and help European citizens appreciate their past and present in a holistic manner. In light of the goals of CrossCult, the issue of evaluation was a central challenge and a large focus of the research carried out throughout the project. This paper analyses how the specific goals of each use case of CrossCult—and the venues where it took place—determined the evaluation protocol for gauging informal learning and reflection (our working definition of the term is discussed in Section 4). Specifically, the paper focuses on the two thinking 'out loud' protocols tested in CrossCult and discusses our findings after we deployed them in practice.

2 EVALUATING THE VISITOR EXPERIENCE

The study of museum visitors has been a developing field since the 1980s but few works focused on the dynamics of the visitor experience up until the early 2000s, when visitor evaluation studies provided a number of well documented, large-scale studies [33, 40, 51]. These large-scale studies use multiple data collection methods to enhance our understanding of the overall experience of visitors of museums or heritage sites. Most of the methods and techniques available to researchers in the field derive from research in cognitive psychology, sociology, artificial intelligence and informal learning [26]. Museum case studies that take into account the visitor's perspective [60, 63] have shown that it is not possible to interpret a visitor's behavior solely from telemetry or from direct observation.

The findings of different systematic literature reviews in surveys on the evaluation of serious games—which included papers focused on the assessment of their educational effectiveness—show that questionnaires are the main assessment method (in 90% of cases surveyed), followed by interviews [16], or quantitative approaches such as the quasi-experimental design [15]. The most commonly assessed quality characteristics include usability, learning outcomes, and user engagement [19]. Visitors' lived experience (pleasant-social-learning) is such a complex element that it cannot be evaluated by questionnaires, as they give emphasis on interaction and thus solely measure user satisfaction and acceptance. When questionnaires are used as an evaluation instrument to assess educational effectiveness, they need to be standardised [55] or mapped to an accepted education framework [18].

Quantitative quasi-experimental methods often adopted in visitor studies to assess games' learning effectiveness [6] may not have the rigour of controlled experiments, but they maintain the argument and logic of experimental research. Whether the research design is experimental or quasi-experimental, the most common strategy in this type of investigation is a comparison between groups: one group provides baseline information (acting as the control group), whilst the other group is given the experimental treatment [31, 70].

Another key method in quantitative research is the (large- or small-scale) survey [2] as it provides information on the distribution of a wide range of respondent characteristics. Large-scale surveys can establish general characteristics of CH venues' visitors (such as duration of the visits declared by visitors and their actual visit time, the frequency of museum visiting and its variation between visitors with different social characteristics) [64]. However, it is argued [27] that small-scale surveys often under-report the methodology and findings; most importantly, such surveys remain within the institution, inaccessible and unavailable.

One of the most important sources of information in qualitative research is the interview, which can take a number of forms, including open-ended, focused, or survey [42]. In-depth interviews are typically a means of understanding visitors' experience and gaining insights from individuals. In the field of visitor studies, in-depth interviews have also been adapted from the methodology used by developmental psychologists in clinical settings [34]. Sometimes, interviews use focus groups when the aim is to detect visitors' behavioural patterns or insights into their attitudes and perceptions. Significant problems common to all semi-structured and unstructured interviews are that memories of the visit are partial and visitors cannot always remember what they have just seen. In such cases, visitors are eager to please and will offer answers they think the interviewer expects to hear [48].

Since cultural heritage is an informal learning environment, we need to address the variety of experiences that allow visitors to learn: experiences that allow them to *absorb* (see, hear, feel, taste, smell), to *do* (perform an activity) and to *interact* (socialise) [72]. The multidimensional nature of the visitor experience is augmented by visitors' expectations based on previous experience. However, research has shown very little about how e.g. visitors integrate new information into their existing beliefs and knowledge [65]. Emphasis needs to be given to a more holistic approach to research, encompassing the personal, contextual, time related aspects of a CH visiting experience [59]. It is clear that new approaches are necessary to move existing methods of evaluating usability (often originating from computer science) further, in order to cover the multifaceted cultural experience of users that encompasses learning and affective elements.

3 THE OUT LOUD PROTOCOLS

Qualitative approaches that have been applied to visitor studies share three assumptions: (a) an over-arching view that seeks to understand phenomena in their entirety in order to develop a complete understanding, (b) an inductive approach that begins with specific observations and moves towards the development of general patterns as the research progresses, and (c) a naturalistic enquiry that seeks to understand phenomena in their natural environment [50]. Qualitative research usually takes place with small samples of people, nested in their context and studied in-depth.

Observing visitors as a method in qualitative visitor user studies research has its origins in anthropological field research, traditionally associated with the practice of participant observation [56] and pioneered by researchers such as Malinowski in 1922 [71].

Tracking studies have the great advantage that they can be administered unobtrusively. Some studies using observational methods in visitor research pioneered (i.e time-lapse photography used by Nelson in 1946 [54]) but identified that visitors themselves remain the least known, most unpredictable and most difficult aspect of museum work to study. Although tracking techniques using video or audio recording instruments augment the work of the researcher (who must record at different times on different days), a substantial amount of data can be gathered and subsequently analysed [32]. Technology makes it possible to record and analyse with a degree of particularity not available to the field researcher.

The value of using video footage or audio recording provides a non-intrusive method of studying visitors' behaviour and their thoughts, and of collecting data that can be re-examined and replicated. However, the method has the potential to prioritise what is captured on video or audio at the expense of the lived experience of the participant. A critical part in conducting this type of evaluation is whether an active or passive role for the observer is adopted. This section examines and compares two Think 'out loud' protocols, which leads to a qualitative evaluation study (in Section 4) in which they are applied to assess visitors' reflection while they interact with gamified CH applications.

3.1 Think Aloud

Think Aloud is a method in which participants speak aloud any words in their mind as they complete a task. The type and level of difficulty of the research task, the degree of prompting which is appropriate, the use of other data to support inferences from think-aloud protocol need to be decided by the experimenter [17]. The theoretical framework underpinning the Think Aloud method was proposed by Ericsson and Simon [29] in order to address shortcomings of traditional interview-based methods within mainstream psychology. Rankin [58] suggested treating each Think Aloud participant as a "small, tightly focused" case study. The objectives of a Think Aloud interview is to observe the participants' behaviour and to record their thoughts and (re)actions formulated aloud. Once completed and recorded, these itinerant open-ended interviews are transcribed and analysed via thematic content analysis [53].

One of the theoretical challenges, however, resides in bridging the temporal gap between the CH experience of the visitor and the verbalisation of this experience that will reveal visitors' internal thinking and understanding. As Schmitt points out [61]:

> If visitors are questioned at regular intervals during their visit, we interrupt the flow of their experience, engaging them in a reflective analysis of their actions. This practice thus modifies the activity and leads us further away from a visit in a natural setting. If, on the contrary, we preserve the natural setting of the visit and question visitors at the end of their tour, we collect a summarised a posteriori discourse devoid of the finesse and precision of the lived experience.

3.2 REMIND

The REMIND approach developed for museums by Schmitt [63] aims to understand, in situ, the physical/sensitive, but also cognitive and emotional components of the museum experience. The goal of REMIND is to investigate how these components are constructed by audiences to create meaning for their journey through an exhibition. REMIND stands for Reviviscence, Experience, Emotions and sEnse MakINg micro Dynamics. The latter is a method of qualitative studies to understand the dynamics (the cognitive and physical dimensions) of the experience of a museum visitor. The epistemological framework that supports REMIND derives from cognitive, perception and linguistic theories. According to Maturana and Varela [49], the words are the clues that the language uses to coordinate our actions but the language does not record the world, it does not allow us to make objective descriptions—objective being void of meaning in the enactive framework. The language has no power to reveal the world; language allows for the coordination of actions that form the world that is ours when we 'speak'. Maturana and Varela view language as a consensual domain, and linguistic behaviour as connotative rather than denotative (i.e. objectively indicating and pointing to an external world). The observed communication of meaning and the practical efficacy of language do not reside in the words and terms themselves but reflect similarities in the organisms' structures developed through their history of interactions [52, p. 329].

Schmitt [62] focuses on the visitor experience from an enactive perspective, one that makes sense from the point of view of visitors during their visit. His method is based on the theory of enaction [60], according to which perception and action are coupled: we interact with our environment according to our history, our expectations, etc. The term enaction, introduced in 1989 by Varela [68], is an epistemological approach whose fundamental idea is that "the cognitive faculties are inextricably linked to the history of what is lived, in the same way as a path beforehand nonexistent appears when walking" [69, p. 111]. The enactive approach establishes a circularity between action and knowledge: "all action is knowledge, and all knowledge is action " [49, p. 12].

Schmitt [60] states that an analysis of the visitor's 'course of experience', carried out by means of a subjective re-situ interview, allows the researcher to describe and understand the basic units of the visitors' experiences, as well as the overall experience of the visit with precision. In addition to the study of the visitor's 'course of experience', REMIND also highlights the actions and cognitive paths of visitors which contribute to the construction of the knowledge. Thus, we do not 'see' the world, but we 'live' it. The REMIND method postulates that this world experience can be observed by asking the subjects to relive their experiences. According to this theory, revising the film of a lived experience sets in motion the same neural networks as the experience itself: the interviewee does not 'see' the experience again, she 'lives it again', to be able to describe it more deeply.

In terms of analysing the transcripts collected from users while they view and narrate their recorded experience, Schmitt [62] argues that when someone is asked to explain their activity *a posteriori*, they spontaneously break it down into units of activity that are significant from their point of view. The codes are as follows:

Representamen (REP) corresponds to what is considered by the respondent in the environment (the museum object, a feature in the app, or a building in the city, etc.)

Engaments (ENG) refers to the type of engagement that the user has with the focus of their attention (Representamen). It is often an action verb (to look, do, try, activate, turn, click…).

Expectations (EXP) is the meaning of the engagement in relation to representation, the reason why the person does this with that, and what they expect to happen during the engagement.

Reference frame (REF) is the knowledge of the moment, which incorporates the interviewee's past experience which encompasses knowledge, beliefs, judgments, past experiences and personal knowledge.

Interpreten (INT) is the knowledge built, validated or invalidated during the experience.

Unit of experience (UE) is the part of the activity that is shown, told or commented on by the interviewee. It can be a symbolic construction, an action (practical or communicative) or an emotion. It is the intellectual journey of the visitor in his experience.

4 CASE STUDY: CROSSCULT

Both Think Aloud and REMIND have been applied in the context of the European project CrossCult, which aims to foster reflection on cultural heritage and history through interconnections among cultural digital resources, physical venues and citizen viewpoints. This section presents the CH apps developed for the four different use cases (UC) of CrossCult, focusing on our approach for evaluating informal learning and reflection.

Reflection is a process that occurs when we act upon different information to synthesise and evaluate it [14]. Apps developed for CrossCult attempt to stimulate this process by encouraging the discovery of historical content in a new interactive way, looking for commonalities, differences, and interrelations beyond their superficial elements. This can often involve the use of higher levels of cognitive and affective processes. Reflection in the context of CrossCult can be more or less profound. Our working definition considers that reflection is an internal process occurring when a person thinks deeply, or seriously considers a topic or an experience; it often includes deep cognitive and affective processes, which encompass derived notions such as our prior knowledge [7, 67], our emotions [4], our personal experiences [13], our beliefs [65], our interpretation of the content based on own world view [41].

Due to the multifaceted nature of reflection as a concept, the challenge of evaluating if the CrossCult apps met this goal in and of itself was complex. We had to find evidence that when using the apps, participants engaged in a thought process—an intellectual journey operationally defined as a train of thoughts— which led to deep thinking. The goal of each use case presented in this Section was to detect and assess whether the designed apps could trigger reflection and/or reinterpretation on crosscutting topics of history and cultural heritage by building upon cognitive processes such as critical thinking, contemplative thoughts, analysis or synthesis of information or effective processes such as expressions of feelings or empathy.

4.1 Reflection and the CrossCult apps

Due to the different venues and experiences targeted by each use case of CrossCult, the result of the participants' intellectual journey varied in type and form. In CrossCult, the participants' experience may lead to the reinterpretation of their city, the affirmation of existing knowledge, the discovery of new information or new learning, the expression of an opinion, the expression of connections between objects and stories, etc. To encourage this reflection journey, each CrossCult app (CC app) used different stimuli and forms of interaction, while it also considered different facets of the concept of reflection. A common goal for all CC apps, however, was to encourage participants to think about historical narratives. Participants' thought process was triggered by: (a) providing content that participants personally connect to their personal experiences through inter-historical stories and/or (b) exploring reflective topics and stories which facilitate consideration of historical content and narratives. This was achieved with the following playful interactions with CH content that participants encountered before, during and after their visit to the CH venue:

(1) Gamification elements [24] in the apps such as badges and level-up systems (see UC4 for more details). These elements would motivate the player to co-create local knowledge about the city, to bring new knowledge or affirm exiting knowledge, to make connections to one's own experiences and opinions, provoke and confront the opinions of other players.

(2) Content provided in various forms (videos, audio, games, images, text) (see UC2 for more details). These elements would assist people of different cognitive needs and allow the exploration of content at a later point in time, after the in-situ experience.

(3) Autotelic creation games (see UC1 for more details). These games would offer intrinsic rewards from the aesthetically pleasing and personalised arrangement of paintings on a virtual wall, and encourage visitors to think about the current display of a collection and reinterpret it by building their personalised virtual version.

(4) Adver and Profiling games played on social media (see UC3 for more details). These games would capture visitors' attention, profiles the users, advertise the venue and the apps, and attract different types of visitors beyond traditional visitors of archeological museums.

4.1.1 Use Case 1: Large multi-thematic venue. The first use case (UC1) took place at the National Gallery (NG) in London, United Kingdom. UC1 aimed to use the broad collection of a single large institution in order to illustrate the connections between paintings, painters, places and events across European history and to stimulate public reflection on different themes that connect paintings. To achieve this, the CC app (see Fig. 1) aims to stimulate reflection by recommending groups of paintings according to user preferences, based on a user model trained via a number of annotation tasks the first time the CC app is used. Moreover, the CC app recommends stories from a set of reflective topics that (a) encourage personalised discovery of the NG collection and (b) enable a different way of viewing the art collection. Reflection is triggered by encouraging visitors to think about topics that personally interest them, and by
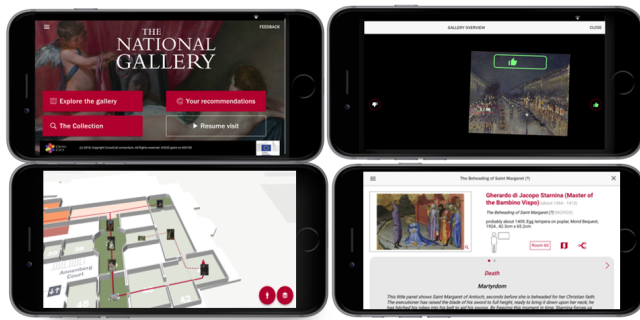
Figure 1: In the UC1 app, a player customises their user model through annotation tasks (top right). The app then exploits the user model to recommend routes in a virtual NG map (bottom left) or present stories on reflective topics such as Death (bottom right).
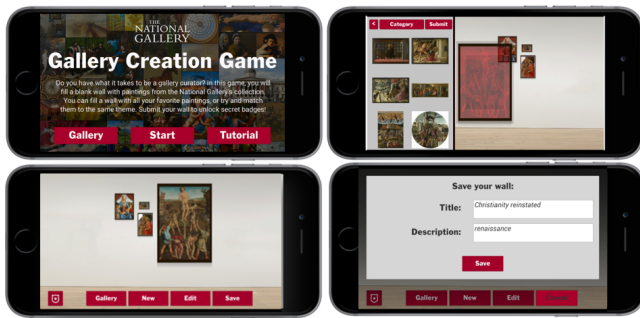

Figure 2: The 'Gallery Creation Game' for UC1 prompts users to place paintings from the NG collection (grouped according to the UC1 reflective topics) on a virtual wall (top right). When a player saves the wall into the device's memory, they provide a title and a short description (bottom right) which prompts reflection on their theming priorities.

connecting seemingly disparate paintings together [44]. In the first evaluation phase, the goal was to assess the perceived value of the app. Findings from this first phase—which involved both experts and visitors–highlighted the need of a more intuitive user experience that will seamlessly link the digital to the physical experience.

As a companion to the UC1 app, a freeform creation game was designed and developed. This 'Gallery Creation game' [43] acted as (a) a tool that visitors can use before, during or after the visit to familiarise with the paintings in the collection, (b) an additional opportunity for personalisation. The UC1 game extends the physical experience of the gallery into the virtual, enabling users to recreate their experience by rearranging the paintings into personalised virtual rooms (see Fig. 2). During the first evaluation phase, a quantitative user study was performed remotely, where users completed the task of creating a virtual room with NG paintings and provided a title and a description. Some of the participants' comments were encouraging, e.g. "*I liked that a game involves art and personal interaction, I liked the fact that it demands organised thinking*" [43]. The evaluation provided some initial evidence that

Figure 3: The UC2 app allowed users to read stories on reflective topics such as water and health in Antiquity (top left). It also uses graphs to show associations among places and prompts multiple-choice questions (top right). Reflections can be provided at all times (bottom left).

the UC1 app, which allowed users to discover information about the paintings and learn new things, was complementary to the UC1 game that provoked them to think and proactively create. However, we need to further gauge users on which elements of the game prompted them to think, how their reflective process started, and which criteria they used to hang paintings on the virtual wall.

4.1.2   Use Case 2: Many small venues. The second use case (UC2) connected four small venues, namely the Roman healing spa of Lugo (Spain), the Roman healing spa of Chaves (Portugal), the archaeological site of Montegrotto Terme (Italy) and the ancient sanctuary of Epidaurus (Greece). UC2 aimed to highlight the connections among the respective bodies of history and culture, as well as traits of human behaviour (captured in the archaeology of the sites) that are still recognisable in our current society. This goal was achieved through a CC app targeted for a post-visit CH experience, which engaged visitors in a quiz game based on the exploration and completion of graph visualisations of relevant concepts and interconnections (see Fig. 3). The CC app was a multi-venue mobile experience that utilises game play and content discovery to stimulate public reflection on the topic of archaeology, water and healing. The first evaluation phase provided us with feedback on the gameplaying modes of the app, the graphs and visual components; a significant number of participants expressed positive emotions [22], triggered mainly by correct answers when interacting with the graph, discovery of new content and reading the stories [9, 47]. For the final version of this CrossCult app, synchronous social interactions and multi-user features were removed, since playing in a team was not appreciated by the test users.

4.1.3   Use Case 3: One venue, non-typical transversal connections. The third use case (UC3) targeted a peripheral and largely unknown venue, the Archaeological Museum of Tripolis (AMT) in Greece, which remains unpopular despite its unique and important collection. UC3 designed a holistic experience that supports the visitor before, during and after her visit. It used games, a social site, and a mobile application to achieve its purpose, i.e. to stimulate public reflection on the topic of women's social place in antiquity. It achieved this by encouraging people to think about topics related to
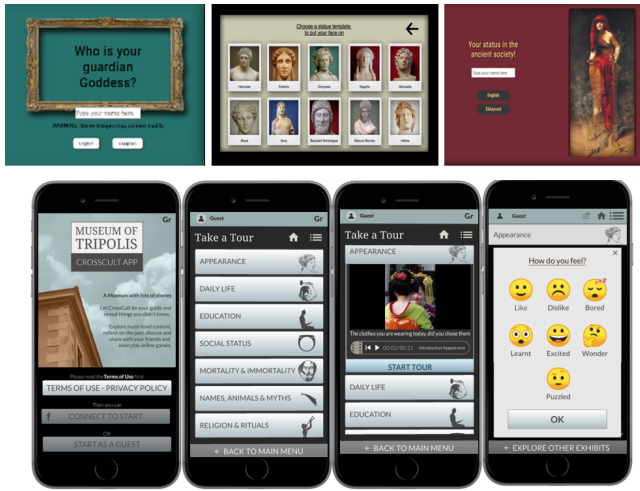
Figure 4: UC3 had a number of mini-games on the browser (top), from left to right: "Who is your protecting goddess?", "Face in a statue", "Your ancient status". The main app for the AMT (bottom) offered personalised tours (based on the mini-games) that combined museum artefacts, digital media, and stories. As in UC2, users provided feedback at all times.

women and by connecting physical and digital cultural items, other venues, current social phenomena, etc. (see Fig. 4, bottom). Such connections were discovered algorithmically within knowledge-bases, using word embeddings and players' personal preferences [21]. A number of mini-games played on the browser (see Fig. 4, top) would act (a) as tools for user profiling by identifying specific visitor interests and needs to be supported during the actual museum visit, (b) as media to familiarise the visitors with the museum content and the narratives that could be explored during the visit, and (c) as an advertising technique not only for the venue but also the CrossCult suite of apps [8].

A quantitative user e-study revealed that the mini-games were indeed very effective as advertisement and profiling tools. Regarding profiling, the mini-games were successful in predicting participants' personality styles (based on actual psychometric tests) in 88% of instances [3]. Based on this satisfactory result the mini-games were connected to the app, through an online social site of the Cross-Cult ecosystem, for profiling purposes in order to deliver a more personalised experience to the visitors. The profiling service these games provided was an invaluable input to pre-select the topics that the visitors could be most interested in from among the many narratives available in the mobile app.

4.1.4 Use Case 4: Multiple cities, "Past & Present" interplay. Use case 4 (UC4) focused on the outdoors, taking place at the cities of Valletta (Malta) and Luxembourg (Luxembourg). The premise of UC4 was the use of location-based game play to stimulate public reflection on a variety of topics such as migration, art and architecture, industrial heritage, language and identity etc. Through the promotion of citizen participation (co-creation), we imagined that such a playful and thought-provoking game could contribute
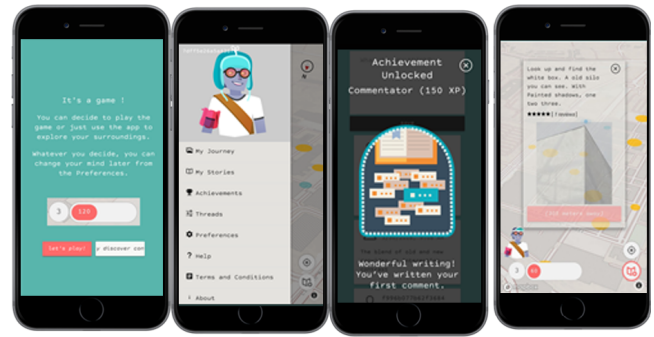

Figure 5: The mobile location-based game of UC4 involved avatars, badges, levels, which motivated the player to walk around Luxembourg or Valletta to discover authored stories at nearby points of interest (3rd image) or user-created stories in the city. Users could rate, answer questions and comment on all authored stories, and rate user stories.

to a collective reflection and re-interpretation of history. Interaction with the CC app for UC4 brought the topic and the city to the forefront of people's mind (even if for a short time) and led to a re-familiarisation with history that is otherwise not part of participants' everyday city experience.

During the first evaluation phase, participants reported that the CC app gave them the opportunity to "*walk around and discover with some fun, entertainment and learning*". When asked to describe the app, the most frequently used word was 'discover' in connection with words such as 'fun', 'explore', 'history', 'city' and 'interest' [39]. This location-based app used board games as a valuable tool of its co-design process [37, 38]. The app's iterative design and development cycles—supported by regular testing and fast prototyping—allowed for more flexibility to add/remove features according to priorities and playtest results [39]. For the gaming elements, the final version of this CC app allowed for two modes of interacting with content creation by switching between game play or simple story discovery.

4.2 Evaluating Informal Learning in CrossCult

The CrossCult project adopted a mixed-method evaluation approach, depending on the phase of the evaluation, the technology readiness of the applications, the type of application and venue that it was designed for. This mixed-methods approach provided complementary results and shed light on different aspects of the cultural experience. During early stages of the evaluation, it was critical to assess the perceived usability, usefulness, learnability of the CC apps. This evaluation also showed that participants believed that they learnt or they reflected, although this was not formally assessed. In addition, the apps received positive feedback and were largely accepted by the participants [35].
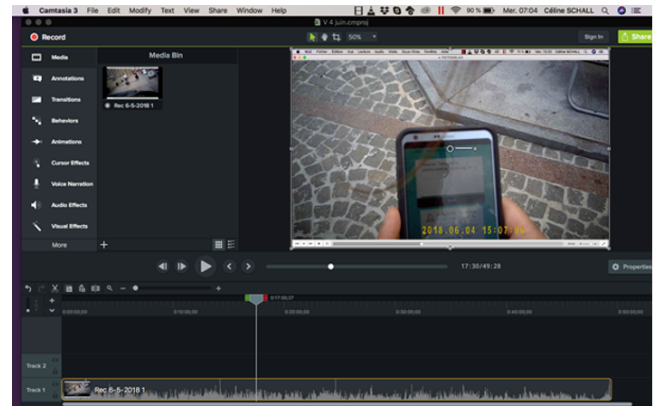
The second evaluation phase aimed to identify the elements of reflection triggered from the intellectual journey of the participant's experience. We wanted to explore in more detail the process of reflection that is triggered by the CC apps, e.g. if they stimulate changes in conceptions, thought process, or comparisons. All of these changes encompass lateral thinking [23] and can contribute

to learning, appreciation, or appropriation of knowledge. One of the questions for this evaluation phase was: *What are the cognitive and affective process that are activated when users experimented with the CrossCult gamified apps?* Evaluating the application simply by observing the traces left by users (visit time, stopping points, use of different functionalities etc.) was insufficient; we could not interpret if people were really engaging in deep thought from such logs or from direct observation.

In this direction, the second evaluation phase followed a multi-pronged approach using different forms of qualitative research methods, developing new or modifying existing evaluation methodologies. We developed a protocol considering the specificities of the apps, the type of CH venue as well as the skills and resources available. We accessed the experience of the different app users, and more specifically their thoughts, by adopting two different but complementary qualitative approaches: (a) an adapted version of the REMIND protocol (UC1 and UC4) and (b) the Think Aloud protocol (UC2 and UC3). Where feasible, we used Likert scale questionnaires or sentence completion questionnaires as an interview probe to support the main data gathering methods.

Through the in-depth qualitative analysis we were able to assess the extent to which the app experience supports reflection in our participants, and thus assess how well we met our project objectives. We analysed the participants' interactions with the CC apps via a combination of methods (questionnaires, post evaluation debriefs, classification of open-ended questions, Think Aloud and REMIND). The complete report of the experiments results and analysis can be found in the project's dedicated deliverable [36], once it becomes publicly available. For UC4 we conducted a primary, detailed qualitative study on 13 users via the REMIND method, and another study using the Think Aloud method with 5 users who engage with the app for the first time. These experiments in UC4 provided the guidelines for the other use cases. The REMIND protocol was then used in UC1 with 12 participants as there are many parallels between the UC4 and UC1 applications and it was considered suitable. The REMIND protocol could not be used in the venue of UC3, due to copyright restrictions on the original artefacts, and was partially used for UC2 in combination with the Think Aloud protocol, *after* a visit (REMIND is designed for evaluating in situ experiences).

To apply the adapted REMIND method to the CC apps, we equipped participants with glasses that had an integrated video camera and let them visit the museum/city whilst using the CC apps (Figure 7). After 20 to 30 minutes, we halt the visiting experience and ask them to review the recorded film. As they review the film, they think aloud and describe their experience based on what they see in the film. This interview is recorded (with the Camtasia[1]), which was then used to synchronise the participant's description and the video of the experiment (see Fig. 6a). The think aloud recording is then transcribed and analysed alongside the film and the app interactions. To transcribe, we used Advene [5] which synchronises the video and the transcription and facilitates the development of qualitative coding (see Fig. 6b).

(a) Camtasia tool for recording participants' responses

(b) Outcome of coding verbatims from interviews

Figure 6: Tools used for the REMIND protocol in UC1 & UC4

4.3 Application of REMIND in UC1

To provide a better context and dig deeper into the evaluation process, we present an experiment for the last phase of evaluation conducted for UC1. The process and findings showcase the rich evidence that emerged from the data collection when the REMIND protocol was followed.

4.3.1 Participants. Experiments using the REMIND protocol were made with 12 volunteers (10 female), with ages ranging from 30 to 65 (P numbers are 1 to 13 since P04 withdrew). Experiments took place in June, August and December of 2018. We used a snowball method of recruiting participants via the researchers' professional networks. Participants were informed that they will be testing an application which will allow them to discover paintings from the National Gallery collection—customised to their preferences—and stories related to the paintings (CrossCult app for UC1, see Fig. 1).

4.3.2 Data collection. After the participants signed the consent form, the experimenter would turn on the camera of the eye-glasses and give them to participants to put them on, so that the experiment would start. The experiment was split into 3 parts: In the first part, we asked participants to follow the profiling process and complete all the steps until they receive recommendations by the app. In the second part, participants were asked to use the digital map to physically explore and follow—if they want to—the recommended

(a) Participant interacting with both the exhibits and the CC app

(b) Participants watching a video and walking through their experience
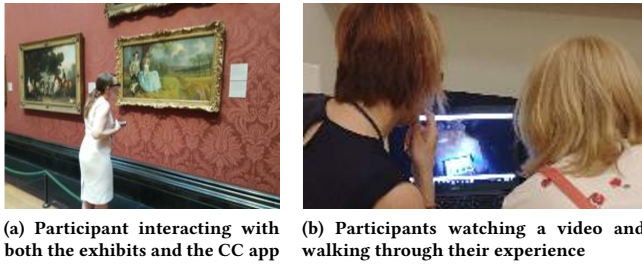
Figure 7: Evaluation of UC1 in the National Gallery of London using the REMIND protocol.

group of paintings that the app chose for them (see Fig. 7a). In the third part, they were asked to select and follow one of the app's stories. The whole process lasted 30 to 45 minutes. The visit was then halted, the camera on the eye-glasses was turned off and we asked participants to watch the film of their interaction: during this period they were invited to comment out loud on their experience, as captured by the video recording (see Fig. 7b).

4.3.3 Data processing. The interview transcripts were processed in order to understand the reflection processes that were triggered while the UC1 app was used. Transcripts were coded so that we could identify and deconstruct the user experience. Data analysis with the REMIND protocol applied the 6 codes presented in Section 3 while the Interpreten (INT) code was broken down further into three different units:

Emotional reactions (ER) as the explicit emotion expressed by the user during the pilot app experience (grouped as positive, negative, neutral).

Trace of reflection (TR) is the part of the intellectual journey that corresponding to reflection. Each use-case standardized the concept of reflection using open axial coding to establish a set of codes for each pilot (examples of UC1 codes are 'Understanding', 'Remembering', 'Analysing', 'Evaluating').

Interactions (INR) is the activity that participants engaged with during the unit of experience (e.g. comments, tags, a response to a reflective question).

These codes emerged from the data and the verbalisations of reflection. Finally, we added a code associated to the physical place, **Localisation (LOC)**, to identify the observed status of participants when the unit of experience is occurring (e.g. if people are walking or they have stopped).

4.3.4 Findings. During the REMIND experiments, the use of the UC1 app triggered an **emotional reaction** to participants. We observed 95 explicitly expressed emotions stimulated while the participants interacted with the different features of the app; 54 of these were positive (57%) [36]. Participants most often expressed surprise, interest, intrigue, degree of comfort and engagement. These positive emotions were **associated with the act of information discovery** while reading the painting description on the app and looking at the physical painting itself, or exploring other paintings in the same room. For example, P03 notes: "*and generally I was surprised that the more detailed text was specifically about the*

painting, I was surprised in a good way…"; P11 notes: "*this one I remember being surprised by …I am looking at it at the app and then I find it next to the other painting and I realise that it is tiny*". Positive emotions were triggered by the recommended groups of paintings, when participants found artworks they had not noticed before which sparked their curiosity. For example P09 notes: "*I was happy because I've never paid attention to this painting before*".

The discovery of personalised paintings led participants to **elaborate** upon the **connections between personal experiences** and what they have read or observed about the artworks. This fostered their curiosity by encouraging discovery of other paintings in the same room. During this personalised discovery, the participants expressed positive emotions—sometimes associated with personal feelings and memories—which supported their intellectual journey and allowed them to **reflect on prior learning and previous experience together with new information**. For example, P07 comments: "*it was absolutely fascinating to see all three of them. You can see me here looking, and I stand back and look at the likeness between the two brothers and the mother. Historically I know (but not that she was a widow) that she was in Exile for so long. So it's actually the story about the people I'm interested in …and then I was thinking about her sad life which is often the case.*". Here the participant connects the CC app experience with their prior knowledge and learning, discovering new information which triggers an **empathetic reflection**[2] about the persons in the painting.

The personalisation algorithms of the UC1 app provoked participants to **think about their art-related preferences**. For example, P13 notes: "*understanding in a way what kind of profile is this building up and what kind of pictures I like or I dislike. For example I am liking things that are more 20th century versus 16th century. As you are doing it, that makes me think*". Participants also **expressed their opinion and judgement** on how this process might change the way they experience the collection. The recommended groups of paintings also encouraged participants to think about the art collection from a different viewpoint and supported them in a process of **contemplation of their choices**. For example, some of them articulated clearly their **refamiliarisation** with the paintings of the collection, e.g. P12: "*I feel I know the paintings but again …it stopped me and made me think about the fabrics. He is painting fabrics in paintings extraordinary, and because I had fabrics in mind I was looking at those aspects of the painting too, I think more that I haven't before. These are paintings I walk past every day…*".

Dewey [25] thought of reflection as a form of problem solving that connected several ideas together by linking each idea with its predecessor in order to resolve an issue and test hypotheses. Originally, Dewey defines reflection as an "active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and further conclusions to which it leads" [25]. According to Dewey, reflection is a mandatory step towards effective learning, suggesting that, "We don't learn from experience. We learn from 'reflecting on experience'." He regarded reflection as a process of making meaning of experience in a systematic way, through community with others [10]. Thus, reflective thinking in this sense might be considered as both a

---

[2]Bloom [12] understands it to be 'the act of feeling what you believe other people feel/experiencing what they experience'(p 3), while cognitive empathy is further defined as 'if I understand that you are feeling pain without feeling it myself'(p 17).

process and an output: the result of a close, prolonged, precise examination of a given belief or hypothetical form of knowledge in light of the arguments supporting it and the conclusions to which it leads.

The findings from the UC1 experiment are consistent with the aforementioned view of learning [1]; this has considerable merit and utility for researchers investigating the development of knowledge and understanding which emerges from experiences in informal educational settings. In CH venues, we identified the importance of prior knowledge as a major factor in mediating the new experience [34], as participants associated new knowledge with what they already knew and had experience of [13]. The reflection traces we observed—and used as the experiment's operational definition—were most often in the form of a reaction (by commenting on feelings towards the learning experience), an elaboration (by comparing reactions with other experiences) or contemplation (focusing on constructive personal insights, attitudes, ethical matters, or moral concerns) [66].

5 COMPARING THE THINK ALOUD AND REMIND PROTOCOLS IN PRACTICE

Both the Think Aloud and the REMIND evaluation protocols require the participants to use the apps in a real-world setting and to describe their experiences. Think Aloud provides a commentary on the 'here and now' as it happens during the experience, whilst REMIND provides a commentary through a reliving of the experience as participants watch a video back and describe what they were thinking. There are positive and negative aspects to both protocols and deciding which one to use depends on many factors, discussed in this Section. The primary factors identified are: available resources (technical, staff, expertise, time, money), type of venue (indoor, outdoor, archaeological excavation site, museum, open and unstructured such as a city), size of venue (large with some noise, quiet gallery rooms), type of technology and readiness of the app, cultural background and personality traits of the participants (levels of introversion/extroversion), demographic characteristics of the participants and venue- or country-specific policies regarding artifacts' copyrights.

Think Aloud requires the commentary to happen in situ as the participant uses the application in the venue/city (UC4). It is only effective if venues permit talking and if participants are comfortable talking out loud. There is also a cultural dimension as some users in specific locations seem more reluctant to talk aloud. In UC4, it was necessary for the interviewer to walk alongside the participant, which we believe helped reduce this self-consciousness of talking to oneself. REMIND requires participants to talk about the video of their experience. While this activity requires a quiet space, laptop and specific software, during the visiting experience (UC1) the interviewer was only following the participants at a close range to observe whether the equipment (camera glasses) was behaving appropriately. If the technology is stable, then the visiting experience can be carried out without any supervision or intrusion, similar to a normal visit.

Both protocols are dependent upon participants who are talkative as they require minimal prompts from (or conversation with) the interviewer. The REMIND protocol requires that participants are

not given any details about the app they are testing. This is helpful in that it provides insights into the end-to-end experience of the participant, similar to how they would interact with the product after its launch. However, during the evaluation of the prototypes participants often focused on known usability issues. While leaving testers without any instructions avoids introducing biases in their behaviour and in the analysis of their behaviour, when participants test an application that is still a prototype, they face usability issues that act as barriers to a normal behaviour and focus on them. For this reason, we recommend the use of REMIND on finished products and not during the early development phases.

Compared to the Think Aloud participants, REMIND participants provided a more holistic commentary of their train of thought. However, they were more disconnected from the physical experience: for instance, participants were saying "I am reading" or "I read the comments". During Think Aloud experiments, on the other hand, when reading others' comments, participants sometimes offered their thoughts about the comments that they were reading (UC4), or how they went about answering a question (UC2 and UC3). This evidence contributed to an internal dialogue that formed part of the process of reflection. REMIND offered rich evidence of engagement but it was not always clear if this internal dialogue contributing towards reflection was occurring during the experience itself or as a post-experience while watching their video. When we were able to detect participants' reflection using the REMIND protocol, their observations about the paintings or thoughts triggered by narratives (UC1) must have been strong and powerful enough to stay in their short term memory (e.g. as highlights of the overall visiting experience).

The Think Aloud protocol applied in UC2, UC3 and UC4 seemed to provide reasonably good data. For example, due to museum IPR issues in UC3, cameras were not allowed to be used inside and therefore, audio recordings were considered the best available option in order to provide rich qualitative data. People watched the videos of the exhibits and told us their thoughts and feeling, immediately after. Since cameras were not used, but only a small recording device, users started talking freely, as if they forgot the recording was occurring. This allowed them to provide a lot of information about their cognitive and affective processes. The protocol seemed to be less intrusive than others that would require wearing cameras. In addition, users seemed to focus on the reflective aspect of the pilot and not so much on usability issues. However, it must be noted that the Think Aloud protocol is probably not appropriate for all cultural spaces, since it could disturb other visitors. It was used easily at the premises of the Tripolis museum, since the museum is peripheral and largely unknown with very few visitors, whereas in UC2 and UC1 the experiments took place at the end of the users' visits, in quiet parts of the venue.

We can draw the conclusion that REMIND is suitable for both ambulatory experiences (indoor and outdoor, for which it had never been tested before) but should be used on a final product rather than on an early technology prototype. It is a truly valuable tool that forms part of the co-design toolkit for understanding usability and user experience but comes at quite a significant resource cost in its analysis. If it is being used with prototype tools, we would suggest a more flexible methodology, where the participant is given some minimum instructions (and warnings regarding certain technical

issues) in order to focus their attention on the content and not on the technical aspects of the experience.

Based on the overall evaluation performed to measure traits of the users' reflection, we conclude that reflection is admittedly an elusive concept to define, and one that is particularly difficult to gather evidence of. The authors of [45] capture through a number of projects the different working definitions of the term 'reflective practice' in informal science learning settings and museums. In one of them, the REFLECTS project, reflective practice involved "museum staff reviewing their own video taped interactions, in order to collaborate more effectively with visitors from culturally and linguistically diverse populations." On the other side of this two-way relationship, do visitors reflect in museums? Do informal learning settings inspire lifelong motivation for visitors to learn? Is the intellectual journey of the visitor facilitated by the use of digital storytelling? Are the connections between the visitor's train of thoughts, stimulated by the CH apps, reinforced? The application of the REMIND protocol allowed us to shorten the temporal gap between the CH experience of the visitor and the verbalisation of this experience, and detect these connections, bringing us one step closer to answer the remaining questions.

Future work will focus on evaluating the gamified experience of the UC1 app in more depth using the REMIND protocol in terms of reflection. Reflection tasks in gamified CH experiences tend to favour analysis of questions and possible answers, evaluation of alternatives, considering clues available in the neighborhood and concepts learned previously during the game [11]. We aim to identify whether the UC1 gamified elements facilitate the participants during their CH experience to make use of their memory and critical reasoning cognitive skills.

6 CONCLUSION

This paper primarily compared how qualitative thinking 'out loud' methods were actually implemented and tested in the field. Based on our experience administering the evaluations and compiling the findings, we felt that REMIND could be an appropriate technique for probing the internal process of reflection that we aim to trigger with the CrossCult applications. The evaluation of the UC1 application in a large multi-thematic venue such as the National Gallery of London highlighted that it stimulated an intellectual journey for participants. During this journey, participants made connections between paintings, painters and contextual information related to the painting's creation. During the evaluations of the CrossCult apps the 'out loud' protocols showed evidence that when users are encouraged to participate and are immersed in the experience of discovery, in the absence of usability issues, they reflect.

ACKNOWLEDGMENTS

REFERENCES

[1] David Anderson, Lucas Keith, B., and Ginns Ian, S. 2003. Theoretical perspectives on learning in an informal setting. *Journal of research in science teaching* 40, 2 (2003), 177–199.

[2] Roberto Andreoli, Angela Corolla, Armando Faggiano, Delfina Malandrino, Donato Pirozzi, Mirta Ranaldi, Gianluca Santangelo, and Vittorio Scarano. 2017. A Framework to Design, Develop, and Evaluate Immersive and Collaborative Serious Games in Cultural Heritage. *ACM Journal on Computing and Cultural Heritage* 11, 1 (2017).

[3] Angeliki Antoniou. 2019. Predicting Cognitive Profiles from a Mini Quiz: A Facebook Game for Cultural Heritage. In *Proceedings of the Games and Learning Alliance conference.*

[4] Angeliki Antoniou, Jamie O'Brien, Tiphaine Bardon, Andrew Barnes, and Dane Virk. 2015. Micro-augmentations: situated calibration of a novel non-tactile, peripheral museum technology. In *Proceedings of the 19th Panhellenic Conference on Informatics.* 229–234.

[5] Olivier Aubert, Yannick Prié, and Daniel Schmitt. 2012. Advene, a versatile hypervideo authoring tool. In *Proceedings of the ACM Symposium on Document Engineering.* 79–82.

[6] Per Backlund and Maurice Hendrix. 2013. Educational games - Are they worth the effort? A literature survey of the effectiveness of serious games. In *Proceedings of the International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES).*

[7] Eva L Baker. 1994. Learning-based assessments of history understanding. *Educational Psychologist* 29, 2 (1994), 97–106.

[8] Stavroula Bampatzia, Ioannis Bourlakos, Angeliki Antoniou, Costas Vassilakis, George Lepouras, and Manolis Wallace. 2016. Serious games: valuable tools for cultural heritage.. In *Proceedings of the International Conference on Games and Learning Alliance.*

[9] Maddalena Bassani, Ahmed Dahroug, Abdullah Daif, Silvia González-Soutelo, Paola Zanovello, Susana Reboreda-Morillo, and Omar Gustavo Bravo-Quezada. 2018. Advanced Visual Interfaces to Represent Cultural and Historical Facts and Associations in the CrossCult EU project. In *Proceedings of AVI-CH 2018 Workshop on Advanced Visual Interfaces for Cultural Heritage co-located with 2018 International Conference on Advanced Visual Interfaces (AVI 2018).* Castiglione della Pescaia, Italy.

[10] Catherine Beauchamp. 2007. *Understanding reflection in teaching: a framework for analysing the literature.* Ph.D. Dissertation. Library and Archives Canada = Bibliothèque et Archives Canada. OCLC: 432399677.

[11] Francesco Bellotti, Riccardo Berta, Alessandro De Gloria, Annamaria D'Ursi, and Valentina Fiore. 2012. A serious game model for cultural heritage. *Journal on Computing and Cultural Heritage* 5, 17 (2012), 1–27.

[12] Paul Bloom. 2016. *Against empathy: the case for rational compassion.* Ecco. OCLC: 1125070017.

[13] David Boud, Ruth Cohen, and David Walker (Eds.). 1993. *Using experience for learning.* Society for Research into Higher Education & Open University Press, Buckingham.

[14] David Boud, Rosemary Keogh, and David Waker. 1996. *Promoting reaction in learning. A model. Reflection: Turning experience into learning.* Routledge in association with The Open University, London, 32–56.

[15] Elizabeth A Boyle, Thomas Hainey, Thomas M Connolly, Grant Gray, Jeffrey Earp, Michela Ott, Theodore Lim, Manuel Ninaus, Claudia Ribeiro, and João Pereira. 2016. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education* 94, C (2016), 178–192.

[16] Alejandro Calderón and Mercedes Ruiz. 2015. A systematic literature review on serious games evaluation: An application to software project management. *Computers & Education* 87 (2015), 396–422.

[17] Elizabeth Charters. 2003. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education: a Journal of Educational Research and Practice* 12, 2 (2003), 68–82.

[18] Tanguy Coenen, Lien Mostmans, and Kris Naessens. 2013. MuseUs: Case study of a pervasive cultural heritage serious games. *Journal on Computing and Cultural Heritage* 6, 2 (2013).

[19] Thomas M Connolly, Elizabeth A Boyle, Ewan Macarthur, Thomas Hainey, and James M Boyle. 2012. A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education* 59, 2 (2012), 661–686.

[20] Marijana Ćosović and Belma Ramić Brkić. 2019. Game-based learning in museums—cultural heritage applications. *Information* 11, 1 (2019).

[21] Ahmed Dahroug, Andreas Vlachidis, Antonios Liapis, Antonis Bikakis, Martin Lopez-Nores, Owen Sacco, and Jose Juan Pazos-Arias. 2019. Using dates as contextual information for personalized cultural heritage experiences. *SAGE Journal of Information Science* (2019). accepted.

[22] Abdullah Daif, Ahmed Dahroug, Martin López-Nores, Silvia González-Soutelo, Maddalena Bassani, Angeliki Antoniou, Alberto Gil-Solla, Manuel Ramos-Cabrer, and Jose Juan Pazos-Arias. 2019. A mobile app to learn about cultural and historical associations in a closed loop with humanities experts. *Applied Sciences* 9, 1, Article 9 (2019), 14 pages.

[23] Edward De Bono. 2010. *Lateral thinking: Creativity step by step.* Harper Collins, New York.

[24] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining "Gamification". In *Proceedings of the International Academic MindTrek Conference.*

[25] John Dewey. 1933. *How we think : a restatement of the relation of reflective thinking to the educative process.* (2nd ed.). D.C.Heath.

[26] Judy Diamond, Jessica J. Luke, and David H. Uttal (Eds.). 2009. *Practical evaluation guide tool for museums and other informal educational settings* (2nd ed. ed.). AltaMira Press.

[27] Hooper-Greenhill Eilean (Ed.). 1999. *The educational role of the museum* (2nd ed.). Routledge, London ; New York.

[28] Christos Emmanouilidis, Remous-Aris Koutsiamanis, and Aimilia Tasidou. 2013. Mobile guides: Taxonomy of architectures, context awareness, technologies and applications. *Journal of Network and Computer Applications* 36 (2013), 103–125.

[29] Karl Anders Ericsson and Herbert A. Simon. 1993. *Protocol analysis : verbal reports as data* (revised ed.). MIT Press.

[30] Alessandro Foni, George Papagiannakis, and Nadia Thalmann. 2010. A taxonomy of visualization strategies for cultural heritage applications. *Journal on Computing and Cultural Heritage* 3, 1 (2010).

[31] Josef Froschauer, Dieter Merkl, Max Arends, and Doron Goldfarb. 2013. Art History Concepts at Play with ThIATRO. *Journal on Computing and Cultural Heritage* 6, 2 (2013).

[32] Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research: Analysing Social Interaction in Everyday Life.* Vol. 23. SAGE Publications, Inc.

[33] George Hein. 1994. *Evaluation of museum programmes and exhibits.* Routledge, 306–312.

[34] George Hein. 1998. *Learning in the museum.* Routledge.

[35] Catherine Emma Jones, Angeliki Antoniou, Martin López-Nores, Kalliopi Kontiza, Celine Schall, Yannick Naudet, and Louis Deladiennée. 2018. Evaluation and Best Practices Report. (2018). http://www.crosscult.eu/en/resources/deliverables/ CrossCult Deliverable 5.2, First version, April 2018.

[36] Catherine Emma Jones, Angeliki Antoniou, Martin López-Nores, Kalliopi Kontiza, Celine Schall, Yannick Naudet, and Louis Deladiennée. 2019. Evaluation and Best Practices Report. (2019). http://www.crosscult.eu/en/resources/deliverables/ CrossCult Deliverable 5.3, Final version, January 2019.

[37] Catherine Emma Jones, Antonios Liapis, Ioanna Lykourentzou, and Daniele Guido. 2017. Board game prototyping to co-design a better location-based digital game. In *Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems.* 1055–1064.

[38] Catherine Emma Jones and Konstantinos Papangelis. 2020. Reflective practice: lessons learnt by using board games as a design tool for location-based games. In *Geospatial Technologies for Local and Regional Development*, Phaedon Kyriakidis, Diofantos Hadjimitsis, Dimitrios Skarlatos, and Ali Mansourian (Eds.). 291–307.

[39] Catherine Emma Jones, Stathis Theodosis, and Ioanna Lykourentzou. 2019. The Enthusiast, the Interested, the Sceptic, and the Cynic: Understanding User Experience and Perceived Value in Location-Based Cultural Heritage Games Through Qualitative and Sentiment Analysis. *Journal on Computing and Cultural Heritage* 12, 1, Article 6 (2019), 26 pages.

[40] Jane Peirson Jones (Ed.). 1993. *Gallery 33 : a visitor study.* Birmingham Museums and Art Gallery.

[41] Yasmin Kafai and Mitchel Resnick (Eds.). 1996. *Constructionism in practice : designing, thinking and learning in a digital world.* Lawrence Erlbaum.

[42] Kristian Kiili, Harri Ketamo, and Timo Lainema. 2007. Reflective thinking in games: triggers and constraints. In *Proceedings of the European Conference on Games Based Learning.* 169–176.

[43] Kalliopi Kontiza, Antonios Liapis, and Joseph Padfield. 2018. Capturing the Virtual Movement of Paintings: A Game and A Tool. In *Proceedings of the Digital Heritage International Congress (DigitalHeritage).*

[44] Kalliopi Kontiza, Olga Loboda, Louis Deladiennee, Sylvain Castagnos, and Yannick Naudet. 2018. A Museum App to Trigger Users' Reflection. In *Proceedings of the Mobile HCI Workshop on Mobile Access to Cultural Heritage.*

[45] Lynn Uyen Tran Laura W. Martin and Doris Ash (Eds.). 2019. *The reflective museum practitioner: expanding practice in science museums.* Routledge.

[46] Antonios Liapis, Georgios N. Yannakakis, Constantine Alexopoulos, and Phil Lopes. 2016. Can Computers Foster Human Users' Creativity? Theory and Praxis of Mixed-Initiative Co-Creativity. *Digital Culture & Education (DCE)* 8, 2 (2016), 136–152.

[47] Martin López-Nores, Omar Gustavo Bravo-Quezada, Maddalena Bassani, Angeliki Antoniou, Ioanna Lykourentzou, Catherine Emma Jones, Kalliopi Kontiza, Silvia Gonzalez-Soutelo, Susana Reboreda-Morillo, Yannick Naudet, Andreas Vlachidis, Antonis Bikakis, and Jose Juan Pazos-Arias. 2019. Technology-powered strategies to rethink the pedagogy of history and cultural heritage through symmetries and narratives. *Symmetry* 11, 3 (2019).

[48] Sharon Macdonald. 2007. Studying Visitors. In *A Companion to Museum Studies.* Blackwell Publishing Ltd, 362–376.

[49] Humberto R Maturana, Francisco J. Varela, and François-Charles Jullien. 1994. *L'arbre de la connaissance.* Ed. Addison-Wesley France. http://www.sudoc.fr/ 003273784

[50] Matthew B. Miles, Michael Huberman, and Johnny Saldana. 2020. *Qualitative data analysis: a methods sourcebook* (fourth ed.). Sage, Los Angeles.

[51] Roger S. Miles and Lauro Zavala. 1994. *Towards the Museum of the Future: New European Perspectives.* Routledge.

[52] John Mingers. 1991. The cognitive theories of Maturana and Varela. *Systems Practice* 4, 4 (1991), 319–338.

[53] Roger Mucchielli. 2006. *L'analyse de contenu: Des documents et des communications.* ESF, Paris.

[54] L. C. Nielsen. 1946. A technique for studying the behavior of museum visitors. *Journal of Educational Psychology* 37, 2 (1946), 103–110.

[55] Ioannis Paliokas and Stella Sylaiou. 2016. The Use of Serious Games in Museum Visits and Exhibitions: A Systematic Mapping Study. In *Proceedings of the International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES).*

[56] Catherine Palmer. 2001. Ethnography: a research method in practice. *International Journal of Tourism Research* 3, 4 (2001), 301–312.

[57] Michael Pickering. 2008. *Research methods for cultural studies.* Edinburgh University Press, Edinburgh.

[58] Judy M. Rankin. 1988. Designing Thinking Aloud Studies in ESL Reading. *Reading in a Foreign Language* 4, 2 (1988), 119–132.

[59] Leonie Rennie and David Johnston. 2004. The nature of learning and its implications for research on learning from museums. *Science Education* 88 (2004), S4–S16. http://search.proquest.com/docview/194944920/

[60] Daniel Schmitt. 2012. *Expérience de visite et construction des connaissances : le cas des musées de sciences et des centres de culture scientifique.* Ph.D. Dissertation. Université de Strasbourg.

[61] Daniel Schmitt. 2016. Describing and understanding the experience of visitors. In *Visiting The Visitors. An Enquiry Into the Visitor Business in Museums*, Ann Davis and Kerstin Smeds (Eds.). transcript Verlag, Bielefeld, 55–69.

[62] Daniel Schmitt. 2018. Enaction, a fertile epistemological framework for Information and Communication Sciences research. *Les Cahiers du numérique* 15, 2018/2 (2018), 93–112.

[63] Daniel Schmitt and Olivier Aubert. 2017. REMIND : a method to understand the micro-dynamics of the museum visitors experience. *Revue des Interactions Humaines Médiatisées (RIHM)* 17, 2 (2017), 43–70.

[64] J. Mark David Schuster. 1991. *The audience for American art museums.* Seven Locks Press, Washington.

[65] Barbara J. Soren. 2009. Museum experiences that change visitors. *Museum Management and Curatorship* 24, 3 (2009), 233–251. https://doi.org/10.1080/ 09647770903073060

[66] Elaine Surbeck, Eunhye Han, and Joan E. Moyer. 1991. Assessing Reflective Responses in Journals. *Educational Leadership* 48, 6 (1991), 25–27.

[67] Sigmund Tobias. 1994. Interest, prior knowledge, and learning. *SAGE Review of Educational Research* 64, 1 (1994), 37–54.

[68] Francisco J. Varela, Paul Bourgine, and Paul Dumouchel. 1989. *Autonomie et connaissance essai sur le vivant.* Éditions du Seuil, Paris. http://www.sudoc.fr/ 001396218

[69] Francisco J. Varela and Pierre Lavoie. 1989. *Invitation aux sciences cognitives.* POINTS.

[70] Georg Volkmar, Nina Wenig, and Rainer Malaka. 2018. Memorial Quest - A Location-Based Serious Game for Cultural Heritage Preservation. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts.* 661–668.

[71] Sarah Stahlke Wall. 2014. Focused ethnography: a methodological adaptation for social research in emerging contexts. *Forum Qualitative Sozialforschung* 16, 1 (2014).

[72] Anna Wertenbroch and Thierry Nabeth. 2000. Advanced learning approaches & technologies: The CALT Perspective. *The Center for Advanced Learning Technologies, Web Site, retrieved on February* 27 (2000).