# Do VR and AR versions of an immersive cultural experience engender different user experiences? ☆

Isabelle Verhulst [*], Andy Woods, Laryssa Whittaker, James Bennett, Polly Dalton

*Royal Holloway University of London (StoryFutures), United Kingdom*

## ARTICLE INFO

## ABSTRACT

Although Virtual Reality (VR) and Augmented Reality (AR) user experiences have received large amounts of recent research interest, a direct comparison of different immersive technologies' user experiences has not often been conducted. This study compared user experiences of one VR and two AR versions of an immersive gallery experience 'Virtual Veronese', measuring multiple aspects of user experience, including enjoyment, presence, cognitive, emotional and behavioural engagement, using a between-subjects design, at the National Gallery in London, UK. Analysis of the self-reported survey data (N = 368) showed that enjoyment was high on all devices, with the Oculus Quest (VR) receiving higher mean scores than both AR devices, Magic Leap and Mira Prism. In relation to presence, the elements 'spatial presence', 'involvement', and 'sense of being there' received a higher mean score on the Oculus Quest than on both AR devices, and on 'realism' the Oculus Quest scored significantly higher than the Magic Leap. Cognitive engagement was similar between the three devices, with only 'I knew what to do' being rated higher for Quest than Mira Prism. Emotional engagement was similar between the devices. Behavioural engagement was high on all devices, with only 'I would like to see more experiences like this' being higher for Oculus Quest than Mira Prism. Negative effects including nausea were rarely reported. Differences in user experiences were likely partly driven by differences in immersion levels between the devices.

## 1. Immersive storytelling in cultural institutions using immersive technologies

Storytelling is important to all ages and cultures, and can use different forms, ranging from oral storytelling, drawings, the written word, theatre and TV, to cutting edge, immersive technologies such as virtual reality (VR) and augmented reality (AR) (Gröppel-Wegener & Kidd, 2019).

Many different definitions of AR and VR exist, with a notable lack of consistency (Kardong-Edgren et al., 2019). Those of Milgram and Kishino (1994) have been adopted for this research. They defined AR as "augmenting natural feedback to the operator with simulated cues", whereas "a VR environment is one in which the participant-observer is totally immersed in a completely synthetic world" (Milgram et al., 1995, p. 283). In other words, in AR elements of the 'real' environment are combined with computer generated images, whereas in VR everything the participant sees is computer generated. The Reality-Virtuality (RV) Continuum shown in Fig. 1, can be used to visualise the difference

between AR and VR. As yet, not only is nomenclature unclear in this developing medium, but also the extent to which AR and VR versions of an immersive experience can lead to different user experiences; the focus of this study.

Immersive storytelling uses virtually generated content of places, peoples and objects that can be richly informative, emotive and memorable, appealing to a variety of audiences (Azuma, 2015). Immersive stories are being told for different purposes and in different settings, including for example in manufacturing, medical training, gaming, psychological treatment and cultural institutions like museums and galleries.

The use of AR and VR at cultural institutions has been argued to have economic, experiential, social, epistemic, cultural, historical, and educational value. For example, tom Dieck and Jung (2017) asked 24 stakeholders of a UK museum about their perception of the value that AR brings to cultural heritage sites, which included preserving history, telling engaging stories of past events, creating word-of-mouth recommendations, delivering a positive learning experience and stimulating
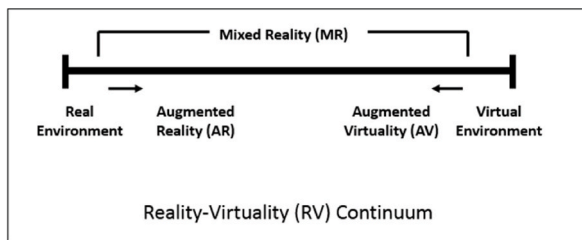
**Fig. 1.** Reality-Virtuality (RV) Continuum. *Note.* Source: Milgram et al. (1995), recreated with permission from the first author.

both existing and new visitors' satisfaction. Given this range of potential benefits, it is not surprising that museums and galleries have increasingly been using immersive technologies to tell the stories of their works. For example, The Kyoto National Museum (Japan) in 2018 used a holographic monk to tell the 400-year old story of the Folding Screen of Fujin and Raijin to visitors wearing a Microsoft's HoloLens (Sylaiou et al., 2018). In 2017; visitors of the Art Gallery of Ontario, Canada, could use their phones to see the subjects of a painting come alive (AGO, 2017), and the Dali Museum in Florida (US) used VR to let people explore the painting 'Archaeological Reminiscence of Millet's Angelus' in 'Dreams of Dali' (Graham, 2016).

The increasingly widespread use of VR and AR for creative storytelling in cultural and other settings raises the question of how these different immersive technologies compare in delivering a story experience. The current study addresses this question by directly comparing user experience across VR and AR versions of a similar immersive experience in the National Gallery in London, UK.

In a recent review of the state of immersive technology based on 54 articles on AR or VR use in education, marketing, business, and healthcare, Suh and Prophet (2018) report three significant limitations in current knowledge about immersive technology use: (1) a lack of comparative work; (2) a lack of 'real world' studies; (3) a lack of analysis of any technology drawbacks.

Suh and Prophet (2018) first claim, that: "… very little research has examined the different effects of diverse technological stimuli on multiple aspects of user performance" (p. 87), highlights the fact that most previous studies in this area have investigated the user experience of only *one* type of immersive technology. For example, in two VR studies with close to 1000 participants, Tussyadiah et al. (2018) found that an increased feeling of being there (presence) increased enjoyment of VR experiences. He et al. (2018) investigated the role of AR in enhancing museum experiences and purchase intention. They found that in AR, dynamic verbal cues can increase willingness to pay, especially when environmental augmentation creates a strong feeling of presence. These and other findings suggest that presence is an important factor in user experiences, for both VR and AR. However, it is presently unknown whether user experience, including the level of presence, is likely to differ between VR and AR versions of the same activity.

Studies that have used both AR and VR have not typically *compared* the two technologies directly. For example, Sylaiou et al. (2010) created a virtual museum with VR and AR elements, based on a gallery in the Victoria and Albert Museum, London, UK. Users were encouraged to manipulate 3D models of cultural artifacts in VR and then in AR. VR and AR presence were not compared, although they found a positive correlation between enjoyment and both VR and AR object presence. Another example is provided by See et al. (2017), who investigated user experience of a four-sided AR and VR showcase pillar about boat builders of Pangkor. The AR side of the pillar offered AR-based videos, maps, images and text, whereas the VR side showed reconstructed 3D-subjects and locations in a 360° virtual environment. Based on user ratings, the authors reported that their 360° VR was experienced as more natural to use than AR's static image and text, and more useful than all other elements. However, the AR and VR elements displayed were too different in content and style to be compared. Similarly, the effects of VR and AR

on visitor experiences were investigated in a mixed reality (VR and AR) mining museum by Jung et al. (2016). Their results showed that social presence in mixed environments with VR and AR elements was a strong predictor of four experience realms ('education', 'aesthetics', 'entertainment' and 'escape'). In all of these, except 'aesthetics', presence was shown to predict the overall tour experience rating and intention to revisit. Even though the user experience was measured consistently and separately for AR and VR, the AR and VR experiences were very different, with AR offering text, images and audio explaining the museum, while the VR application allowed participants to experience a lift ride down the mining shaft, precluding direct comparison.

Only a few comparative studies have used multiple devices with the same experience. Some of these studies found differences; others did not. Examples of the former include Dow et al. (2007), who reported findings from a qualitative study of three different versions of an emotional drama (desktop 3D using typed text, desktop 3D using speech, and an interactive AR head mounted device (HMD) version in which participants could interact with virtual characters in an apartment representation). Presence was higher when using the AR interface than when using either of the desktop interaction approaches. In this study higher presence did not increase engagement, which according to the authors may have been because participants preferred the safe distance from the emotionally charged drama that the less immersive desktop versions offered. Another comparison study identifying device differences was reported by Voit et al. (2019). The authors measured immersion levels, using the Augmented Reality Immersion (ARI) questionnaire (Georgiou & Kyza, 2017) to understand whether immersion varied across different technologies/situations. These included an online survey with a video displaying the relevant interaction, for example watering a plant (called 'Online'), a lab study in which the VR controller was used to grab a simulated watering can in VR (labelled 'VR'), a lab study using a regular watering can in AR ('AR'), a lab study with the actual watering can ('Lab'), and an in-situ study in participants' homes with their own watering can ('In-Situ'). 60 participants tested prototypes of four smart objects (a cup, mill, plant, and speaker). The authors found that 'In Situ' was given the highest immersion score, followed by 'VR', then 'Lab', then 'AR' and then 'online'. 'AR' was rated as producing significantly lower immersion than 'VR'.

In contrast, there are comparative studies not finding significant device differences. These include Aslan et al. (2019), who assessed the effect of three AR devices (HoloLens, iPad Pro tablet, and iPhone X smartphone) on presence while viewing an AR treehouse. They measured presence using the Presence Questionnaire (PQ) (Witmer & Singer, 1998). Presence was found to be similar between these devices, with the only difference on 'possibility to examine an AR object', which was rated highest for the tablet, then the smartphone, and lowest for the HoloLens. According to participants' qualitative feedback this could be because the smartphone and tablet allowed the additional possibility to move the device without having to move themselves, and the tablet's larger screen. A study by Loizides et al. (2014) investigated user experience of two types of VR virtual museums (head mounted displays (HMD's) and large screen stereoscopic projections) and found that participants gave both types the same rating for overall experience. It is possible that the studies not finding significant user experience differences had fairly similar device immersion levels, or that the effects of individual pros and cons of different devices netted out in similar overall scores.

Suh and Prophet's (2018) second criticism was that most existing studies used lab experiments, with students as participants. This is in line with findings from Dey et al. (2018) who concluded from their systematic review of ten years of AR usability studies that participant populations are generally dominated by young, male participants and that between-subject designs are scarce. Using real life user settings and non-student {subjects might increase study validity and engender a fuller understanding of immersive user experiences (Kidd & Nieto, 2019; Kim et al., 2018a).

To summarize, this study aims to help fill three gaps in the existing literature. The first, the lack of comparative work, is addressed by directly comparing the user experience of three different versions (one VR, two AR). The second, the lack of real world studies, is addressed by using regular visitors to the National Gallery in London, UK, as participants. And the third, the lack of analysis of the drawbacks of immersive technologies, is addressed by including negative side effects (for example, nausea and feeling uncomfortable) in the outcome measures.

The research question is: Do VR and AR versions of an immersive cultural experience engender different user perceptions of presence, engagement, enjoyment, and do they have negative side-effects?

### 1.1. Measuring immersive user experience

Different concepts have been used to measure the user's experience in immersive research projects, including presence (Cummings & Bailenson, 2016; Hein et al., 2018), engagement (Boyle et al., 2012) and enjoyment (Dey et al., 2018).

Immersion is defined as the technical aspects of a device, its 'physics' or "how well it can afford people real-world sensorimotor contingencies for perception and action" (Slater & Sanchez-Vives, 2016, p. 37). It is the extent a device is able to provide our human senses with computer-generated input that allows the brain to create a full, believable picture of its surroundings through a perceptual fill-in mechanism. Examples of important aspects of immersion include the resolution of the display, wide field of view vision, number of Degrees of Freedom (DOF), low latency from head tracked movement to display (for vision) and quality (stereo/surround) sound. Input on the senses touch, force, smell and taste are less often used due to current technical limitations. Different devices, therefore, have different levels of immersion. Thus "system A is more immersive than system B if A can be used to simulate the perception afforded by B but not *vice versa*" (Slater & Sanchez-Vives, 2016, p. 5).

Presence relates to the human response to immersion and is defined as "the propensity of people to respond to virtually generated sensory data as if they were real" (Slater et al., 2009, p. 194). Factor analysis showed that presence has three components: "(1) the relation between the virtual environment (VE) as a space and the own body (spatial presence), (2) the awareness devoted to the VE (involvement) and (3) the sense of reality attributed to the VE (realness)" (Schubert et al., 1999).

Engagement is a concept shared in many domains (Bouvier et al., 2014), although no generally agreed definition exists in the immersive literature. The definition suggested by Attfield et al. (2011) created in a web application context, was adopted for this study as it covers multiple relevant psychological aspects; "the emotional, cognitive and behavioural connection … between a user and a resource."

## 2. Methods

### 2.1. Apparatus

The immersive experience tested, Virtual Veronese, told the story of Veronese's painting 'The consecration of St. Nicolas' as it would have been seen in its original location, a chapel in Italy in 1562. The experience was made available in AR and VR. The AR version of the experience was presented on two headsets; Magic Leap One Creator Edition (Magic Leap, Inc., 2019) and Mira Prism (Mira Labs, Inc., 2017) with an inserted iPhone 8 (Apple, 2017). Whilst Magic Leap places a transparent display between eyes and reality to display content on and thus augment reality, the Mira Prism uses a semi-transparent mirror to partly reflect content from the iPhone screen. The VR version was presented on Oculus Quest (Facebook, 2019). The experience was built in Unity version 2019.2.2. Stereo headphones (Monoprice brand) were used by all participants for audio delivery. Fig. 2 shows the three devices, one of which is shown worn in Fig. 3.



**Fig. 3.** *Participant with Apparatus in Situ. Note.* Participant wearing a Magic Leap and headphones in experience.



**Fig. 2.** *Apparatus Images. Note.* Figure shows the three devices used in the study.

## 2.2. Experience consistency

The experience was kept as similar as possible between the devices, however building the experience on different devices (platforms) did create certain aesthetic and user interface changes. Similarities and differences are presented below, as these may have impacted user experience.

### 2.2.1. Experience similarities

The images, audio and story of the painting were based on Dr Rebecca Gill's (Ahmanson Research Fellow at the National Gallery) research about Paolo Veronese's painting *The Consecration of St Nicholas* and the creative treatment of this research by the StoryFutures team. The experience replaced (in VR) or blended (in AR) the environment of the National Gallery in London, with that of the Church of San Benedetto al Po in 1562, where the painting originally hung.

The core content for the experience was identical across the headsets, comprising of a 3D scan of the chapel rendered into the headsets using the Unity game engine, stereo video of the story actors, a 3D audio sound track (choral music), voice over and text based language subtitles. The actor/avatar videos, music audio, and subtitles were identical across all platforms.

The same video was used for all the headsets. The video capture was performed in a green screen studio. The camera system was a pair of Blackmagic Microstudio 4 K cameras setup side by side and synchronized. The camera separation mimicked an average human interpupillary distance (IPD). The captured videos were composited side by side and post produced to remove shadows and create a single colour green background. In the application the video is decoded and the left hand side played on a billboard to the left eye and right side to the right eye, using the green as a transparent chroma key.

All participants could choose between two narrative versions: one story-led, narrated by actors portraying the Abbot and a monk living in the church of San Benedetto al Po in 1562 (henceforth called 'Abbot'); and one factual, narrated by the present day curator Dr Gill (henceforth called 'Curator'). Participants selected their preferred narrative version in the headset by looking at the narrator of their choice for a few seconds, which activated the selected content (see Fig. 4). For the Mira Prism the visitor assistant (VA) made the selection for the user. In a similar way they could select (or turn off) subtitles, which were available in six languages. The narrative versions and subtitles were the same in AR and VR.

### 2.2.2. Experience differences

The Magic Leap and Quest versions included higher resolution textures inside the chapel. The Mira is a three degrees of freedom (3DOF) headset and the Quest and Magic Leap offer six degrees of freedom (6DOF). The higher DOF means that the user had some level of ability to move around in the chapel in the Quest and Magic Leap versions of the experience. However, most participants only moved their heads as they were asked to stand still to prevent touching the actual painting or other



**Fig. 4.** *Narrator Selection. Note.* Image from narrator selection in experience.

participants. The Quest version showed the adjoining room from the chapel and church as it was in 1562 through the doorways, whereas Mira Prism and Magic Leap users saw the present day National Gallery through the doorways. The Quest included a computer generated painting, whereas both AR headsets had a cut-out through which the real painting was seen by users.

Image tracking was used to position real world objects by the Mira Prism, whereas depth cameras and accelerometers created persistence tracking used to position real world objects by the Magic Leap. Image tracking and persistence tracking are tools used to spatialise digital content overlaid onto real world environments. Image tracking 'attaches' digital augmentations to real life images or targets and matches the augmentations to the movement of the target. Persistence tracking uses collected information about the surrounding environment and the perceived movement of the device to place a digital augmentation and keep this placement consistent despite the movement of the viewer (Wikitude, 2021a, Wikitude, 2021b; Magic Leap Developer, 2020).

The voiceover prompts for the user were customised to each headset. The on-boarding (e.g. how to select a narrator and subtitles) was different for each headset: When participants were going to be using the Mira Prism, the VA asked the user which narrator they wanted and whether they wanted subtitles, the selection was entered by the VA on the phone and the headset with phone in it handed to the viewer. The viewer was asked to stand in front of the painting and look at the top of the picture - the application then located the picture and started the experience. For the Magic Leap, the location of the picture was set up once by the VA staff at the start of each session (or day). The VA staff would start the application and hand the headset to the viewer, who would choose the narrator and whether they wanted subtitles by pointing the headset at images overlaid on the chapel. For the Oculus Quest, the VA staff set up a guardian area on the headset at the start of each day (more often if required). The viewer put the headset on outside the guardian area - at which point they could see the gallery in black and white. They were then asked to step into the guardian area, which put them in the VR chapel, where the viewer would then choose the narrator and whether they wanted subtitles, by pointing the headset at images overlaid on the chapel.

## 2.3. Design

Between July 23 and July 29, 2019 (week 1) all National Gallery visitors had the opportunity to participate in the AR experience. Between July 30 and August 5, 2019 (week 2) visitors could participate in the VR experience. Participants could not select VR or AR, as which device they were presented with depended on which week they visited the National Gallery, to prevent self-selection bias. The AR and VR versions were presented in the same room of the National Gallery London. The experience took place in a cordoned off section of approximately 8 × 8 metres, in front of the painting The Consecration of Saint Nicholas, as shown in Fig. 5. A large bench was made available for people to re-acclimatise or take the post-experience questionnaire.

During the trial, two participants took the experience in parallel, standing next to each other, each assisted by one VA. A dedicated team of individuals from the National Gallery coordinated the experience, and were responsible for on-boarding (which included explaining the experience, how to select a narrator and subtitles, how to put the headset and headphones on, adjust the volume and answer any questions); providing assistance during the experience if needed; and off-boarding (which included helping people take the headset and headphones off and, if need be, to sit down to re-acclimatise, and inviting people to fill out the research questionnaire). An additional National Gallery host answered questions from people while they were queuing. During off-boarding, researchers asked participants if they would like to fill in an anonymous web-based survey on an iPad. The survey was made available in twelve major languages based on the expected National Gallery visitor profile. There was no financial incentive for taking part in the
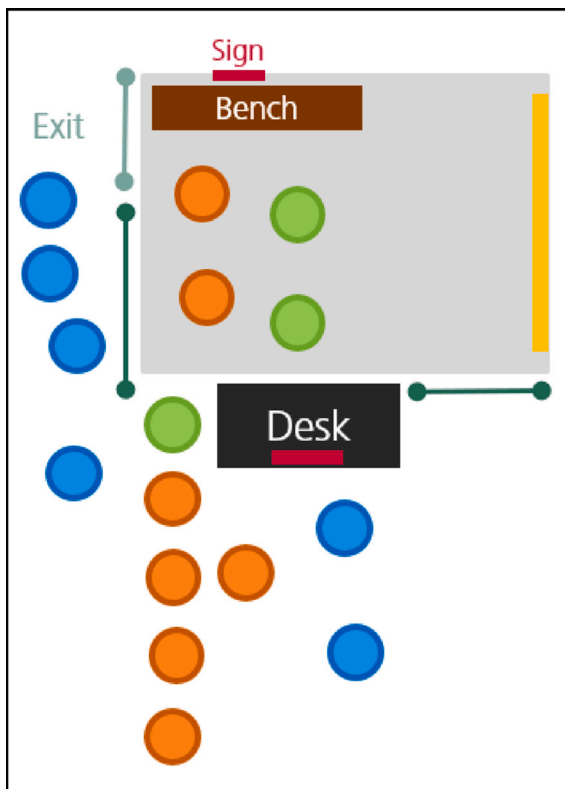
**Fig. 5.** *Experience Layout. Note.* A schematic of the layout of the experience. Participants are marked as orange dots (inside the greyed out experience area), visitor assistants (VA) as green dots, people queuing before participation as orange dots (outside the greyed out experience area), other visitors as blue dots and the painting as a yellow rectangle. Source: National Gallery.

experience nor questionnaire. Queuing took between zero and 30 min, on-boarding approximately 2 min, the AR or VR experience 6 min, off-boarding 1 min and the questionnaire 6 min. The work described has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

### 2.4. Instrument

Self-report data was captured in an online questionnaire on an iPad, immediately after the experience. The questionnaire items used to measure key concepts in this study were informed by a literature review and a recent immersive industry toolkit (Freeman et al., 2019), which was validated with industry stakeholders and audiences as described in the report "Evaluating Immersive User Experience and Audience Impact" (Nesta & i2 Media Research, 2018).

Enjoyment was measured by asking participants to rate the overall experience they had on a five point Likert scale (labelled with one to five stars), using the question 'Overall, how much did you enjoy the experience?', in line with multiple existing studies (Weibel & Wissmath, 2011). This will be referred to as 'enjoyment'.

Presence has been measured in multiple ways, most often using questionnaires (Grassini & Laumann, 2020; Hein et al., 2018). The Igroup Presence Questionnaire (IPQ, Schubert et al., 2001) consists of fourteen questions, defined by factor analysis of four previous questionnaires. For brevity, and with advice from the main author of the IPQ (personal correspondence, April 26, 2019), one question per subscale was selected: spatial presence ('I felt present in the virtual space'); involvement ('I was completely captivated by the virtual world'); realism ('how real did the virtual world seem to you?'); plus the general item 'sense of being there' ('in the computer generated world I had a sense of

"being there"'). Participants' responses to these questions were measured using a five point Likert scale. For brevity, these items will be referred to as 'spatial presence', 'involvement', 'realism' and 'being there'. Overall presence was calculated as the average of the four presence questions.

Engagement was separated into its cognitive, emotional and behavioural elements, for measurement.

Cognitive engagement questions were based on relevant narrative questions from the transportation scale–short form (TS–SF) (Appel et al., 2015). The word 'narrative' was replaced with 'story' for user friendliness. Questions were: 'I was mentally involved in the story while experiencing it' and 'I wanted to learn how the story ended', measured on Appel et al. (2015)'s four point Likert scale. The items will be referred to as 'mentally involved' and 'how story ended', with a cognitive engagement score calculated as their unweighted average.

Emotional engagement was measured by asking for participants' agreement with these statements, measured on a four-point Likert scale: 'The story affected me emotionally' (referred to as 'affected emotionally'), 'I felt happy/relaxed/kama muta (Sanskrit for 'moved by love')' (the average of happy/relaxed/kama muta is referred to as 'positive emotions'), 'I felt angry/sad/fearful' (the average of which is referred to as 'negative emotions'). These were based on a condensed version of the 'discrete emotion questionnaire' (Harmon-Jones et al., 2016), adapted in the following ways; including a four-point scale rather than seven, to keep the scale in line with other scales employed; using one item per specific emotion rather than four, for brevity; and using a subset of emotions highly relevant to the study. The National Gallery was interested in understanding the effect of the experience on a specific emotion called kama muta, operationalised as the average of three agreement scores: 'the experience was heart-warming', 'the experience moved you' and 'you were touched by the experience' (Fiske et al., 2019). An unweighted average of the seven emotional items was calculated for overall emotional engagement.

Behavioural engagement was measured by asking for agreement with different kinds of behavioural intention statements: 'I would like to repeat this experience', 'I would like to see more experiences like this one', 'the Gallery should have more interactive experiences', 'I will tell my friends about this experience' and 'I plan to look for more information about this painting in the future'. These will be referred to as 'repeat this experience', 'more like this experience', 'Gallery interactive experiences', 'tell friends' and 'look up more information'. Repeatability and 'tell friends' questions are common in marketing and user experience testing, and 'look up more information' is in line with VR studies from Slater et al. (2018) and Steed et al. (2018), who also investigated intention to follow up in VR studies. An unweighted average of the five behavioural items was calculated for overall behavioural engagement. In addition to these behavioural intent items, an additional item was measured: 'I knew what I was supposed to do the whole time'. Because this combines both cognitive ('knew') and behavioural ('do') elements, it is not included in the cognitive or behavioural engagement averages, but is presented separately. All these items were measured with a five point Likert scale.

Negative effects on users were measured by asking to which extent people felt nauseous, disconnected from their body and uncomfortable, each on a five point scale. These items were selected from questions of the three subscales (nausea, discomfort and disorientation) of the 16-item virtual reality sickness questionnaire (VRSQ; Kim et al., 2018b).

### 2.5. Sample

Approximately 400 participants experienced Virtual Veronese, 368 of whom completed the StoryFutures Audience Insight survey (approximately 90 % of participants). Participants came from over 40 countries, with most coming from Italy (18 %), UK (15 %), USA (13 %) and China (6 %). There were more female participants (56 %) than male participants (44 %). Virtual Veronese attracted a wide range of different ages

(13–77), with a 50-50 split between those under and those 35 and older (by chance). The minimal age limit of 13+ was imposed on recommendation from the headset manufacturers. Most participants had A-level education (a UK school-leavers' qualification for normally 16–18 year olds), or an international equivalent. Participants were fairly new to immersive technologies, with 79 % stating they had never or only once or twice used a VR/AR headset before. Of all participants, 33 % used Magic Leap (these were participants in week 1 without glasses), 17 % Mira Prism (participants in week 1 with glasses) and 50 % Oculus Quest (all participants in week 2), therefore the overall AR-VR split between visitors was 50-50.

Since individual differences can shape the effects of technological and content stimuli on users' cognitive and affective reactions (Suh & Prophet, 2018) differences in user characteristics between participants using different devices were investigated. As shown in Table 1, significance testing (using a chi-square for categorical variables 'gender', 'previous AR/VR head mounted display (HMD) usage', 'education' and 'previous National Gallery visits', and using an independent Kruskal-Wallis test for the continuous variable 'age', as recommended by Field, 2013) showed that most user characteristics were not significantly different between the devices ($p > .05$).

Only the participants' countries of origin differed between devices, with participants from US, Italy and Spain being slightly over-represented in the VR group (on average 6 more participants per country in the VR group than expected) and participants from the large group of 'other' countries being underrepresented in the VR group (23 fewer than expected). Offering subtitles in 6 languages in the immersive experience, and offering the survey in 12 languages as mentioned earlier, helped to minimise a language effect on the participants' understanding of the experience and/or survey. It is therefore assumed unlikely that the difference in countries of origin would have influenced the findings to the extent that the outcomes cannot be compared between devices.

### 2.6. Data assumptions

Kolmogorov-Smirnov and Shapiro-Wilk normality tests suggested data normality issues, as shown by $p < .05$ on multiple variables (Field, 2013). Visual checks using histograms showed that this was due to most participants giving highly positive ratings on all variables. Levene's statistic was not significant, so homogeneity of variance between devices can be assumed. However, because of the non-normality of the data, non-parametric tests were used.

**Table 1**
Individual differences between devices significance test.

| Variable | Significance Test |
| --- | --- |
| Age | $H(2) = 2.563, p = .278$[a] |
| Gender | $\chi^2(2) = 3.729, p = .161$[b] |
| Education (UK level or equivalent) | $\chi^2(6) = 3.073, p = .803$[b] |
| Previous AR/VR HMD usage | $\chi^2(6) = 2.413, p = .888$[b] |
| Previous NG visits | $\chi^2(6) = 11.032, p = .085$[b] |
| Country of Origin | $\chi^2(28) = 55.866, p = .007$*[a] |

*Note.* [a] asymptotic significance (2-sided) [b] exact significance (2-sided).
* denotes significance at $p < .05$ level.

**Table 2**
Narrative version frequencies per device.

| Device | Guide | | |
| --- | --- | --- | --- |
| | Abbot | Curator | Total |
| Prism Mira | 38 | 26 | 64 |
| Magic Leap | 71 | 51 | 122 |
| Oculus Quest | 117 | 65 | 182 |
| Total | 226 | 142 | 368 |

*Note.* Table shows how often either narrative version was selected by participants, per device.

### 2.7. Abbot and curator narrative versions

As mentioned previously, all participants could choose between two narrative versions, either told by the Abbot or Curator. To assess if this choice may have influenced the survey responses, additional analyses were conducted. Independent samples Kruskal-Wallis tests showed that there was no significant difference between the two narrative versions in terms of enjoyment ($H(1) = 0.665$, $p = .415$), overall presence ($H(1) = 0.001$, $p = .981$), cognitive engagement ($H(1) = 0.618$, $p = .432$), emotional engagement ($H(1) = 1.883$, $p = .170$), behavioural engagement ($H(1) = 0.844$, $p = .358$) or negative effects ($H(1) = 0.000$, $p = .998$). Of all individual questionnaire items only the mean of 'I knew what to do' between the Curator ($M = 4.36$, $SD = 0.951$) and the Abbot ($M = 4.07$, $SD = 1.135$) was significantly different ($H(1) = 6.405$, $p = .011$ after Bonferroni correction) in favour of the Curator. The effect size for this analysis (d = 0.245, $\eta^2 = 0.015$) (Lenhard & Lenhard, 2016) was in line with Cohen's (2013) convention for a small effect (d = 0.20). This difference may have been due to the Curator version being more factual and more in line with traditional gallery information. In addition, a number of participants who had selected the Abbot stated after the experience that they were unsure if the experience had ended when the actors left the chapel to go to mass. A chi-square test showed that there was no significant effect of device on narrator selection ($\chi^2(2) = 1.279$, $p = .536$). The frequencies are presented in Table 2. Therefore, for the purpose of this article the data from the two narrative versions are combined.

## 3. Results

### 3.1. Initial analysis

Initial analyses were first run at overall levels (for example on overall presence), followed up by analyses on lower levels (for example individual items relating to 'being there', 'felt real', etc.). To test if means were significantly different between the three devices, independent-samples Kruskal Wallis tests were conducted. Pairwise post hoc comparisons used adjusted $p$-values for Bonferroni correction for multiple tests. The means are presented in graphs and test statistics in tables, per concept.

### 3.2. Enjoyment

As shown in Fig. 6, participants enjoyed the experience on all devices, with the post hoc analysis shown in Table 3 showing that the Oculus Quest received significantly higher mean scores than both AR devices.

### 3.3. Presence

As shown in Fig. 7, on overall presence (average score of all presence questions), the Oculus Quest received a significantly higher overall mean score than both the Magic Leap and the Mira Prism. These significant differences were also visible on the individual presence items, other than realism, where the Oculus Quest scored only significantly higher than the Magic Leap, as shown in Table 4.

Qualitative insight on user experiences was gathered through informal conversations with 18 participants, 7 hours of observation and an open-ended question in the survey. Regarding presence, participants commented that they felt like they were present in the chapel, or like they were transported to another space or time. One woman remarked positively about a sense of being alone with the painting in the middle of a crowd. A teenage Italian male stated 'It's so real, so cool. I feel like I was there.' A young South Korean male (age 13, Quest, Abbot Story) stated 'People in the VR were realistic, and talking as they are real people who lived at that century. Realistic.'
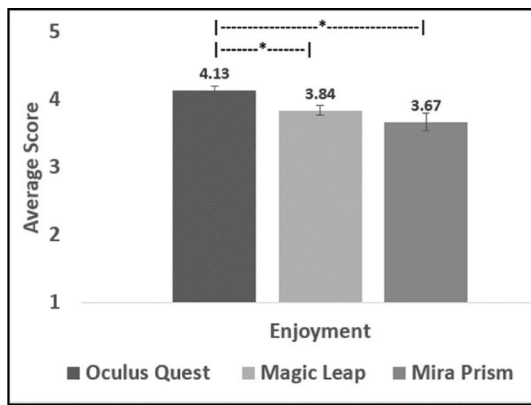
**Fig. 6.** *Enjoyment per Device Note.* Average enjoyment ratings over the different headsets. Error bars are 2 SEM.
* denotes significance at p < .05 level.

**Table 3**
Enjoyment per device significance test.

| Questionnaire Items | Kruskal-Wallis Test (3 devices) | Post-hoc Oculus Quest vs. Magic Leap | Post-hoc Oculus Quest vs. Mira Prism | Post-hoc Magic Leap vs. Mira Prism |
|---|---|---|---|---|
| Enjoyment | $H(2) = 14.873$, $p = .001*$ | $p = .010*$ | $p = .003*$ | $p = 1.000$ |

*Note.* Asymptotic significances are displayed.
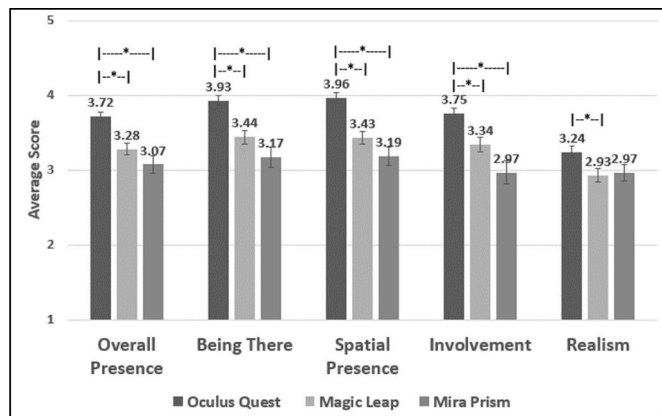* denotes significance at p < .05 level.



**Fig. 7.** *Presence and its Items, Means per Device Note.* Figure shows mean scores for overall presence and its items. Error bars are 2 SEM.
* denotes significance at p < .05 level.

### 3.4. Engagement

#### 3.4.1. Cognitive engagement

As per Fig. 8, cognitive engagement (overall and its component items) was very similar between the three devices and a Kruskal-Wallis test found no significant differences (see Table 5).

#### 3.4.2. Emotional engagement

As shown in Fig. 9, emotional engagement was fairly low and very similar between devices. Very few negative emotions were reported. As shown in Table 6, there were no significant differences between devices.

To illustrate some of the positive emotions experienced, a South African female (33 years old, Prism Mira, Abbot) exclaimed when taking

**Table 4**
Presence and its component items, Significance Tests.

| Questionnaire Items | Kruskal-Wallis Test (3 devices) | Post-hoc Oculus Quest vs. Magic Leap | Post-hoc Oculus Quest vs. Mira Prism | Post-hoc Magic Leap vs. Mira Prism |
|---|---|---|---|---|
| Overall Presence | $H(2) = 33.543$, $p = .000*$ | $p = .000*$ | $p = .000*$ | $p = .535$ |
| Being There | $H(2) = 31.102$, $p = .000*$ | $p = .000*$ | $p = .000*$ | $p = .380$ |
| Spatial Presence | $H(2) = 37.686$, $p = .000*$ | $p = .000*$ | $p = .000*$ | $p = .486$ |
| Involvement | $H(2) = 23.975$, $p = .000*$ | $p = .000*$ | $p = .000*$ | $p = .175$ |
| Realism | $H(2) = 8.412$, $p = .015*$ | $p = .023*$ | $p = .153$ | $p = 1.000$ |

*Note.* Asymptotic significances are displayed. For the post-hoc analyses the significant values have been adjusted by Bonferroni correction.
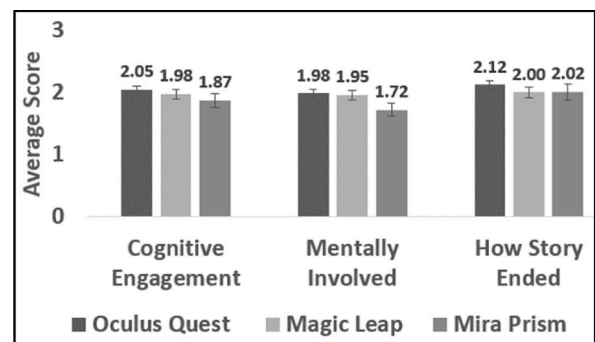* denotes significance at p < .05 level.



**Fig. 8.** Cognitive Engagement and its items, Means per Device Note. Error bars are 2 SEM.
* denotes significance at p < .05 level.

**Table 5**
Cognitive Engagement and its items, Significance Tests.

| Questionnaire Items | Kruskal-Wallis Test (3 devices) |
|---|---|
| Cognitive Engagement | $H(2) = 1.452$, $p = .484$ |
| Mentally Involved | $H(2) = 4.182$, $p = .124$ |
| How Story Ended | $H(2) = 1.069$, $p = .586$ |

*Note.* Asymptotic significances are displayed.
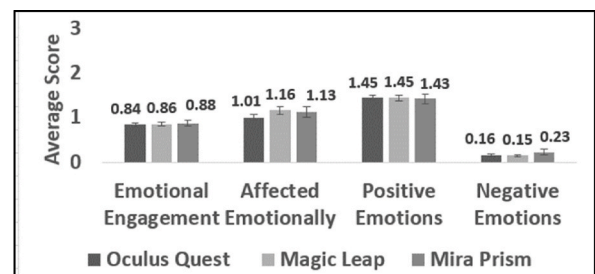* denotes significance at p < .05 level.



**Fig. 9.** *Emotional Engagement and its items, Means per Device Note.* Error bars show 2 SEM.

the headset off, 'That is so moving! I feel kind of emotional!'.

#### 3.4.3. Behavioural engagement

As shown in Fig. 10, all devices were able to create high behavioural engagement. As shown in Table 7, 'I would like to see more experiences

**Table 6**

Emotional Engagement and its items, Significance Test.

| Questionnaire Items | Kruskal-Wallis Test (3 devices) |
|---|---|
| Overall Emotional | $H(2) = .388, p = .824$ |
| Affected Emotionally | $H(2) = 1.842, p = .398$ |
| Positive Emotions | $H(2) = .022, p = .989$ |
| Negative Emotions | $H(2) = .669, p = .716$ |

*Note.* Asymptotic significances are displayed. Multiple comparisons are not performed because the overall test does not show significant differences across samples.

like this' differed significantly between devices, with a higher mean score for Oculus Quest than Mira Prism. 'Tell friends' also differed significantly across the devices overall, however the post-hoc tests did not identify significant differences due to correction for multiple comparisons. 'Knew what to do' was also high for all devices and significantly higher for Oculus Quest than Mira Prism, as shown in Fig. 10 and Table 7.

### 3.5. Negative effects

As shown in Fig. 11, few negative effects were reported by participants. As shown in Table 8 'felt uncomfortable' was low on average but significantly higher on Mira Prism than on the other two devices, possibly because this AR device was only used by participants wearing (prescription) glasses - wearing AR glasses over prescription glasses may have felt uncomfortable for some participants. 'Felt disconnected' was low on average but significantly higher on the Oculus Quest than the Magic Leap, possibly because the Quest is a fully enclosed system, visually disconnecting people from their body and surroundings, whereas the other two visually mix reality and computer generated elements.

### 4. Discussion

This study examined whether VR and two different AR versions of an immersive cultural experience can lead to different user enjoyment, presence, engagement, and/or negative side-effects.

Although enjoyment was high on all devices (with average scores of around 4 out of 5), VR received higher enjoyment scores than both AR devices. The VR version also scored more highly than both AR devices on all presence elements, other than realism where the VR score was only significantly higher than that for Magic Leap.

These effects may have been driven by variation in immersion between the devices. Even though immersion was not measured objectively between the devices, it can be assumed that the Oculus Quest and the Magic Leap had higher immersion than the Mira Prism, as the first two benefitted from higher resolution textures inside the chapel and

**Table 7**

Behavioural Engagement and its items and Knew What to Do, Significance Tests.

| Questionnaire Items | Kruskal-Wallis Test (3 devices) | Post-hoc Oculus Quest vs. Magic Leap | Post-hoc Oculus Quest vs. Mira Prism | Post-hoc Magic Leap vs. Mira Prism |
|---|---|---|---|---|
| Overall Behavioural | $H(2) = 3.946$, $p = .139$ | a | a | a |
| More Experiences Like This | $H(2) = 7.867$, $p = .020*$ | $p = .245$ | $p = .025*$ | $p = .728$ |
| Gallery Interactive Experiences | $H(2) = .504$, $p = .777$ | a | a | a |
| Repeat This Experience | $H(2) = 4.343$, $p = .114$ | a | a | a |
| Tell Friends | $H(2) = 6.441$, $p = .040*$ | $p = .125$ | $p = .109$ | $p = 1.000$ |
| Look Up More Information | $H(2) = .680$, $p = .712$ | a | a | a |
| Knew What To Do | $H(2) = 8.541$, $p = .014*$ | $p = .183$ | $p = .019*$ | $p = .762$ |

*Note.* Asymptotic significances are displayed. For the post-hoc analyses the significant values have been adjusted by Bonferroni correction. [a] Multiple comparisons are not performed because the overall test does not show significant differences across samples.
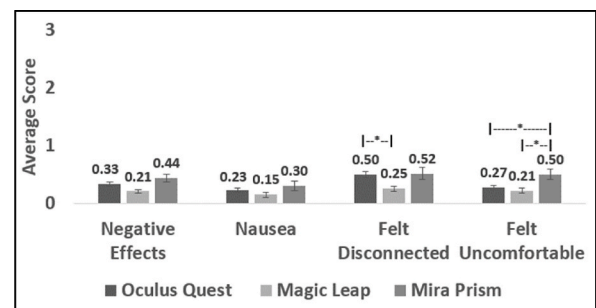* denotes significance at $p < .05$ level.



**Fig. 11.** Negative Effects and its items, Means per Device Note. Error bars show 2 SEM.
* denotes significance at $p < .05$ level.

offered 6DOF, whereas the Mira Prism had 3DOF. In addition, VR is generally seen as more immersive than AR because VR uses fully closed off HMDs, thereby replacing 'real world' sensory input more effectively. According to Slater and Steed (2000), 'place illusion', a key factor in presence, can be broken when a user 'flips' between being in the virtual world and in the real world. Such 'breaks in presence' may be more prominent in AR than in VR, as AR presents users with a mix of sensory
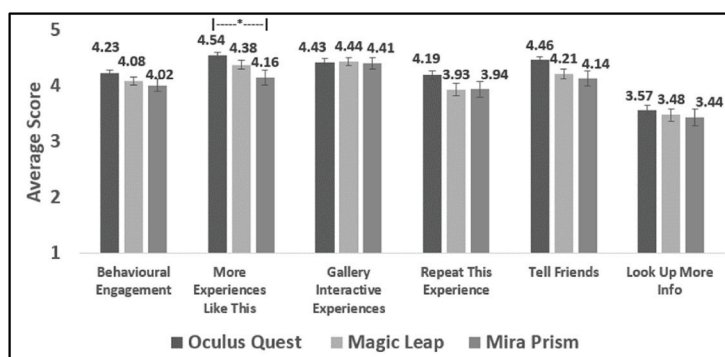


**Fig. 10.** *Behavioural Engagement and its items, Means per Device.* Error bars show 2 SEM.
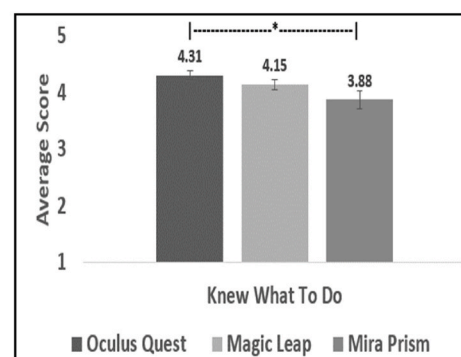* denotes significance at $p < .05$ level.

**Table 8**
Negative Effects and its items, Significance Tests.

| Questionnaire Items | Kruskal-Wallis Test (3 devices) | Post-hoc Oculus Quest vs. Magic Leap | Post-hoc Oculus Quest vs. Mira Prism | Post-hoc Magic Leap vs. Mira Prism |
|---|---|---|---|---|
| Negative Effects | $H(2) = 6.837$, $p = .033*$ | $p = .118$ | $p = 1.000$ | $p = .053$ |
| Nausea | $H(2) = 3.701$, $p = .157$ | [a] | [a] | [a] |
| Uncomfortable | $H(2) = 10.098$, $p = .006*$ | $p = .720$ | $p = .047*$ | $p = .005*$ |
| Disconnected | $H(2) = 9.678$, $p = .008*$ | $p = .007*$ | $p = 1.000$ | $p = .178$ |

*Note.* Asymptotic significances are displayed. For the post-hoc analyses the significant values have been adjusted by Bonferroni correction. [a] Multiple comparisons are not performed because the overall test does not show significant differences across samples.
* denotes significance at $p < .05$ level.

input (e.g. images and light) from the real and virtual world, whereas VR offers only computer generated images. Higher immersion scores for VR than AR were reported by Voit et al. (2019) in their study of immersion levels between five different methods. Therefore, it can be assumed that in the current study, immersion was highest for Oculus Quest, followed by Magic Leap, and Mira Prism. This order is in line with most of the user experience ratings in this study, supporting the possibility that user experience differences may have been driven, at least in part, by differences in the level of immersion provided by the different devices.

Emotional engagement was fairly low, especially for the negative emotions (sadness, anger and fear), possibly due to the factual nature of the content (research about why the painting was commissioned), and/or due to the calm delivery of the content. Emotional engagement did not vary significantly between the different devices, nor between the narrative versions, other than for 'happy' which was slightly higher (statistically significant, $p < .05$) for the Curator than the Abbot ($M_{\text{curator}} = 1.65$, $SD_{\text{curator}} = 0.91$, $M_{\text{Abbot}} = 1.42$, $SD_{\text{Abbot}} = 0.98$, $p = .033$).

Even though there were differences between the experiences delivered by each of the three technologies, the generally high mean scores on most items and the low number of reported negative effects suggest that all the tested devices created an overall positive user experience. It indicates that both VR and AR can be effective immersive storytelling tools in a cultural institution. This is in line with earlier research suggesting that both VR and AR can create positive user experiences, in cultural and other settings. More research is required that actively compares other types of immersive technologies and/or more traditional technologies like 2D computer monitors, to expand our understanding of the ways in which immersive technologies might affect user experiences.

One limitation of the current study relates to the sampled population – the National Gallery's on average highly educated, art- and culture-interested visitor profile limits the generalizability of the findings to other visitor profiles. On the other hand, the close to 50-50 gender split, 40+ countries of origin and wide age range (13–77) improve generalizability by comparison with much of the existing work in this field. Another limitation relates to the duration of the experience and the low number of reported negative effects. The experience lasted for only 6 min, and it is possible that more negative effects (such as fatigue and feeling uncomfortable) could have been reported if the intervention had lasted longer.

Future research can look into the importance of the Gallery setting on user experience, and to what extent the VR user experience is influenced by the physical presence of the (unseen in VR) physical painting. These factors could influence how well the experience may be received in other environments or geographical locations. In addition, since data were collected immediately after the experience, future research could cover additional time points, for example pre-experience (user expectations), and user experiences during, immediately after the experience,

and longer term (Olsson, 2013).

### Credit author statement

Author contributions were as follows: Literature review by Andy Woods and Isabelle Verhulst. Study design by Andy Woods, Polly Dalton, James Bennett and Isabelle Verhulst. AR and VR prototypes by Focal Point VR and StoryFutures. Narrative design by Catherine Skinner, Rebecca Gill, James Bennett, Will Saunders and Lawrence Chiles. Data collection by Laryssa Whittaker, Isabelle Verhulst and Andy Woods. Data analysis by Isabelle Verhulst. Manuscript written mainly by Isabelle Verhulst with review and contributions by all other authors.

### Acknowledgements

### References

AGO. (2017). *Reblink*. Retrieved from www.ago.ca/exhibitions/reblink.

Appel, M., Gnambs, T., Richter, T., & Green, M. C. (2015). The transportation scale–short form (TS–SF). *Media Psychology, 18*(2), 243–266.

Aslan, I., Dang, C. T., Petrak, B., Dietz, M., Filipenko, M., & André, E. (2019, November). Viewing experience of augmented reality objects as ambient media-a comparison of multimedia devices. In *European conference on ambient intelligence* (pp. 324–329). Cham: Springer.

Attfield, S., Kazai, G., Lalmas, M., & Piwowarski, B. (2011, February). Towards a science of user engagement (position paper). In *WSDM workshop on user modelling for Web applications* (pp. 9–12).

Azuma, R. (2015). *11 location-based mixed and augmented reality storytelling*.

Bouvier, P., Lavoué, E., & Sehaba, K. (2014). Defining engagement and characterizing engaged-behaviors in digital gaming. *Simulation & Gaming, 45*(4–5), 491–507.

Boyle, E. A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior, 28*(3), 771–780.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309.

Dey, A., Billinghurst, M., Lindeman, R. W., & Swan, J. (2018). A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI, 5*, 37.

tom Dieck, M. C., & Jung, T. H. (2017). Value of augmented reality at cultural heritage sites: A stakeholder approach. *Journal of Destination Marketing & Management, 6*(2), 110–117.

Dow, S., Mehta, M., Harmon, E., MacIntyre, B., & Mateas, M. (2007). Presence and engagement in an interactive drama. In *Conference on human factors inComputing systems – proceedings (san Jose, CA)* (pp. 1475–1484).

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Fiske, A. P., Seibt, B., & Schubert, T. (2019). The sudden devotion emotion: Kama muta and the cultural practices whose function is to evoke it. *Emotion Review, 11*(1), 74–86.

Freeman, J., Lessiter, J., Borden, P., Turner-Brown, L., Kurta, L., & Ferrari, E. (2019, September). *The immersive audience experience evaluation toolkit*. IBC.

Georgiou, Y., & Kyza, E. A. (2017). The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings. *International Journal of Human-Computer Studies, 98*, 24–37.

Graham, P. (Jan 20, 2016). *Salvador Dali museum launching VR experience*. VR Focus. www.vrfocus.com/2016/01/salvador-dali-museum-launching-vr-experience/.

Grassini, S., & Laumann, K. (2020). Questionnaire measures and physiological correlates of presence: A systematic review. *Frontiers in Psychology, 11*, 349.

Gröppel-Wegener, A., & Kidd, J. (2019). *Critical encounters with immersive storytelling. Routledge.]*.

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. *PloS One, 11*(8). e0159915.

Hein, D., Mai, C., & Hußmann, H. (2018). The usage of presence measurements in research: A review. In *Proceedings of the international society for presence research annual conference (presence). The international society for presence research*.

He, Z., Wu, L., & Li, X. R. (2018). When art meets tech: The role of augmented reality in enhancing museum experiences and purchase intentions. *Tourism Management, 68*, 127–139.

Jung, T., tom Dieck, M. C., Lee, H., & Chung, N. (2016). Effects of virtual reality and augmented reality on visitor experiences in museum. In *Information and communication technologies in tourism 2016* (pp. 621–635). Cham: Springer.

Kardong-Edgren, S. S., Farra, S. L., Alinier, G., & Young, H. M. (2019). A call to unify definitions of virtual reality. *Clinical Simulation in Nursing, 31*, 28–34.

Kidd, J., & Nieto, E. (2019). *Immersive experiences in museums, galleries and heritage sites: A review of research findings and issues*.

Kim, K., Billinghurst, M., Bruder, G., Duh, H. B. L., & Welch, G. F. (2018a). Revisiting trends in augmented reality research: A review of the 2nd decade of ISMAR (2008–2017). *IEEE Transactions on Visualization and Computer Graphics, 24*(11), 2947–2962.

Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018b). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics, 69*, 66–73.

Lenhard, W., & Lenhard, A. (2016). Berechnung von Effektstärken [Calculation of effect sizes]. *Dettelbach: Psychometrica*. Available online at: https://www.psychometrica.de/effektstaerke.html. (Accessed 15 February 2020).

Loizides, F., El Kater, A., Terlikas, C., Lanitis, A., & Michael, D. (2014, November). Presenting cypriot cultural heritage in virtual reality: A user evaluation. In *Euro-mediterranean conference* (pp. 572–579). Cham: Springer.

Magic Leap Developer. (February 26, 2020). *Content persistence: Local – unity*. https://developer.magicleap.com/en-us/learn/guides/persistence-tutorial-local-unity.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems, 77*(12), 1321–1329.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995, December). Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies* (Vol. 2351, pp. 282–292). International Society for Optics and Photonics.

Nesta & i2 Media Research. (2018). *Evaluating immersive user experience and audience Impact summary report A report produced by for digital catapult*. https://www.immerseuk.org/wp-content/uploads/2018/07/Evaluating_Immersive_User_Experience_and_Audience_Impact.pdf.

Olsson, T. (2013). Concepts and subjective measures for evaluating user experience of mobile augmented reality services. In *Human factors in augmented reality environments*. New York, NY: Springer.

Schubert, T., Friedmann, F., & Regenbrecht, H. (1999). Embodied presence in virtual environments. In *Visual representations and interpretations* (pp. 269–278). London: Springer.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). *Igroup presence questionnaire*.

See, Z. S., Sunar, M. S., Billinghurst, M., Dey, A., Santano, D., Esmaeili, H., & Thwaites, H. (2017, November). An augmented reality and virtual reality pillar for exhibitions: A subjective exploration. In *ICAT-EGVE* (pp. 79–82).

Slater, M., Lotto, B., Arnold, M. M., & Sánchez-Vives, M. V. (2009). How we experience immersive virtual environments: The concept of presence and its measurement. *Anuario de Psicología, 2009, 40*, 193–210.

Slater, M., Navarro, X., Valenzuela, J., Oliva, R., Beacco, A., Thorn, J., & Watson, Z. (2018). Virtually being lenin enhances presence and engagement in a scene from the Russian revolution. *Frontiers in Robotics and AI, 5*.

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI, 3*, 74.

Slater, M., & Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators and Virtual Environments, 9*(5), 413–434.

Steed, A., Pan, Y., Watson, Z., & Slater, M. (2018). 'We wait'-the Impact of character responsiveness and self embodiment on presence and interest in an immersive news experience. *Frontiers in Robotics and AI, 5*, 112.

Suh, A., & Prophet, J. (2018). The state of immersive technology research: A literature analysis. *Computers in Human Behavior, 86*, 77–90.

Sylaiou, S., Kasapakis, V., Dzardanova, E., & Gavalas, D. (2018, September). Leveraging mixed reality technologies to enhance museum visitor experiences. In *2018 international conference on intelligent systems (IS)*. IEEE.

Sylaiou, S., Mania, K., Karoulis, A., & White, M. (2010). Exploring the relationship between presence and enjoyment in a virtual museum. *International Journal of Human-Computer Studies, 68*(5), 243–253.

Tussyadiah, I. P., Wang, D., Jung, T. H., & tom Dieck, M. C. (2018). Virtual reality, presence, and attitude change: Empirical evidence from tourism. *Tourism Management, 66*, 140–154.

Voit, A., Mayer, S., Schwind, V., & Henze, N. (2019, May). Online, VR, AR, lab, and in-situ: Comparison of research methods to evaluate smart artifacts. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–12).

Weibel, D., & Wissmath, B. (2011). Immersion in computer games: The role of spatial presence and flow. *International Journal of Computer Games Technology, 2011*.

Wikitude GmbH. (2021a). Extended tracking augmented reality. AR SDK for extended marker-based AR experiences https://www.wikitude.com/augmented-reality-extended-tracking/.

Wikitude GmbH. (2021b). *Object & scene tracking augmented reality*. https://www.wikitude.com/augmented-reality-object-scene-recognition/.

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