frontiers | Frontiers in Computer Science

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fcomp.2022.927177

Developing a play-anywhere handheld AR storytelling app using remote data collection

Gideon Raeburn*, Martin Welton and Laurissa Tokarchuk

School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom

Check for updates

OPEN ACCESS

EDITED BY
Carlos Vaz De Carvalho,
Polytechnic Institute of Porto, Portugal

REVIEWED BY
Matjaž Kljun,
University of Primorska, Slovenia
Kamran Khowaja,
Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology, Pakistan

*CORRESPONDENCE
Gideon Raeburn
g.raeburn@qmul.ac.uk

SPECIALTY SECTION
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

RECEIVED 23 April 2022
ACCEPTED 23 August 2022
PUBLISHED 15 September 2022

CITATION
Raeburn G, Welton M and Tokarchuk L (2022) Developing a play-anywhere handheld AR storytelling app using remote data collection. *Front. Comput. Sci.* 4:927177. doi: 10.3389/fcomp.2022.927177

Immersive story experiences like immersive theater productions and escape rooms have grown in popularity in recent years, offering the audience a more active role in the events portrayed. However, many of these activities were forced to close at the start of the COVID-19 pandemic, arising from restrictions placed on group activities and travel. This created an opportunity for a story experience that users could take part in around their local neighborhoods. Five mobile applications (apps) were developed toward this goal, aiming to make effective use of available local map data, alongside virtual content overlaid on users' surroundings through Augmented Reality (AR), to offer additional story features not present in the real environment. The first two apps investigated the feasibility of such an approach, including the remote field testing of the apps, where participants used their own devices across a variety of locations. Two follow-up apps further aimed to offer an improved user experience, also adopting a more standardized testing procedure, to better ensure each app was completed in an intended manner by those participating remotely. Participants rated their experience through immersion and engagement questionnaire factors that tested for their appropriateness to rate such experiences, in addition to providing their feedback. A final app applied the same AR story implementation to a curated site-specific study, once pandemic restrictions had eased. This combination of remote studies and subsequent curated study offered a reverse methodology to much previous research in this field, but was found to offer advantages in corroborating the results of the remote studies, and also in offering new insights to further improve such an AR story app, that is designed to be used at an outdoor location of the user's choosing. Such an app offers benefits to those who may prefer the opportunity to take part in such an activity solo or close to home, as well as for storytellers to develop an outside story for use at a variety of locations, making it available to a larger audience, without the challenges and costs in migrating it to different locations.

KEYWORDS
augmented reality, immersion, storytelling, location-based experiences, mobile applications

1. Introduction

COVID-19 pandemic restrictions caused major issues for experiences involving a live audience. This included a variety of immersive story experiences, that encourage greater participation in the story's events, such as through interaction with actors, props, or through the use of a mediating device, as the audience navigate around a physical set or public space (Benford et al., 2009; Kasapakis and Gavalas, 2015). Such experiences have also demonstrated benefits in the heritage sector, encouraging new visitors who may not typically choose to visit a traditional museum (Vayanou et al., 2014; Krzywinska et al., 2020). Attempting to fill the gap caused by the temporary closure of such experiences, the current research investigated whether the sensors present in modern mobile phones could provide sufficient data about the user and their surrounding location, to offer them an immersive story that could be experienced outdoors from almost any choice of starting location. This follows a precedent to build stories around available location data or transposing a site-specific narrative to different locations procedurally. However, constructing a narrative from a limited database of place information has struggled to offer a coherent story, given the need to link the available data together in a natural and meaningful way (Tregel et al., 2017). Attempts to translate a narrative from one location to another, have included WeQuest (2011), and LAGATO's LeGAL (2019), which aimed to identify equivalent story sites with either comparable map features, or from available user reviews (Macvean et al., 2011; Ferreira et al., 2019). Such methods invariably affected the amount of walking involved, with each new site commonly also offering its own distinct atmosphere.

The solution considered by this research set out to avoid the requirement to find an exact match for each story site visited, through the use of Augmented Reality (AR). AR is available through modern high end smartphones and can overlay virtual content on the camera view of the real world, visible on the phone's screen. AR was, therefore, used to enhance real world locations with virtual content relating to the story. This form of handheld AR has become more widely known through the release of Niantic's Pokémon GO (2016), a Location-Based Game (LBG), where real world Points of Interest (POIs) provide opportunities to collect in-game items, and catch virtual creatures (Dunham et al., 2018). Some of the game's POIs were identified through the crowdsourcing and peer review of user submitted suggestions, providing additional information like safety and accessibility considerations, and contributing to Niantic's detailed *Portal Network* of worldwide map POIs. This database is not publicly available for use as part of this research but has allowed Niantic to offer a more coherent user experience of its games across a variety of locations (Laato et al., 2019). However, Niantic still introduced changes to their games during the pandemic, to make some challenges less dependent on collaboration and walking to each POI, in order to trigger game events. Some players reported that these changes encouraged a renewed interest to pick up the games again (Ellis et al., 2020). Crowdsourced data like that used as part of Niantic's Portal Network has been found capable of being of equivalent accuracy compared to data collected in the lab, as well as providing data from a wider range of demographics (Manzo et al., 2015). Used in a game, such data may offer user rewards in terms of recognition or added personal significance, though care may be needed when a financial reward is offered for submitting data, with risks some users might try to obtain it with minimal effort (Detering, 2014).

An analysis of previous AR studies suggests the majority are lab-based or curated field studies, commonly collecting data on task completion, and/or subjectively rating an experience through appropriate questionnaires or user feedback (Dey et al., 2018). However, in a location-based experience intended for use at a variety of locations, there is also the viewpoint that only testing in a single controlled environment would miss the unpredictable elements that are often central to such an experience, and may only become evident during *in the wild* testing. Partially influenced by ongoing pandemic restrictions, a method of remote field testing was ultimately adopted with four out of five of the mobile applications created, subsequently referred to as apps, with participants downloading and testing each app on a personal device. A Design Research Method was adopted to carry out this research, that aims to improve both design and research practices concurrently, and has been adapted from the field of education to video games, through iterative rounds of design, implementation, analysis, and redesign (Blessing et al., 2009; Viudes-Carbonell et al., 2021). In the wild testing has been found to pose additional challenges in this regard, due to the risks of missing important information, where questionnaires and feedback provide the primary measures of data collection (Ducasse et al., 2020; Lehman et al., 2020). Remote studies also risk potential hardware variations, as well the feedback covering a wide range of themes, that can risk influencing the research on tangential paths (Ratcliffe et al., 2021; Zhao et al., 2021).

Five different app designs were ultimately tested, each offering its own unique story experience relating to the features each app was investigating, which followed the results collected with the previous version. The testing procedure was simultaneously updated for each app, following a discussion in associated literature around the current lack of tools and study processes in AR research, that lags behind those for Virtual Reality (VR) (Kim et al., 2018). A detailed analysis of the user feedback to each app tested, therefore, played an important role in corroborating or following up any contradictions observed with the questionnaire responses received. The first two apps offered a preliminary examination of using handheld AR and location data to tell an immersive story, while testing a variety of existing questionnaire factors for their appropriateness. These first apps highlighted a wariness from some users in taking part in an outside activity during the pandemic, and the need to implement additional features to ensure data was only kept from those who had completed each study as intended. Two follow-up apps aimed to adopt a more standardized testing procedure, while also focusing more on particular aspects of the AR implementation, to compare how these affected users' engagement and immersion in each story. A fifth and final app reapplied the lessons learned to a new story as part of a curated site-specific study, once pandemic restrictions had eased. In this user study, all participants took part at the same location with a moderator present beforehand, and the majority adopted to use a single device. The results from this curated study largely agreed with the data collected from the earlier remote studies, while suggesting some potential benefits to such a reverse methodology, where a curated study is conducted later in the design process. The ability to observe participants using the app during the curated study also offered new directions for improving the user experience, many of these ideas having counterparts as part of a non site-specific experience.

2. Literature review

The current research follows an investigation into previous work around real world stories and LBGs, including those using AR, which here played a significant role in providing virtual content to bring users' surroundings closer in line with the story events. A discussion of measuring user immersion and engagement with such experiences is also presented here, with some associated questionnaires from related fields presented, which were ultimately adopted in testing the five apps that were developed as part of this research.

2.1. Pervasive games and AR experiences

The current research ultimately sets out to blend the proposed story world with the user's physical environment. This closely relates to the field of pervasive games and AR experiences, which aim to offer and blend appropriate content alongside the user's current surroundings and activity (Grubert et al., 2017). LBGs are one such example, where gameplay relates to players' real world location and movement, often detected using a suitable device (de Lange, 2009; Leorke, 2018). Such experiences commonly encourage users to attach new meanings to their surroundings, creating ambiguity that may give rise to Reid's *magic moments*, such that the fictional events are perceived to mirror those in the real world (Reid et al., 2005; Dansey, 2008). Such moments have been associated with a temporary increased sense of immersion, though tend to rely on a detailed knowledge of the user's location to deliver them effectively (Reid, 2008). Adopting a new technology like AR also requires sufficient on-boarding, to overcome its unfamiliarity and novelty, in order for users to then be more open to a deeper experience (Papagiannis, 2017). Earlier studies have offered guidelines for such experiences, connected to both pragmatic and aesthetic considerations, both to maximize users' ease of navigation, as well as to ensure the physical location supports users in feeling part of the unfolding events (Packer et al., 2017). Taking part in public offers additional challenges through possible distractions, and self-consciousness around the existing behavior each user considers appropriate for the space, tied to Goffman's notion of *frames* (Goffman, 1986; Montola, 2011). This contributes to the short, moderately easy tasks commonly adopted in such experiences, with a mediating technology like a mobile phone also proving uncomfortable to hold for long periods (Korhonen et al., 2008).

Such games and story experiences often incorporate a variety of different elements, leading to challenges in studying them in relation to what factors proved most significant. Murray suggested that any new storytelling medium must offer a balance of agency, immersion, and transformation (Murray, 1998). The current research does vary some aspects of user agency, but to simplify the experiences offered, each app tended to adopt a single often linear narrative. There is evidence that as long as a story continues in its original context while offering curiosity and suspense, an audience can be willing to suspend their disbelief, even when adopting a more passive role, this being tied to adults' desire for immersion (Brooks, 2003; Tanenbaum and Tanenbaum, 2009). Here, definitions of immersion may vary according to the quality of the multi-sensory experience offered, or in terms of users generating meaning from an experience, as they actively apply their imagination as to how the events may play out. This may relate to the story, its theme, the rules of an experience, or an opportunity for role-playing, with outstanding questions around the use of virtual elements, such as whether they act primarily as props in this regard, similar to the choice of words in a literary story (Deterding, 2016; McDonnell and Wildman, 2019). Such pervasive stories have the potential to also offer various interaction opportunities, borrowing elements from both video games, where on-screen tasks and exploration help an emergent narrative to form in the mind of the player, as well as immersive theater, where using multiple senses can aid understanding (Qin et al., 2009; Machon, 2011). Several heritage projects have demonstrated benefits from offering technology mediated experiences. This parallels a growing trend to have stories offered in an original way such as using AR, rather than just recounting the history of a space (Ballagas et al., 2007; Bryon, 2012). Mobile phones also offer additional opportunities due to them being widely available and more comfortable than current AR headsets, with new sensors such as Lidar and depth cameras offer further tools to identify users' surroundings or offer the potential for more embodied methods of interaction (Baker, 2017; Du et al., 2020).

2.2. Measuring users' immersion and engagement

Pervasive AR experiences and LBGs typically incorporate features that are hard to capture in questionnaires designed for other mediums, due to the real world not being tailored specifically for them, unlike a typical video game environment (Denisova et al., 2016). The Pervasive GameFlow Model suggests that the criteria that typically give rise to Csíkszentmihályi's optimal *flow* state, are unlikely to all be met simultaneously in a pervasive game. This occurs, e.g., because a player will always need to focus some attention on their surroundings when navigating, in order to maintain their safety (Jegers and Wiberg, 2006). A model appropriate to such experiences may be seen in Benford's trajectories model for mixed reality experiences, which considers how all the real and virtual elements contribute to a user's unique path through the experience, that will inevitably deviate from the designer's intended route (Benford et al., 2009). Collecting sufficient data as part of a remotely conducted study to improve the user experience is challenging. Questionnaires offer a common method of data collection for such studies, though they are subjective and risk interrupting users' immersion, or rely on memory if completed later (Schirm et al., 2019). Various validated questionnaires also highlight the lack of agreed definitions for engagement and immersion as applied to different systems. Slater suggests immersion should focus on the extent to which a system offers sense modalities that support the proposed environment, while presence should refer to the subjective quality by which a user then feels part of that world (Slater, 1999). This differs slightly from Brown and Cairns' model, which considers three increasing levels of immersion, each related to the removal of successive barriers, culminating in the highest level of total immersion associated with a sense of both presence and flow (Brown and Cairns, 2004). This model has been validated for use in AR through the Augmented Reality Immersion (ARI) questionnaire (Georgiou and Kyza, 2017).

An attempt to quantify user experience across various mixed reality systems, can be found in the ITC-Sense of Presence Inventory (ITC-SOPI) from Goldsmiths University, which identifies four sub-scales in terms of users' sense of physical space, engagement, the environment's perceived realness, and any negative effects experienced (Lessiter et al., 2001). The degree to which a user feels transported to a proposed location through a system has also been related to the concept of telepresence, extending from earlier research into telecommunication systems. This forms the basis of the Temple Presence Inventory (TPI) (Lombard et al., 2009). Related research around telling stories in new mediums has also been associated with users' level of transportation to the story world. Transportation extends from literary narratives, being related to becoming less aware of one's surroundings, and engaging in active thought about how the events might transpire. Questionnaires in this area applied to literary or video narratives include the Narrative Transportation Scale (NTS), and Story World Absorption (SWA) scale (Green and Brock, 2000; Kuijpers et al., 2014). In pervasive games and AR experiences, immersion has been considered separately in terms of perceptual and psychological immersion (Mcmahan, 2003). Perceptual immersion may vary according to the degree that a user's environment matches that described, found to vary in the study of iLand (2012), where participants watched a video narrative at either a matching or unrelated location (Karapanos et al., 2012). Psychological immersion occurs through a willing desire to be mentally absorbed in events, and so may vary between different people. Identifying relevant player types to offer a more personalized experience has been a significant area of video game research. Tools in this regard include the OCEAN Big-5 Personality Framework, which rates five key personality factors (John and Srivastava, 1999). In VR, a relationship was observed between a user's sense of presence in the VR world, and their immersive tendency, as measured with the Immersive Tendencies Questionnaire (ITQ) (Witmer and Singer, 1998). Sections of the various questionnaires described form part of the testing process in the current research, with these quantitative measures compared against users' feedback, to offer evidence for their appropriateness in relation to such a handheld AR experience.

3. Method

Table 1 offers a breakdown of the features of the five mobile applications developed and tested as part of this research, which are subsequently referred to as apps ① to ⑤. All five apps were created in the Unity game engine and made available to download for both iOS and Android handsets, capable of supporting AR through the camera (except app ③ which was only released for iOS phones containing a depth camera). AR functionality was implemented through the *ARFoundation* plugin, which includes the SLAM tracking of any virtual objects placed relative to the real world environment (Klein and Murray, 2009). Each of the five apps was created with a new story, features, and modified testing procedure, according to the results and participant feedback on the previous version. A summary of this development process across the five apps is presented in Section 4, through the main results collected from the study with each app. Table 1 also demonstrates how most of the apps were tested in multiple versions as part of a mixed methods study design, with the different versions enabling a direct comparison of particular features. Each app underwent a period of pilot testing, with participants using the apps at their own choice of a location similar to the final user studies. This pilot testing resulted in various modifications being made to each app before it was launched. Questionnaire data and feedback provided at various points within each app were uploaded directly to a Google Form on the app's completion.

TABLE 1 Summary of all five apps relating to this research and their respective features.

App design feature		Original app name and version reference code number								
		Map Story 1	Map Story 2		Home Story		Map Story 3			Map Story 4
		①	2a	2b	3a	3b	4a	4b	4c	⑤
Remote study		✓	✓	✓	✓	✓	✓	✓	✓	
Recruitment	Mailing lists	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Social media	✓	✓	✓	✓	✓	✓	✓	✓	
	Prolific		✓	✓	✓	✓	✓	✓	✓	
Device	User's Android device	✓	✓	✓			✓	✓	✓	✓
	User's iOS device	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Curated iOS device									✓
Location adaptive		✓	✓	✓			✓	✓	✓	✓
Map API	Mapbox	✓	✓	✓			✓	✓	✓	✓
	Google Places		✓	✓			✓	✓	✓	✓
Radius of local search area /metres		500	250	250			250	250	250	
Walking between sites		✓	✓	✓			✓			✓
Virtual background					✓	✓			✓	
AR tasks	Imagination	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Interaction			✓	✓	✓	✓	✓	✓	✓
Gesture interaction					✓	✓				
Audio	Sound effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Text to speech	✓	✓	✓	✓	✓	✓	✓	✓	
	Recorded						✓	✓	✓	✓
On-boarding	Instructions	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Initial story chapter				✓	✓	✓	✓	✓	✓
	Face to face									✓
Pre-questionnaires	NEO-FFI-3 Big-5 Personality Inventory	Extroversion, openness (2/5)								
	Immersive tendency Questionnaire (ITQ)		All 3 factors		Becoming involved in an activity (1/3)		All 3 factors			All 3 factors
	ITC sense of presence Inventory (ITC-SOPI)				All 4 pre-questions					
Post-questionnaires	Augmented reality Immersion (ARI)	All 6 factors	Usability (1/6)		All 6 factors		All 6 factors			All 6 factors
	Temple presence Inventory (TPI)		Mental immersion (1/8)							
	Story world absorption Questionnaire (SWA)		Transportation, emotional engagement (2/4)							
	ITC sense of presence Inventory (ITC-SOPI)				Engagement (1/4)					
	Narrative transportation Scale (NTS)						All 3 factors			All 3 factors

The table assigns each app version a code descriptor that is adopted throughout the rest of this paper. The table also lists the existing questionnaires that were tested alongside each app, indicating the factors used when only a subset of the existing questionnaire's factors was adopted. In addition, a series of bespoke questions appropriate to each app was also used as part of the pre and post questionnaires, as well as after each story site (or virtual object interaction).

The five apps tested can be best considered in three different groupings. The first two apps offered largely preliminary studies, also investigating the use of a remote testing procedure. This included investigating the general feasibility of the research direction to use AR alongside location data, to provide an immersive story experience suitable for most outdoor locations through accessible map data. Prior research has identified the lack of questionnaires specifically designed for pervasive games and AR stories. As part of these preliminary studies, a variety of questionnaire factors were employed to investigate their applicability to such an experience. Adopting only a subset of a questionnaire's factors obviously does not provide a complete picture in respect of how the questionnaire was validated. In some cases, this decision was taken because the questions present in the other factors would not make logical sense when applied to an experience set in the real world. In other cases, the questionnaire stated that an appropriate subset of its factors could be used according to the system being tested, with a subset of factors also sometimes adopted purely as a first test of the questionnaire's suitability. Pilot testing also revealed issues in completing too lengthy a questionnaire in a public space due to the distractions posed for participants. Apps ③ and ④ aimed to offer a more standardized testing procedure, re-using the same questionnaires to directly compare users' ratings across the subsequent apps, which were again tested remotely while the pandemic restrictions continued. App ⑤ differed by offering a site-specific story, where participants met a researcher before starting, and also largely used a single mobile device. This user study collected the same complete questionnaire responses and user feedback within the app as in the previous app ④ while providing the additional opportunity to observe participants using the app, as well as speaking to all users on its completion to gather additional feedback. The questionnaire factors adopted with each app's study are presented in Table 1.

3.1. Participant recruitment

The methods of participant recruited varied across each user study, with a breakdown of the participant demographics whose data was retained shown in Table 2. App ① was initially made available for download through internal university mailing lists, this invitation extended to social media groups in the United Kingdom (UK), where users expressed an interest in immersive storytelling. Successful data collection with app ① was made more challenging by a variety of factors, including some users' hesitation in taking part in an outside activity during the pandemic, and issues with the app itself that are expanded upon in Section 4. To further encourage participation, all subsequent studies offered a financial reward for completing each app. The funds available limited the total number of apps that were ultimately created, as well as the number of participants testing each version, aiming for around 30 users, removing data points that showed evidence of instructions not being followed correctly, or through evidence of data straight lining. Apps ② - ④ were also made available on the public study recruitment site, Prolific, for participants in the UK. As discussed in relation to the results for each study, Prolific proved beneficial in making each app available to a wide range of demographics, though also introduced some issues in making sure that those who took part had correctly followed the instructions associated with each study. Due to its anonymity, data from participants recruited on Prolific was primarily collected through each app, though some participants were contacted through the site's anonymous messaging service to clarify particular observations. Participants recruited through other recruitment channels were invited after completing an app, to take part in an optional semi-structured interview specific to the app, to provide additional details regarding their experience. Feedback collected both through the app and in post interviews was later combined, and thematic coding was applied according to Braun and Clarke's method (Braun and Clarke, 2006). This identified common themes, that were compared against the quantitative findings from the questionnaires, to inform the direction taken in subsequent apps, and to improve the testing procedure.

4. Results

This section presents an overview of the user study conducted with each of the five apps, including how the significant results and feedback collected, led to a modification of the testing procedure and direction taken in the design of the subsequent app. For brevity, each app summary presents these results without a detailed breakdown of the statistical tests that were performed on the questionnaire responses. Such an analysis is only presented in terms of a comparison between the data collected from app (4a), the final remotely tested version incorporating an outdoor walk at a user's selected locations, and the curated site-specific app ⑤, where all participants took part at the same place.

4.1. Preliminary user studies

4.1.1. App ①

This app and user study offered the first attempt at remotely testing a story combining AR and local map data, that participants could download and use at their choice of outdoor location. The Mapbox API was used to search for six sites commonly found around urban areas, such as a public house and a church, looking within 500 m of the user's location. These formed the story sites that users would be required to visit in order to locate lost diary pages (displayed virtually through AR), relating to the main story character's disappearance. In

TABLE 2 Demographics of participants whose data were retained after they completed a version of each app.

		Participant count with each app version								
		①	2a	2b	3a	3b	4a	4b	4c	⑤
Gender	Male	15	13	17	12	13	9	10	7	16
	Female	6	14	12	14	18	21	19	21	19
	Other	2	1	1	1	0	0	1	1	0
Age	18–29	9	16	8	19	27	16	22	16	33
	30–39	8	8	16	5	1	9	5	12	2
	40+	6	4	6	3	3	5	3	1	0
Previous AR experience	Limited	17	20	25	20	26	26	25	25	26
	Significant	6	8	5	7	5	4	5	4	9
Recruitment channel	University mailing lists	15	10	8	7	6	9	8	6	35
	Social media	8	2	3	4	3	2	2	1	0
	Prolific website	0	16	19	16	22	19	20	22	0
Number of participants		23	28	30	27	31	30	30	29	35

The table also demonstrates the number of participants recruited from each recruitment channel.

cases where one of the six sites could not be identified close-by based on local map data, a location was selected where users would instead overlay a virtual model of the missing feature on top of the location using AR. These replacement locations were selected based on their size and the land use metadata. In this way, most users would visit a mixture of physical and augmented locations in order to uncover more of the story, with these stages of gameplay shown in Figure 1. In this way, the app offered a single version, with users answering a series of bespoke questions regarding how successful each of the six real or augmented locations visited was at maintaining their engagement in the story. Users also completed the full Augmented Reality Immersion (ARI) questionnaire on the app's completion, providing a measure of their immersion with the AR experience.

App ① suffered from a small conversion rate of app downloads to those who successfully completed the app. In some cases, this came down to some users' unfamiliarity with taking part in academic studies, such that they were put off by the difference in quality between the app and commercially available apps, as well as the request to answer questions relating to their experience while taking part. Additionally, app ① suffered from the selected Mapbox API offering limited map data in some areas, increasing the walking distance for those participants, based on the local sites identified. This led to several users who started the app choosing not to finish it. The Mapbox API had been selected based on an earlier analysis of different map APIs' benefits, with Mapbox offering a higher number of free API calls per month, various free customization tools, and real-time navigation features. Based on the pre-questionnaire data collected from all users who downloaded the app without necessarily completing it, there was a wider range of demographics compared to those recruited just using university mailing lists, highlighting a potential key benefit of conducting such a study remotely. The pre-questionnaire data also suggested no evidence of a relationship between users' Big-5 personality scores and their immersion rated by the ARI questionnaire, though only two of the personality factors were adopted here in terms of users' levels of extroversion and openness, as part of an exploratory investigation. Several users also expressed a desire that their answers to these personality themed questions would have fed back into the events of the story, to personalize it in some way.

The real buildings (pubs, churches, etc...) visited scored higher across a number of the questions answered at each story site visited when compared to those that were alternatively augmented with a virtual model of the missing feature. This included ratings for usability, as well as users' ease of imagining the story events unfolding at the location. However, users' attention to the story was rated significantly higher when they visited the augmented locations, with feedback suggesting that AR had the potential to offer a new type of experience around locations the user already knew well. AR also proved successful in the use of overlays to let users know they were at the correct location, which offered a sense of comfort rather than limiting the belief that was actually part of the story world. However,

10.3389/fcomp.2022.927177

FIGURE 1
App ① gameplay where users visited real world locations marked on a local map, some of which were augmented with a large virtual model of a missing real world feature **(A)**. At each location, the user was required to locate a hidden diary page **(B)**, in order to uncover new story details **(C)**.

the use of large virtual models created challenges for users to overlay them on their surroundings without occlusion issues, also requiring users to stand further back to see and place them on the small field of view offered by a phone screen. This also limited the potential for interaction opportunities with the virtual objects that several users requested in their feedback. The ARI questionnaire rated users' experience in terms of Brown and Cairns' three increasing levels of immersion in form of engagement, engrossment, and total immersion. Each successive level was found to be scored significantly lower across all participants, highlighting the significant barriers to immersion in this initial app design. Feedback suggested immersion only occurred significantly in brief moments of synchronicity, where events mentioned in the story or audio delivered while walking, were closely aligned with observed events in their surroundings, paralleling Reid's magic moments (Reid et al., 2005).

While the story was designed to continue automatically, based on identifying the user had arrived at the next story site through their GPS location, an option to trigger the story to continue manually was found to be necessary during pilot testing, in case the target location was inaccessible, or somewhere the user did not feel safe going. Unfortunately, this led to some participants using this feature at every site visited, making it difficult to verify that they had actually engaged in an outdoor walk, with their data subsequently removed if they had used this feature more than twice. The design of the study with app ① also made it challenging to directly compare different users' experiences since most users had visited a mixture of both physical and augmented locations. Such observations led to subsequent apps adopting between subject conditions, randomly assigning users to one or more versions of the app, to directly compare the differences in their experience. An overall mixed methods design enabled additional features present in the various versions to also be examined through within subject tests.

4.1.2. App ②

This app again offered a preliminary examination of the research direction, given the range of issues identified with app ①, based on the app's varying performance at different starting locations. In this latest app, the Mapbox API was retained for its map and real-time navigation features, but six story sites were now identified using the more extensive Google Places API, now searching for local places of significance based on a variety of criteria, rather than searching for specific map features like particular buildings. This allowed the radius of the search area to be reduced from 500 to 250 m around the user's starting position, and the app to still detect enough suitable sites in all but the most remote locations. App ② also adopted smaller virtual models for users to overlay on their surroundings at each location visited, with the models placed in a set order as the user progressed through the story, tied to uncovering a lost memory.

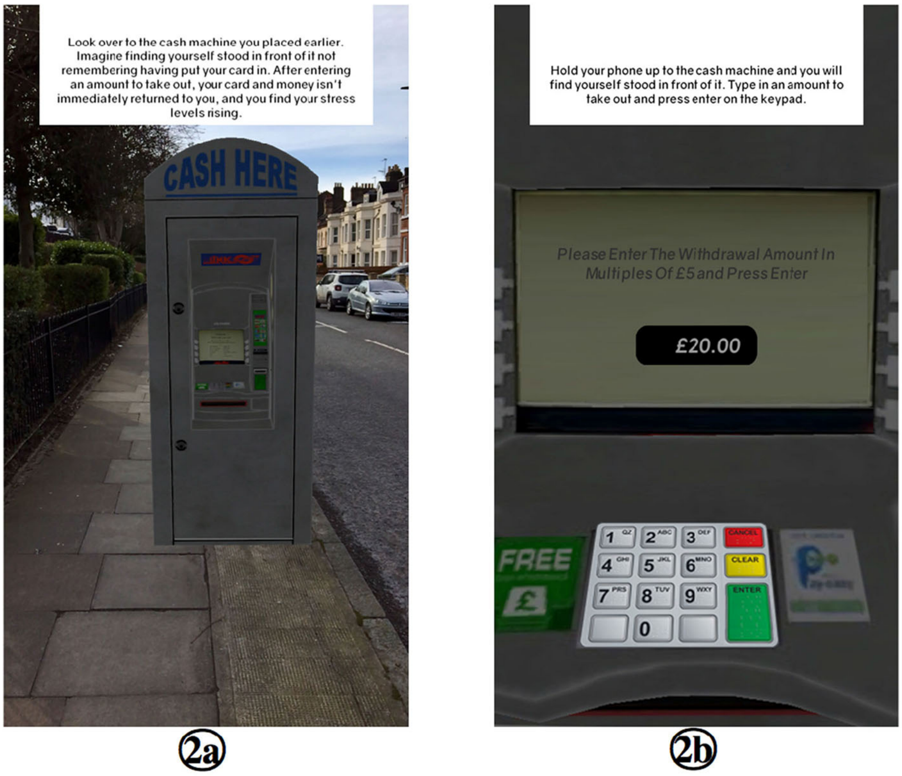

FIGURE 2
App ② gameplay where users were guided to walk to locations identified on a local map, and overlay virtual objects there using AR, at similar positions to those described in the story. In version (2a), users would then be asked to imagine the story events playing out at each augmented location, whilst those in version (2b) were asked to interact with the virtual objects to carry out the same events through their phones.

The app also offered improved tools for accurately positioning each object, with users reporting particularly enjoying this aspect of overlaying each virtual object on their surroundings, to accurately match the object's description described in the story audio or on-screen text.

As part of the user study with the new app, participants were randomly assigned to one of two different versions, differing in terms of the way users interacted with the AR scenes they generated, by placing virtual content on their surroundings at each site visited. In version (2a), the AR scenes acted purely as props to guide the user's imagination, with an audio track guiding them through the story events as they unfolded. In version (2b), users were required to interact with the virtual objects through their phones, actively carrying out the same series of events described in version (2a). This followed feedback to app ①, where some users had expressed a desire to interact with each object, though with questions about whether this would lead to greater immersion in the story. Example gameplay from both versions of the app is shown in Figure 2. A further within-subjects aspect of the study investigated whether participants demonstrated a preference for when the name of the real world location where users placed objects (taken from the map API) was covertly incorporated into the story text. Whether consciously recognizing the use of local place names or not, users expressed a significant preference for these particular sections of the story in both versions of the app that referenced their location, though questions remain whether this might offer a novelty factor that would diminish over successive apps reusing the same mechanic.

Highlighting its role as an exploratory study, the user study with app ② adopted a range of sub-factors from questionnaires in related fields, to rate users' experience in each version of the app. The factors used further highlighted the varying definitions used for particular terms, e.g., the engagement factor of the Temple Presence Inventory (TPI), differing from engagement as Brown and Cairns' lowest level of immersion in the ARI questionnaire, the former more closely associated with a sense of presence (the ARI's highest level of immersion). The questionnaires adopted ultimately proved to be less insightful than the full ARI adopted in app ①, where the gradated scale of three increasing immersion levels suggested how successful the app had been at eliminating barriers to deeper immersion. However, app ② did demonstrate a benefit in adopting a questionnaire factor relating to user's engagement in the story in the form of the Story World Absorption (SWA) questionnaire, rather than purely focusing on immersion related to the way it was delivered through the handheld AR medium, measured with the ARI questionnaire. In terms of the four post questionnaire factors adopted, the two versions of app ② did not score significantly differently, tied to the AR scenes offering stationary props, or interactive objects. However, questions remain whether this would have differed if users were aware of the other interaction opportunity, such that this could be a potential *first-trial* effect. Differences between the two versions of the app were observed in the greater variation of scores in the SWA emotional engagement factor in version ②a, with feedback and users' willingness to continue after each site visited also suggesting some users experienced a deeper experience in this version that prioritized the use of their imagination. However, others found this approach more challenging and said they preferred the more clearly defined interactions of version ②b. As part of the study, the Immersive Tendency Questionnaire (ITQ) was adopted as an alternative to the questionnaire factors previously taken from the Big-5 Personality Inventory. However, no relationship was observed with respect to any of the post-questionnaire factors used to rate participants' immersion using the app.

Feedback to app ② suggested the reduced walking between the story sites and easier to place virtual models, had offered a more immersive experience than app ①. However, not adopting the same questionnaires, meant the ratings for the two apps could not be compared directly. Significant areas for further improvement were still suggested in the perceived lack of reward from having walked to each location, with some users suggesting they would have walked further to reach a local site of greater significance, this was limited by the 250 m radius boundary adopted for this app. Users also expressed a desire that such an experience should finish close to where it started, as well as feeling the use of Text To Speech (TTS) software (here necessary to incorporate local place names into the story script audio in real time) often took them out of the experience. To encourage a greater number of users to participate, a financial reward was offered in addition to listing the study on Prolific. Pilot testing revealed that it was necessary to include checks to prevent users from completing the app without engaging in an outdoor walk to the sites suggested. This involved limiting the feature to trigger the story to continue manually (rather than according to the user's GPS position) to a maximum of two uses. This still allowed for unexpected accessibility issues, such as the displayed map not being up to date, while still not collecting users' exact GPS path taken, which would have presented a privacy issue.

4.2. Follow-up user studies

4.2.1. App ③

The first two apps suggested some users enjoyed the use of AR to tell a story set around their local neighborhood, supporting the research direction taken, though with outstanding issues around the most effective way to implement AR as part of an engaging story. This led to the next app prioritizing a focus on different opportunities for interaction with the virtual content and temporarily setting aside the use of local map data. App ③ also aimed to offer a more standardized testing procedure by re-introducing the ARI questionnaire used alongside app ①, while comparing it to the ITC-Sense Of Presence Inventory (ITC-SOPI) that aims to measure users' quality of experience in terms of the presence experienced. With these aims, app ③ offered a home experience that used Unity's ARFoundation tools, to overlay a single virtual environment registered to the user's surroundings using SLAM tracking. They could navigate around the space through their physical movement, interacting with it using their phones, this now being more of a handheld VR experience. The story involved the user attempting to find a way out of a locked room and discover how they got there. Two versions of the app offered varying user agency, with version ③a offering a linear guided experience with the various objects in the room, and version ③b letting users pick which object to interact with next. Offering increased agency aimed to offer a greater reward that some users felt was lacking in the previous apps. Each virtual object in the room would also offer interaction in one of three ways as shown in Figure 3, to further investigate how these affected users' immersion in the story. Alongside the virtual objects acting purely as props, and the touchscreen interactions offered in app ②b, natural hand gesture interactions were also offered, this feature limiting participants taking part to those with an iOS phone containing a depth camera.

This phone requirement was made clear in the recruitment instructions, but pilot testing of the app revealed some users still attempted to download the app and take part without a suitable device, particularly those taking part through Prolific. The app was subsequently modified to record the model of iPhone used to prevent this issue from occurring, though based on contacting some of these individuals, it highlighted some users attempts to take part in a study that sounded interesting without fully reading the instructions, particularly when a financial reward was offered. This also highlighted a lack of features when creating a new study on Prolific's editor, to prevent users from continuing that did not satisfy the requirements. The app also suffered the inability to check that users attempted to place the virtual environment somewhere of sufficient size and with limited clutter, which may otherwise have affected their experience. Based on previous feedback that answering excess questions at each story site inadvertently affected users' immersion in the ongoing story, app ③ limited these questions to just each object interacted with in terms of its enjoyability and ease of use, as well as rating users' desire to continue the story, which has been suggested as a key indication of their enjoyment in related narrative video game research (Schoenau-Fog et al., 2013).

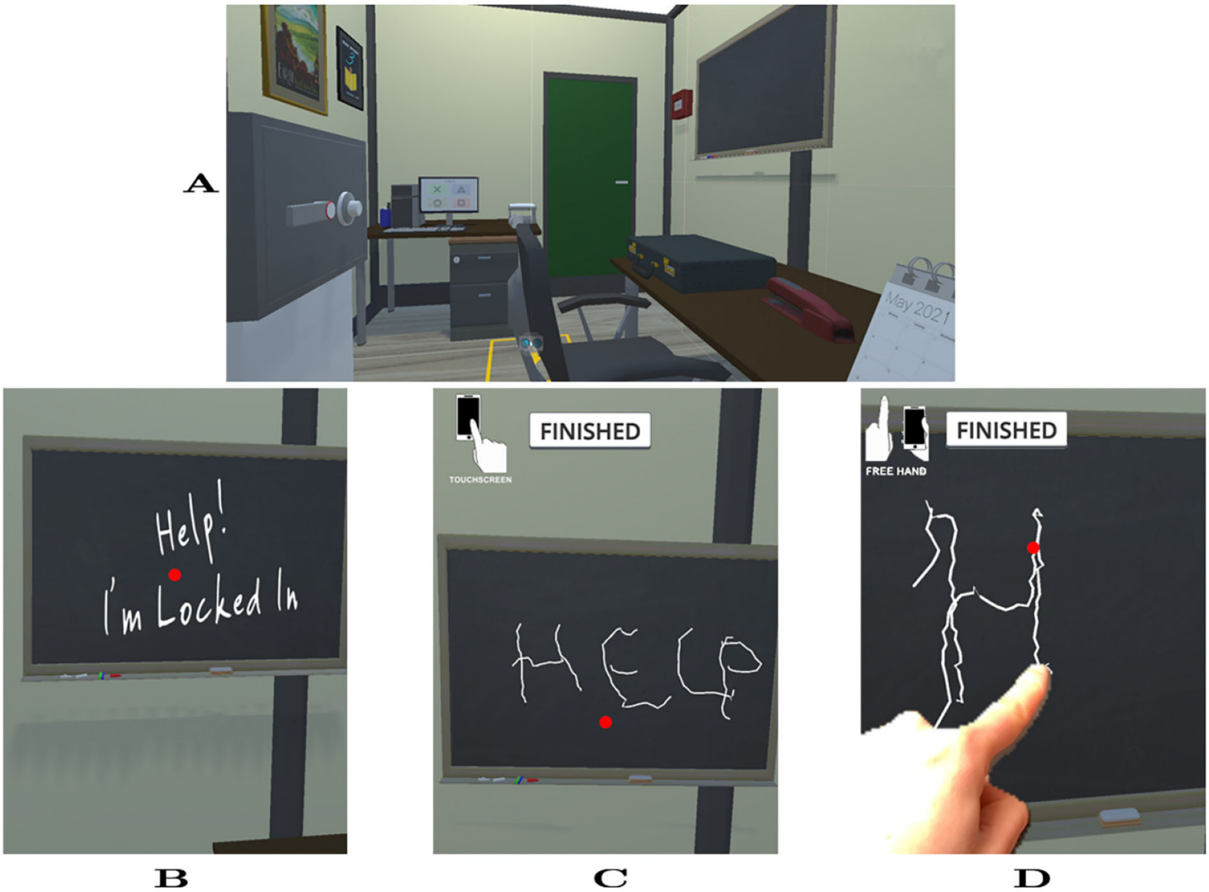

FIGURE 3
App ③ gameplay where users were asked to overlay a virtual environment in their homes in a suitably sized space using AR SLAM technology **(A)**. The virtual room contained a variety of objects to interact with in three different ways, either asking users to imagine the events described **(B)**, interacting with the objects on the phone's touchscreen **(C)**, or detecting their free hand movement **(D)**. The app was also released in two different versions, (3a) and (3b), varying the agency afforded to users.

Comparing the two versions of the app that varied the level of agency offered, the lack of agency in version (3a) was reported as frustration for users who wanted to pick the next object to interact with, having theorized a connection between particular objects in the room. However, version (3b) highlighted the need to keep such choices relatively simple, with some users reporting feeling stuck when they were required to interact with the objects in a particular order, even when clues appeared on a TV screen in the virtual room to help them. Feedback to the two versions appeared to demonstrate that a reduced cognitive load from having to decide what object to interact with next, did lead to more thoughts about the story and how different objects connected. Ultimately there was not a significant difference between the two versions according to users' ratings through the ARI and ITC-SOPI factors completed on reaching the end of the app. One significant variation observed was in terms of the different types of interaction possible with each object, whether with the objects acting as stationary props, interacting with them through the touchscreen, or a gesture based interaction. The most common theme of feedback left related to the gesture hand detection interactions, which users felt were enjoyable but had significant usability issues. Conversely, being asked to imagine the events was rated least popular, now that users were aware they could be interacting with the objects, supporting the first-trial effect why this perhaps did not prove an issue in app (2a). This also highlights both a potential novelty and on-boarding issue with the free hand interactions which is why they were not pursued in later studies, despite offering the potential for a more embodied and immersive experience. Despite the first chapter of the story being set aside for on-boarding of the different interaction opportunities, the unfamiliarity of the gesture controls required greater on-boarding for several users who took part, this requirement varying according to the particular user and so challenging to implement as part of a remote study.

Readopting the ARI questionnaire for app ③ demonstrated that the Brown and Cairns immersion levels of engrossment and total immersion were rated significantly higher for this latest app than app ①, which also used the ARI. However, each successive immersion level was still scored significantly lower in app ③, raising about questions whether the highest level associated with presence and flow could be significantly achieved in such an experience, where the user is required to split their attention between the virtual world while also navigating obstacles in the real world. Subsequent apps would be required to investigate whether the improved immersion experienced by users, was primarily related to taking part without engaging in a walk so reducing some feelings of self-consciousness, the primarily virtual story world, or the variety of interactions now offered with the virtual content. Adopting the ITC-SOPI engagement factor once again demonstrated this measure as closely replicating the measure of presence also measured in the ARI, and differing from the ARI definition of engagement. The factors of the ITC-SOPI not adopted here like perceived realness of the environment and sense of physical space, further suggest the questionnaire is more appropriate to the study of head-mounted VR worlds, where the user's sense of presence may be seen as the most significant factor. This further supported the ongoing use of the ARI questionnaire supported by appropriate questions related to users' story engagement, to cover both these aspects of the user experience. The tendency to get an involved factor of the ITQ was also re-tested as part of this latest study and showed evidence of correlating with some of the ARI factors. Further study is required to investigate if the ITQ also shows some relationship with a walking AR experience, given that app ③ offered something more akin to a handheld VR experience, with the ITQ having been validated for VR in relation to a user's sense of presence experienced (Witmer and Singer, 1998).

4.2.2. App ④

App ④ aimed to incorporate some of the features that led to a more immersive experience in the previous app, such as more complex AR scenes that combined a variety of virtual objects to encourage active imagination, whilst also linking more closely with later scenes, and offering a variety of different interaction opportunities. The latest app continued the path taken by app ③ in terms of a story that could be experienced from home, in addition to comparing this with a version involving walking outdoors to identified sites based on the local map information. Three different versions of the app were ultimately tested, where the user took on the role of an investigator trying to solve a local murder. In version ④a, users selected from the story sites marked on a local map to physically walk to. Versions ④b and ④c adopted the same local map, but users took part from a suitably sized space in their home or garden, instead directing a computer controlled agent to traverse the map to the map location they selected. In version ④b users overlaid virtual objects on top of their home surroundings through AR, investigating whether the disconnect between the virtual objects and users' surroundings significantly affected their immersion. Version ④c offered the addition of a virtual background to overlay the virtual items against, the background displaying a photographic panorama of the local real world site taken from the Google StreetView API. Gameplay in the three different versions is shown in Figure 4. Alongside a variety of interactions with each virtual object placed, in order to investigate users' preferences, users built up a personalized clue board as they progressed through the story, including any photos they took, alongside the names of the real world places where clues were discovered. Similar to app ②, the Mapbox API was used to display a local map and suggested walking route for the user or virtual agent to follow, with the Google Places APIs used to identify the local sites, sometimes fine tuning the selected story site locations, such as through Google's Nearest Road API when an object was found on the street as part of the story.

In addition to the ARI questionnaire, app ④ adopted the full Narrative Transportation scale (NTS), to offer a measure of users' transportation to the story world whilst taking part (Green and Brock, 2000). This questionnaire has previously been applied to location-based video narratives (Karapanos et al., 2012), and was found to be more appropriate to apply in its entirety without significant modifications, compared to the SWA questionnaire. However, Table 3 demonstrates that the NTS factor relating to story immersion recorded a low internal consistency score according to Cronbach's alpha. This can be understood in terms of the questions comprising this questionnaire factor, which suggest story immersion comes from becoming less aware of one's surroundings, which makes sense in the case of a literary story, but will not be the case in such a handheld AR experience. The walking version of the new app, ④a was found to offer significantly higher immersion than both the home versions, across the majority of the NTS and ARI factors collected. User feedback suggested a benefit came from witnessing public activity, as they walked between the different story sites. This better supported users' immersion, as well as offering greater opportunities for *magic moments* and positioning each virtual object in an appropriate location, which could not easily be achieved in versions ④b and ④c. Both home versions rated similarly according to the post questionnaire factors, with the still backgrounds offered in version ④c commonly not even being recognized as locations close-by the user. Additionally, users' immersion was affected by the photos' low resolution and blurred details to protect the privacy of those captured in them, with it also proving difficult to accurately align the 3D models against the static images, compared to the physical locations in version ④a.

In all versions of app ④, the more interactive tasks with the virtual content proved most popular, providing further evidence that when users are aware of the opportunity for interacting with the virtual elements, they will choose this option over the AR scenes just acting as props to guide their imagination. The ITQ was adopted as part of this latest study and once again showed evidence of a relationship with the three immersion

FIGURE 4
App ④ gameplay was tested in three versions. Users walked outdoors to nearby locations where they placed and interacted with AR scenes in version (4a). Alternatively, a virtual agent moved around the map in their place, with version (4b) asking users to overlay the virtual items against their fixed home or garden surroundings, while version (4c) offered virtual backgrounds of each map location to overlay the virtual items against, taken from Google StreetView.

levels of the ARI, with a similar correlation also found between the ARI levels and users' rated enjoyment of walking. However, this result was partly affected by the limited number of users who reported not enjoying walking that took part in version (4a), though this did support the idea that such users are less likely to complete such an experience if they do not enjoy walking for recreation. As an alternative for such users, the mobile VR experience of app ③ would potentially appeal more, given the barriers posed to immersion in versions (4b) and (4c), where the users' surroundings did not support the virtual objects placed as part of a handheld AR experience. Alternatively, further research may look at selecting closer story sites from the map if such users are identified, though this will inevitably vary the amount of story content required when users are walking between locations. In this respect, app ④ used voice recordings rather than TTS audio to deliver a hands free version of the story text, which potentially also resulted in an additional reduced barrier to deeper immersion. Readopting the ARI questionnaire also allowed the ratings for each version of the app to be compared to previous apps adopting the same questionnaire. This is shown in Figure 6 and reveals that app (4a) scored similarly to both versions of app ③ in terms of the ARI immersion levels of engagement and engrossment, only scoring lower in respect of the highest level of total immersion. The subsequent app would examine the extent to which this is the result of using AR, as well as where some features of the real environment did not necessarily fit with the story.

4.3. Conducting a curated user study

4.3.1. App ⑤

The previous apps had all been tested remotely at a variety of participant locations. In spite of a number of confounding variables that this could give rise to, conducting these studies under the conditions such a final product was intended for use, did allow improvements to be made based on user feedback, and comparing ratings to each app through the ARI questionnaire. The previous app (4a) incorporating an outdoor walk between real world locations, was reworked into a new app and story designed for use at a single pre-determined location, that of the Queen Mary University of London (QMUL) campus. This allowed a more controlled experimented using the new app to be conducted. The use of a location known in advance allowed some further changes to be made, such as the audio tailored to fit the walk between particular sites, fitting the walking distance and common occurrences observed along the route, as well as selecting virtual models that were more aligned with the physical surroundings. This curated study offered all participants the

TABLE 3 Internal consistency using Cronbach's alpha, of each of the study measures adopted across all 3 versions of app ④, and the single version of app ⑤.

Questionnaire factor	App code and associated user study			
	④		⑤	
	Scale reliability	Mean score and sd	Scale reliability	Mean score and sd
ITQ - Tendency to maintain focus (6 items)	$\alpha = 0.69$	$M = 5.24$, SD = 0.79	$\alpha = 0.63$	$M = 5.24$, SD = 0.81
ITQ - Tendency to become involved (5 items)	$\alpha = 0.56$	$M = 4.78$, SD = 0.89	$\alpha = 0.63$	$M = 4.80$, SD = 0.93
ITQ - Tendency for emotional engagement (4 items)	$\alpha = 0.66$	$M = 5.01$, SD = 1.09	$\alpha = 0.50$	$M = 4.91$, SD = 0.94
NTS - Level of emotional Involvement (5 items)	$\alpha = 0.71$	$M = 3.32$, SD = 1.12	$\alpha = 0.58$	$M = 4.67$, SD = 0.89
NTS - Level of story immersion (3 items)	$\alpha = 0.27$	$M = 3.99$, SD = 1.06	$\alpha = 0.33$	$M = 3.71$, SD = 1.03
NTS - Level of mental immersion (2 items)	$\alpha = 0.84$	$M = 4.50$, SD = 1.60	$\alpha = 0.81$	$M = 5.57$, SD = 1.12
ARI - Interest rating (4 items)	$\alpha = 0.83$	$M = 5.24$, SD = 1.29	$\alpha = 0.92$	$M = 6.01$, SD = 0.95
ARI - Usability rating (1 item)	$\alpha =$ N/A	$M = 4.97$, SD = 1.69	$\alpha =$ N/A	$M = 5.37$, SD = 1.39
ARI - Emotional attachment rating (3 items)	$\alpha = 0.85$	$M = 4.22$, SD = 1.49	$\alpha = 0.75$	$M = 5.67$, SD = 1.05
ARI - Focused attention rating (3 items)	$\alpha = 0.77$	$M = 4.40$, SD = 1.37	$\alpha = 0.61$	$M = 5.50$, SD = 0.96
ARI - Presence rating (4 items)	$\alpha = 0.90$	$M = 3.25$, SD = 1.55	$\alpha = 0.77$	$M = 4.95$, SD = 1.22
ARI - Flow rating (3 items)	$\alpha = 0.69$	$M = 3.85$, SD = 1.29	$\alpha = 0.76$	$M = 4.81$, SD = 1.18

Here the ARI usability factor was limited to a single question given the additional questions about the app usability answered at each story site visited.

use of a single mobile device, with only a couple of people expressing a preference for using their personal phones that they felt more comfortable with. Some users appreciated the opportunity to ask questions before starting the study given a researcher was present, this appeared to aid on-boarding for users who were unfamiliar with the AR technology, and put them at ease. The same validated questionnaires were adopted from the previous study and completed within the app, alongside the same bespoke questions rated after the AR interaction at each story site, concerning its enjoyability, usability, and users' desire to continue. The presence of a researcher also allowed more detailed feedback to be collected after participants had completed the app, though there was a risk that they may have felt less inclined to leave negative comments when providing face to face feedback in this way. Example gameplay from app ⑤ is shown in Figure 5.

Table 3 shows a comparison of the questionnaires' internal consistency scores using Cronbach's alpha, for apps ④ and ⑤, which demonstrates the similarity between the remote and curated data collection measures in this regard. It again highlights the issue with the story immersion factor of the NTS, related to the difference in the way users' surroundings contribute to their experience of the story compared to a literary narrative. Developing a new questionnaire would require a new factor analysis to validate it for such experiences, based on the issues identified in adopting immersion and engagement questionnaires from other fields of study. Additional data in the study with app ⑤, came from the opportunity to observe some participants taking part. This highlighted differences in some individuals' ease of navigating using the on-screen map of the campus. Some users walked a significant distance before realizing they were heading in the wrong direction, at which point they were out of sync with the recorded audio track, which directly referenced the surroundings they should have been seeing. The use of a known location also allowed virtual objects to be used that more closely matched the surroundings, whilst also increasing the possibility of magic moments, by referencing common activities observed at each story site. These differences

10.3389/fcomp.2022.927177

FIGURE 5
App ⑤ gameplay offers a site-specific handheld AR experience adopting many of the same elements from app (4a), but designed for use only on the QMUL campus. All participants experienced the same app, which also tested a variety of interactions with different sizes and complexity of virtual objects overlaid on users' surroundings, to further investigate which rated more enjoyable and easy to use.

from app (4a), which had been designed for use at a variety of locations, contributed to the higher NTS emotional involvement factor rating, with feedback expressing that users were excited to overlay virtual content where it directly aligned with their surroundings, even when there was little subsequent interaction offered with the virtual objects.

Another effective use of AR in app ⑤ was when users entered a virtual cabin full of objects they could interact with, similar to the virtual room environment offered in app ③, though this time offering a view out of the cabin windows onto the real world environment around them. Feedback in the previous AR apps highlighted the need to make any virtual content placed using AR detailed enough that it both offered a reward for the walk involved to reach it, while also allowing it to be examined or interacted with in a way that encouraged the user to build new meaning in respect of the story. The cabin full of other virtual objects to inspect appeared to successfully offer this more than some of the lone objects placed. Several participants with only limited knowledge of the campus location before using the app ⑤, left feedback regarding their appreciation of exploring the location and being shown hidden areas they had not seen before. This connects to some of the feedback from the remote studies, where users expressed a desire to see parts of their local neighborhood they were not familiar with, particularly if data regarding their historical or social significance could be captured and included in the story in some way.

Users' ARI and NTS factor scores for app ⑤ and the previous studies adopting the same questionnaires are shown in Figures 6, 7. The data was non-parametric in most cases, influencing the choice of statistical tests used. A Kruskal-Wallis test investigated the difference between the ratings for each app. Only the NTS story immersion factor did not vary significantly between the versions of app ④ and app ⑤, with the caveat mentioned earlier that the internal consistency of this factor was too low to offer any meaningful information. NTS mental immersion appeared to vary significantly ($\chi^2 = 12.9$, $\boldsymbol{p} < 0.01$), with the apps intended for use at home, (4b) and (4c), scoring significantly lower than both version (4a) (r = 0.26) and the site-specific app ⑤ (r = 0.36). NTS emotional involvement also showed a difference (χ^2 = 34.7, $\boldsymbol{p} < 0.001$), with app ⑤ scoring significantly higher than both the walking version (4a) (r = 0.37), as well as the home versions of the previous app (r = 0.50 - (4b) and r = 0.65 - (4c)). Performing a similar analysis for the ARI immersion levels, app ⑤ rated highest for engagement alongside the home handheld VR story offered in app ③ (χ^2 = 14.3, $\boldsymbol{p}$ = 0.01). In terms of the next level of immersion namely engrossment, apps ③, (4a) and ⑤, all scored significantly higher than the other apps (χ^2 =

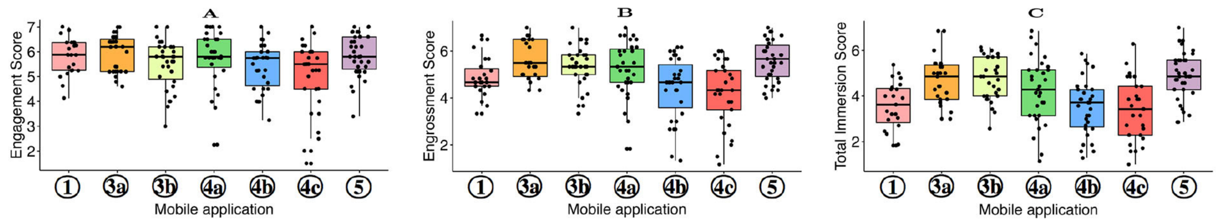

FIGURE 6
User ratings for the six ARI questionnaire factors grouped in pairs according to Brown and Cairns three levels of increasing immersion, namely engagement **(A)**, engrossment **(B)**, and total immersion **(C)**. The scores are shown for all the app's tested that measured user immersion using the ARI.

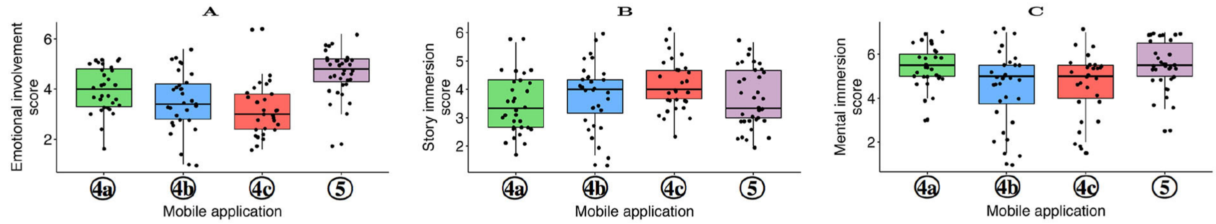

FIGURE 7
User ratings for the three NTS factors for the different versions of apps ④ and ⑤ that adopted it to rate users' engagement with the story. The three factors of the NTS relate to a user's level of emotional engagement **(A)**, their immersion in the story **(B)**, and their level of mental immersion in the events **(C)**.

33.2, $p < 0.001$). For the highest immersion level of ARI total immersion, apps ③ and ⑤ rated similarly, but a significant difference ($\chi^2 = 36.7$, $p < 0.001$) was found with respect to all three versions of app ④ though with varying effect sizes ($r = 0.24$ - (4a), $r = 0.51$ - (4b) and $r = 0.54$ - (4c)). Apps (4a) and ⑤ also demonstrated evidence of a correlation between users' ITQ scores, and their immersion in the apps rated according to Brown and Cairns' three immersion levels using the ARI. The ARI immersion scores similarly showed evidence of correlating with users' appreciation of walking for recreation as rated in the pre-questionnaire. The difference in immersion scores between apps (4a) and ⑤, therefore, only differed significantly for the ARI questionnaire's total immersion level, observed with a small effect size, as well as the NTS emotional involvement score, observed with a medium effect size. This suggests that relating the story to the features of a known location in app ⑤, was most effective in enhancing users' sense of emotional involvement, presence, and flow associated with these factors.

5. Discussion

This section provides a discussion of the lessons learned from adopting a Design Research Method, to create an immersive story experience that could be used at a variety of locations. This was done through the development of several user studies that each tested a new app, revising the testing procedure adopted in each case. This research direction was partly informed by the pandemic restrictions in place at the time, though the easing of these restrictions later allowed the results to be compared against a similar app designed for a known location. Testing the five apps created as part of this research also gave rise to new research directions that have the potential to further improve the user experience.

5.1. Performing robust remote studies and collecting sufficient quality data

- *Issues arising around the early stage testing of a new user experience.*

  The first two preliminary versions of the app suffered from several significant issues in both the app design and their testing, which limited users' ability to deeply engage with them. This included excess walking between the selected story sites as a result of the chosen map API, such that the audio tracks were commonly too short for the walk involved. Some participants also reported feeling that the excess questions asked after each site visit took them out of the story. Some of these issues could have been revealed by more extensive pilot testing, including a direct comparison of a range of map APIs across a greater variety of locations. Some of these issues also highlighted the benefit of conducting remote studies, since such an issue would not necessarily become evident from a study conducted at a single location. These early studies also matched prior research in highlighting the lack of validated questionnaires specifically suited to such AR experiences (Kim et al., 2018), with a benefit also observed in reusing the same tools across subsequent iterations, such that the ratings could be directly compared as was later achieved using the ARI questionnaire.

- *Remote studies benefited from comparing different versions of the same app.*

  Following the study with app ①, all subsequent remote studies used a between subjects approach to compare several versions of each app, with users randomly assigned to each version. This method of direct comparison offered benefits for such remote studies, where the ability to collect sufficient data about all aspects of the user experience is more challenging, and did not allow the level of detail to be collected that would typically form part of a Design Research Methodology (Blessing et al., 2009). Most of the user studies conducted adopted a mixed methods study approach that was successful in allowing several aspects of each app to be tested concurrently, though the limited sample size testing each version made it hard to make conclusions in some cases, due to the limited number of users in certain test sub groups.
- *The use of remote and curated studies to support each other.*

  Central to the testing of such an app designed for use at the user's choice of location, was the observation that both remote and curated studies could be used in conjunction with each other to offer useful insights. Agreeing with prior research, the reliance on questionnaires and feedback alone as part of a remote user study will struggle to offer a suitably detailed picture of the user experience, given the various issues that could be missed or not reported (Lehman et al., 2020). The curated study with app ⑤ produced more detailed feedback through participant conversations with a researcher after the experience, as well as revealing additional issues when participants could be observed taking part. However, the remote studies did offer important feedback regarding how the app functioned across a variety of locations that could be challenging to implement later following purely lab based experiments, this offering a reverse methodology to most AR studies which tend to be curated (Dey et al., 2018). This approach also enabled the more refined curated experience of app ⑤ to highlight the specific benefits of a site-specific experience, which further research could aim to generalize to one made for use at a variety of locations.
- *Remote studies offer a wider reach.*

  Making each remote study available on a study recruitment website like Prolific and social media did make it available to a wider range of demographics, which agrees with similar research published in this regard (Ratcliffe et al., 2021). However, there was still less uptake from older users and those with significant previous AR experience, with it also proving challenging to recruit those who did not necessarily enjoy walking for recreation to take part in such an outdoor study. The curated study data from app ⑤ did not appear to differ significantly from that collected in the remote study with app ④, agreeing with earlier findings as part of a VR experience, that remote and curated study data would largely agree provided hardware variations were minimized and participants pre-screened (Zhao et al., 2021). Social media recruitment proved more challenging than on a site like Prolific, with some users recruited in this way more expectant of testing a polished product, with some also turned off by the request to complete questionnaires that they felt interrupted their experience while taking part.
- *Constraints to make sure remote studies are completed appropriately.*

  Recruiting on Prolific demonstrated the need to implement additional checks to make sure that instructions had been followed and each study completed as intended, as far as this was possible. User feedback suggested this was partly in response to the offer of reimbursement for taking part in a study that sounded interesting and had limited places available. Extra checks performed involved collecting users' model of phone used, and limiting the number of times that a story site was triggered without being at the exact GPS location, this feature implemented to still allow for map errors or unexpected real world issues, while was making sure users had engaged in an outdoor walk. Prolific would benefit from making some changes in this respect to the way studies are presented to its users, so that any bespoke requirements can be more clearly displayed on the listing, or requiring participants to confirm they satisfy each requirement before being able to enlist. Such checks would also benefit from being part of the app itself, given the additional cognitive loads that may be offered by a user's location that is not a suitable size or too cluttered (Shin et al., 2019).
- *Offering appropriate levels of on-boarding.*

  Conducting the curated study with app ⑤ provided greater insight into some users' unease about adopting handheld AR, based on their desire to ask several questions to the researcher before starting the study. This was not possible in the remote studies were instructions were provided when users signed up, as well as through in-app instructions, and on-boarding. This aimed to be implemented seamlessly within the first chapter of the story, often acting as a tutorial involving the guided placing and interaction with one or more virtual objects through AR. This was done most openly in app ③, where the first chapter guided users through interaction in each of the three different interaction types. However, the same app further highlighted the challenge in offering sufficient on-boarding for a wide range of users, where several individuals reported failing to become comfortable with the hand gesture interactions after only this tutorial. This contributed to such gestures not being pursued in later studies, with related research also highlighting the challenges in accurately identifying gestures most users will feel comfortable with, alongside the potential for such embodied interaction to lead to a more immersive experience (Slater et al., 1998; Gračanin, 2018). Such effects as part of a remote study also risk the prevalence of a novelty factor, and users being less open to engaging more deeply with a story (Raybourn et al., 2019).

- *Precautions to protect participants' safety and privacy.*

  Alongside the careful design of the on-boarding phase, it was necessary to include reminders that users take care when using AR in public, due to the various risks posed by roads and pedestrians, which have also been highlighted in related literature (Serino et al., 2016). There was evidence that the offer of a financial reward for taking part sometimes encouraged users to attempt it even in poor weather conditions, with one participant reporting their concern at being asked to follow an uphill route on an icy day. Warnings for such unexpected occurrences need to be included, with the added effect that low light levels would inadvertently affect the quality of user experience, tied to reduced SLAM AR tracking performance. Recording the time of day a study is completed therefore offered an additional statistic in deciding whether a user's data was retained. Some users also expressed concern about the collection of their personal data, with it proving important to classify that their exact GPS position was not recorded as part of any of the studies.

- *Missing important issues in remote studies.*

  The use of remote studies was effective in highlighting the significant issues with each app design, such as those related to excess walking and difficulty to place virtual models in app ①. However, these findings highlighted that most participants only chose to report a couple of the most significant issues experienced when submitting their feedback. In this respect, participants through the Prolific portal tended to submit less detailed feedback than those recruited through university mailing lists and those more aware of academic research. One possible feature to trial would be the use of an onscreen button to report details of an issue when one arose, similar to the button that allowed users to trigger the story to continue manually, which asked them for a short description of why they could not access the intended GPS location when it was used.

5.2. Offering users an immersive handheld AR outdoor story experience

- *Selecting appropriate virtual objects and placing them in the real world.*

  The curated study with app ⑤ demonstrated users' appreciation when the virtual objects overlaid aligned with users' surroundings, even above offering a significant interaction with them. This aligns with various previous location-based experiences, that suggested greater immersion and mental imagery occurred when a story was experienced at a location with at least a matching atmosphere (Karapanos et al., 2012; Packer et al., 2017). Currently, this is much more challenging to achieve in an app that can be used at a variety of locations, though there are potential areas for further research in this regard. Approaches could be to ask users to select the most appropriate 3D model to place from a range of differently styled versions of the same object. Machine learning imaging techniques may also offer the potential to identify relevant features of a location through the camera view, which could then be used to present appropriate virtual content according to the surroundings. A related approach has also been seen in TransformMR, to replace real world objects with virtual ones according to the physical object's pose and movement, the virtual object also in-line with the experience's theme (Kari et al., 2021).

- *Placing and interacting with virtual content.*

  The use of large virtual objects overlaid on users' surroundings caused a barrier to immersion in app ①, from being hard to place accurately and interact with, due to significant occlusion and SLAM tracking errors as users moved closer to them. This arose from them needing to be placed at a significant distance from the user, to fully observe them within the limited phone camera field of view. Conversely, users appreciated the task of finding suitable positions for smaller sized models, that matched their described location in the story. In this way, users were guided to generate meaning from each object, relating back to ideas in Merleau-Ponty's phenomenology of perception (Merleau-Ponty and Landes, 2012). Generating meaning can also be found in creating different opportunities for pretend play that has been linked to the choice of story, theming, framing, and opportunities for role-play (Deterding, 2016). The most effective use of virtual objects reported with each app was either when they were closely aligned with the user's surroundings or were sufficiently detailed to allow their inspection which could lead to new connections being made with the story. This was particularly effective in the case of placing a virtual environment users could move inside of, which allowed the user to explore another section of the story world as in apps ③ and ⑤.

- *Accounting for different users.*

  Different users' levels of engagement and immersion appeared to show evidence of a connection to their immersive tendency measured through the ITQ, as well as users' predilection toward walking. Larger sample size would be required to examine such relationships in greater detail. App ⑤ also revealed users' varying abilities to navigate easily using a map, which caused some users to miss parts of the story as they took time to find their bearings. These are all areas for improvement in order to offer deeper immersion for a wider range of users, though questions remain over the degree to which those who do not enjoy walking would appreciate a story told in this way. The alternative approach suggested by app ③ was that such users may prefer to take part from home with the story instead told through handheld VR, given the issues caused in app ④ when the virtual objects appeared out of place alongside the users' surroundings. Further possibilities to test include reducing the area within which the story sites are selected for these users, aiding navigation through the use of AR markers, a photo-realistic map, the option to replay audio once a user knows the direction to move, or delivering the audio in parts according to their GPS position as they trigger points along the correct path.
- *The benefits of using audio to aid immersion.*

  Users reported that the use of audio including spatial sound effects was of particular benefit in blending the story world alongside the real world, particularly when the audio directly referenced real world features in app ⑤, or generated a sense of ambiguity from closely matching unrelated events happening around them. The effective use of audio has been shown to aid immersion in various digital media experiences (Ekman et al., 2005; Dyson, 2009). Limits to immersion also came from the use of TTS in the first few apps, which is necessary for the story text to be modified in real-time to include local place names. Users also desired that the audio be crafted to last the length of the walk to the next story site, acting as an additional indication they had arrived, as well as providing a more continuous experience. Offering this when the distance between locations varies is a significant area for further research, with the additional desire that map landmarks detected along the user's walking route, could also be incorporated into the story events described as the user passes them. This connects to the research fields of procedural text generation and Natural Language Processing (NLP).
- *Offering opportunities for magic moments to occur.*

  The occurrence of magic moments where users perceived the story as synchronizing with their surroundings, formed significant moments of high immersion for those users who experienced them. Such moments were significantly easier to generate at the known location of app ⑤, matching research suggesting such moments can be feigned if enough is known about a location (Reid, 2008). Several participants reported being willing to accept the inclusion of lots of random details, to raise the probability such moments would occur. Again image recognition techniques through a user's camera might offer an improved way to generate such moments based on recognizing local features or objects.

5.3. Additional areas for further study

- *Developing appropriate research tools for handheld AR experiences.*

  Testing a variety of immersion questionnaires from related fields highlighted the need for questionnaires more appropriate to such experiences, due to the varying internal consistency of some of the factors, and others that did not make sense applied to such an experience. This is supported by a systematic review of current research in the area of AR, which found it lags behind the tools available for VR (Kim et al., 2018). While the ARI questionnaire offered a measure of immersion adapted from Brown and Cairns' immersion model, questions remain about whether terms like flow should be modified in relation to a pervasive AR experience (Jegers and Wiberg, 2006). Appropriate tools would also benefit by having a measure of users' transportation by the story, in addition to the immersion offered through the mediating AR device, since the two operate in tandem in such an experience. The NTS was found to be limited in its current form designed for use primarily with literary stories. Further research would also account for potential variations between different users' experiences, that could be used to better personalize an experience. Some evidence toward this was suggested here by the ITQ (a questionnaire only currently validated for VR), with prior work having looked at personalising narrative video games according to the player (de Lima et al., 2018).
- *Improving the immersion offered with such an AR storytelling app.*

  All five apps failed to offer a high level of total immersion as rated by the ARI and connected with a user experiencing both a sense of presence and flow while taking part. While questions remain over the effectiveness of these questionnaire factors for such an experience, user feedback highlighted several areas where the apps could have been further improved. These include offering more embodied forms of interaction, which users found enjoyable in app ③ but was not continued due to the greater on-boarding requirements. The benefits of multimodal interaction including haptics form much ongoing research, in terms of the potential benefits of offering an interaction that more closely matches reality

(Hecht et al., 2005). The current apps were also limited by a linear story, with several users expressing a desire for a greater agency for their actions to influence the events, this relates back to Murray's suggestion that any new storytelling medium must successfully offer a balance of immersion, agency and user transformation (Murray, 1998). Additionally, there are several approaches that have the potential to deliver virtual objects more appropriately for users' surroundings, including new techniques to match the lighting of composited virtual content to that of an environment's ambient lighting, which could deepen users' immersion in the AR world, through the objects feeling more like they truly exist (LeGendre et al., 2019).

6. Conclusion

A series of remotely tested mobile applications were studied as part of a Design Research Methodology, in order to offer users a handheld AR immersive story, that they could use at their own choice of outdoor location during ongoing pandemic restrictions. The results were compared against a curated site-specific experience, contributing to an overall reverse methodology to much work in the area, and suggesting ways that remote and curated studies could support each other when developing an experience suitable for a variety of locations. The research considered various aspects to ensure that any remotely collected data is admissible, given various ways a user could deviate from the study guidelines, or be influenced by external confounding effects. Suggested further directions are also outlined based on potentially reducing the barriers to deeper immersion for a wider range of users, through improved on-boarding and user navigation, as well as better relating virtual content and the story to the users' surroundings. Conducting both remote and curated AR studies has the potential to better account for the limitations each study approach offers when used in isolation. However, better evaluation tools for such experiences require development, due to the lack of current tools available and challenges in adopting questionnaires from related fields. The results presented aim to support ongoing research to offer new AR story experiences that can be experienced anywhere, whilst also reducing the current challenges for designers to transpose site-specific experiences to a new location, as well as encouraging new audiences who may prefer to take part closer to home.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by QMUL Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

GR oversaw the design and user study with each mobile app outlined with support and direction from both MW and LT. A similar process was adopted for the journal article, which was initially written by GR and revisions made according to feedback from MW and LT. All authors contributed to the article, its revisions, and approving the submitted version.

Funding

The research formed part of the Centre for Doctoral Training in Media and Arts Technology at QMUL, funded by the PSRC and AHRC.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Baker, C. (2017). "Virtual, artificial and mixed reality: new frontiers in performance," in *2017 23rd International Conference on Virtual System Multimedia (VSMM)* (Dublin: IEEE), 1–10.

Ballagas, R. A., Kratz, S. G., Borchers, J., Yu, E., Walz, S. P., Fuhr, C. O., et al. (2007). "REXplorer: a mobile, pervasive spell-casting game for tourists," in *CHI '07 Extended Abstracts on Human Factors in Computing Systems-CHI '07* (San Jose, CA: ACM Press), 1929.

Benford, S., Giannachi, G., Koleva, B., and Rodden, T. (2009). "From interaction to trajectories: designing coherent journeys through user experiences," in *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems (Boston, MA), 709–718.

Blessing, L. T. M., Chakrabarti, A., and Blessing, L. T. M. (2009). *DRM, A Design Research Methodology*. Dordrecht; London: Springer.

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol*. 3, 77–101. doi: 10.1191/1478088706qp063oa

Brooks, K. (2003). *There is Nothing Virtual About Immersion: Narrative Immersion for VR and Other Interfaces*. Technical report, Human Interface Labs, Motorola Labs.

Brown, E., and Cairns, P. (2004). "A grounded investigation of game immersion," in *Extended Abstracts of the 2004 Conference on Human Factors and Computing Systems-CHI '04* (Vienna: ACM Press), 1297–1300.

Bryon, J. (2012). Tour guides as storytellers-from selling to sharing. *Scand. J. Hospit. Tourism* 12, 27–43. doi: 10.1080/15022250.2012.656922

Dansey, N. (2008). "Facilitating apophenia to augment the experience of pervasive games," in *Paper presented at the Breaking the Magical Circle seminar* (Tampere).

de Lange, M. (2009). "From always on to always there: locative media as playful technologies," in *Digital Cityscapes: Merging Digital and Urban Playspaces*, eds A. de Souza e Silva and D. M. Sutko (New York, NY: Peter Lang), 55–70.

de Lima, E. S., Feijó, B., and Furtado, A. L. (2018). Player behavior and personality modeling for interactive storytelling in games. *Entertain Comput*. 28, 32–48. doi: 10.1016/j.entcom.2018.08.003

Denisova, A., Nordin, A. I., and Cairns, P. (2016). "The convergence of player experience questionnaires," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, TX: ACM), 33–37.

Deterding, S. (2014). "Eudaimonic design, or: six invitations to rehtink gamification," in *Rethinking Gamification*, eds M. Fuchs, S. Fizek, P. Ruffino, and N. Schrape (Lüneburg: Meson Press), 305–323. Available online at: ssrn.com/abstract=2466374

Deterding, S. (2016). "Make-believe in gameful and playful design," in *Digital Make-Believe, Human-Computer Interaction Series*, eds P. Turner and J. T. Harviainen (Cham: Springer International Publishing), 101–124.

Dey, A., Billinghurst, M., Lindeman, R. W., and Swan, J. E. (2018). A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Front. Rob. AI* 5, 37. doi: 10.3389/frobt.2018.00037

Du, R., Turner, E., Dzitsiuk, M., Prasso, L., Duarte, I., Dourgarian, J., et al. (2020). "DepthLab: real-time 3D interaction with depth maps for mobile augmented reality," in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event USA. ACM), 829–843. doi: 10.1145/3379337.3415881

Ducasse, J., Kljun, M., and Čopič, Pucihar, K. (2020). Interactive web documentaries: a case study of audience reception and user engagement on iOtok. *Int. J. Hum. Comput. Interact*. 36, 1558–1584. doi: 10.1080/10447318.2020.1757255

Dunham, J., Papangelis, K., LaLone, N., and Wang, Y. (2018). Casual and hardcore player traits and gratifications of pokemon GO, harry potter: wizards unite, ingress. *arXiv [Preprint]*. doi: 10.48550/arXiv.2103.00037

Dyson, F. (2009). *Sounding New Media: Immersion and Embodiment in the Arts and Culture*. Berkeley, CA: University of California Press.

Ekman, I., Ermi, L., Lahti, J., Nummela, J., Lankoski, P., and Mäyrä, F. (2005). "Designing sound for a pervasive mobile game," in *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology-ACE '05* (Valencia: ACM Press), 110–116.

Ellis, L. A., Lee, M. D., Ijaz, K., Smith, J., Braithwaite, J., and Yin, K. (2020). COVID-19 as "game changer" for the physical activity and mental well-being of augmented reality game players during the pandemic: mixed methods survey study. *J. Med. Internet Res*. 22, e25117. doi: 10.2196/25117

Ferreira, C., Maia, L. F., de Salles, C., Trinta, F., and Viana, W. (2019). Modelling and transposition of location-based games. *Entertain Comput*. 30, 100295. doi: 10.1016/j.entcom.2019.100295

Georgiou, Y., and Kyza, E. A. (2017). The development and validation of the ARI questionnaire: an instrument for measuring immersion in location-based augmented reality settings. *Int. J. Hum. Comput. Stud*. 98, 24–37. doi: 10.1016/j.ijhcs.2016.09.014

Goffman, E. (1986). *Frame Analysis: An Essay on the Organization of Experience*. Boston, MA: Northeastern University Press.

Gračanin, D. (2018). "Immersion versus embodiment: embodied cognition for immersive analytics in mixed reality environments," in *Augmented Cognition: Intelligent Technologies, Vol. 10915*, eds D. D. Schmorrow and C. M. Fidopiastis (Cham: Springer International Publishing), 355–368.

Green, M. C., and Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *J. Pers. Soc. Psychol*. 79, 701–721. doi: 10.1037/0022-3514.79.5.701

Grubert, J., Langlotz, T., Zollmann, S., and Regenbrecht, H. (2017). Towards pervasive augmented reality: context-awareness in augmented reality. *IEEE Trans. Vis. Comput. Graph*. 23, 1706–1724. doi: 10.1109/TVCG.2016.2543720

Hecht, D., Reiner, M., and Halevy, G. (2005). *Multi-Modal Stimulation, Response*. London: Time and Presence.

Jegers, K., and Wiberg, M. (2006). Pervasive gaming in the everyday world. *IEEE Pervasive Comput*. 5, 78–85. doi: 10.1109/MPRV.2006.11

John, O., and Srivastava, S. (1999). *The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives., Volume 2 of Handbook of Personality: Theory and Research*. New York, NY: Guildford Press.

Karapanos, E., Barreto, M., Nisi, V., and Niforatos, E. (2012). Does locality make a difference? Assessing the effectiveness of location-aware narratives. *Interact. Comput*. 24, 273–279. doi: 10.1016/j.intcom.2012.03.005

Kari, M., Grosse-Puppendahl, T., Coelho, L. F., Fender, A. R., Bethge, D., Schutte, R., et al. (2021). "TransforMR: pose-aware object substitution for composing alternate mixed realities," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (Bari: IEEE), 69–79.

Kasapakis, V., and Gavalas, D. (2015). Pervasive gaming: status, trends and design principles. *J. Network Comput. Appl*. 55, 213–236. doi: 10.1016/j.jnca.2015.05.009

Kim, K., Billinghurst, M., Bruder, G., Duh, H. B.-L., and Welch, G. F. (2018). Revisiting trends in augmented reality research: a review of the 2nd decade of ISMAR (2008–2017). *IEEE Trans. Vis. Comput. Graph*. 24, 2947–2962. doi: 10.1109/TVCG.2018.2868591

Klein, G., and Murray, D. (2009). "Parallel tracking and mapping on a camera phone," in *2009 8th IEEE International Symposium on Mixed and Augmented Reality* (Orlando, FL: IEEE), 83–86.

Korhonen, H., Saarenpää, H., and Paavilainen, J. (2008). "Pervasive mobile games-a new mindset for players and developers," in *Fun and Games: Second International Conference* (Eindhoven), 12.

Krzywinska, T., Phillips, T., Parker, A., and Scott, M. J. (2020). From Immersion's bleeding edge to the augmented telegrapher: a method for creating mixed reality games for museum and heritage contexts. *CM J. Comput. Cult. Heritage* 13, 21. doi: 10.1145/3414832

Kuijpers, M. M., Hakemulder, F., Tan, E. S., and Doicaru, M. M. (2014). Exploring absorbing reading experiences: developing and validating a self-report scale to measure story world absorption. *Sci. Study Lit*. 4, 89–122. doi: 10.1075/ssol.4.1.05kui

Laato, S., Pietarinen, T., Rauti, S., and Laine, T. H. (2019). "Analysis of the quality of points of interest in the most popular location-based games," in *Proceedings of the 20th International Conference on Computer Systems and Technologies* (Ruse: ACM), 153–160.

LeGendre, C., Ma, W.-C., Fyffe, G., Flynn, J., Charbonnel, L., Busch, J., et al. (2019). DeepLight: learning illumination for unconstrained mobile mixed reality. *Publisher: arXiv Version Number: 1*. doi: 10.1145/3306307.3328173

Lehman, S. M., Ling, H., and Tan, C. C. (2020). "ARCHIE: a user-focused framework for testing augmented reality applications in the wild," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Atlanta, GA: IEEE), 903–912.

Leorke, D. (2018). *Location-Based Gaming: Play in Public Space*. Singapore; Palgrave Macmillan.

Lessiter, J., Freeman, J., Keogh, E., and Davidoff, J. (2001). A cross-media presence questionnaire: the ITC-sense of presence inventory. *Presence* 10, 282–297. doi: 10.1162/105474601300343612

Lombard, M., Ditton, T. B., and Weinstein, L. (2009). "Measuring presence: the temple presence inventory," in *Proceedings of the Twelfth International Workshop on Presence, International Society for Presence Research* (Los Angeles, CA), 1–15.

Machon, J. (2011). *(Syn)aesthetics: Redefining Visceral Performance*. Hampshire; New York, NY: Palgrave Macmillan; Basingstoke.

Macvean, A., Hajarnis, S., Headrick, B., Ferguson, A., Barve, C., Karnik, D., et al. (2011). "WeQuest: scalable alternate reality games through end-user content authoring," in *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology* (Lisbon: ACM Press), 1–8.

Manzo, C., Kaufman, G., Punjasthitkul, S., and Flanagan, M. (2015). "By the people, for the people": assessing the value of crowdsourced, user- generated metadata. *Digit. Humanities Q*. 9. Available online at: digitalhumanities.org/dhq/vol/9/1/000204/000204.html

McDonnell, N., and Wildman, N. (2019). Virtual reality: digital or fictional? *Disputatio* 11, 371–397. doi: 10.2478/disp-2019-0004

Mcmahan, A. (2003). "Immersion, engagement, and presence," in *The Video Game, Theory Reader*, eds M. J. P. Wolf and B. Perron (New York, NY: Routledge; Taylor & Francis Group), 67–86.

Merleau-Ponty, M., and Landes, D. A. (2012). *Phenomenology of Perception*. Abingdon, VA; New York, NY: Oxon; Routledge. doi: 10.4324/9780203720714

Montola, M. (2011). A ludological view on the pervasive mixed-reality game research paradigm. *Pers. Ubiquitous Comput*. 15, 3–12. doi: 10.1007/s00779-010-0307-7

Murray, J. H. (1998). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Cambridge, MA: MIT Press.

Packer, H. S., Hargood, C., Howard, Y., Papadopoulos, P., and Millard, D. E. (2017). "Developing a writer's toolkit for interactive locative storytelling," in *Interactive Storytelling, volume 10690 of Lecture Notes in Computer Science*, eds N. Nunes, I. Oakley, and V. Nisi (Cham: Springer International Publishing), 63–74.

Papagiannis, H. (2017). *Augmented Human: How Technology is Shaping the New Reality*. Beijing: O'Reilly.

Qin, H., Patrick Rau, P.-L., and Salvendy, G. (2009). Measuring player immersion in the computer game narrative. *Int. J. Hum. Comput. Interact*. 25, 107–133. doi: 10.1080/10447310802546732

Ratcliffe, J., Soave, F., Bryan-Kinns, N., Tokarchuk, L., and Farkhatdinov, I. (2021). "Extended reality (XR) remote research: a survey of drawbacks and opportunities," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama), 1–13. doi: 10.1145/3411764.3445170

Raybourn, E. M., Stubblefield, W. A., Trumbo, M., Jones, A., Whetzel, J., and Fabian, N. (2019). "Information design for XR immersive environments: challenges and opportunities," in *Virtual, Augmented and Mixed Reality. Multimodal Interaction, Vol. 11574*, eds J. Y. Chen and G. Fragomeni (Cham: Springer International Publishing), 153–164.

Reid, J. (2008). "Design for coincidence: incorporating real world artifacts in location based games," in *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts* (Athens: ACM Press), 18–25.

Reid, J., Hull, R., Cater, K., and Fleuriot, C. (2005). "Magic moments in situated mediascapes," in *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Valencia: ACM Press), 290–293.

Schirm, J., Tullius, G., and Habgood, J. (2019). "Towards an objective measure of presence: examining startle reflexes in a commercial virtual reality game," in *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts - CHI PLAY '19 Extended Abstracts* (Barcelona: ACM Press), 671–678.

Schoenau-Fog, H., Lim, S. L. T., and Soto-Sanfiel, M. T. (2013). "Narrative engagement in games-a continuation desire perspective," in *Proceedings of the 8th International Conference on the Foundations of Digital Games (FDG 2013)*, *Vol. 8* (Crete: PublisherSociety for the Advancement of the Science of Digital Games), 384–387.

Serino, M., Cordrey, K., McLaughlin, L., and Milanaik, R. L. (2016). Pokémon Go and augmented virtual reality games: a cautionary commentary for parents and pediatricians. *Curr. Opin. Pediatr*. 28, 673–677. doi: 10.1097/MOP.0000000000000409

Shin, J., Kim, H., Parker, C., Kim, H.-,i., Oh, S., and Woo, W. (2019). Is "Any room really OK? the effect of room size and furniture on presence, narrative engagement, and usability during a space-adaptive augmented reality game," in *2019 IEEE International Symposium on Mixed and Augmented Reality* (Beijing: IEEE), 135–144.

Slater, M. (1999). Measuring presence: a response to the witmer and singer presence questionnaire. *Presence* 8, 560–565. doi: 10.1162/105474699566477

Slater, M., McCarthy, J., and Maringelli, F. (1998). The influence of body movement on subjective presence in virtual environments. *Hum. Factors* 40, 469–477. doi: 10.1518/001872098779591368

Tanenbaum, K., and Tanenbaum, J. (2009). *Commitment to Meaning: A Reframing of Agency in Games*. UC Irvine: Digital Arts and Culture.

Tregel, T., Raymann, L., Göbel, S., and Steinmetz, R. (2017). "Geodata classification for automatic content creation in location-based games," in *Serious Games, volume 10622 of Lecture Notes in Computer Science*, eds M. Alcñiz, S. Göbel, M. Ma, M. Fradinho Oliveira, J. Baalsrud Hauge, and T. Marsh (Cham: Springer International Publishing), 212–223.

Vayanou, M., Katifori, A., Karvounis, M., Kourtis, V., Kyriakidi, M., Roussou, M., et al. (2014). "Authoring personalized interactive museum stories," in *Interactive Storytelling, Vol. 8832*, eds A. Mitchell, C. Fernández-Vara, and D. Thue (Cham: Springer International Publishing), 37–48.

Viudes-Carbonell, S. J., Gallego-Durán, F. J., Llorens-Largo, F., and Molina-Carmona, R. (2021). Towards an iterative design for serious games. *Sustainability* 13, 3290. doi: 10.3390/su13063290

Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence* 7, 225–240. doi: 10.1162/105474698565686

Zhao, J., Simpson, M., Sajjadi, P., Wallgrun, J. O., Li, P., Bagher, M. M., et al. (2021). "CrowdXR-pitfalls and potentials of experiments with remote participants," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (Bari: IEEE), 450–459.