Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Immersive *Spring Morning in the Han Palace*: Learning Traditional Chinese Art Via Virtual Reality and Multi-Touch Tabletop

Sheng Jin, Min Fan & Aynur Kadir

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2021.1930389

Immersive *Spring Morning in the Han Palace*: Learning Traditional Chinese Art Via Virtual Reality and Multi-Touch Tabletop

Sheng Jin[a*], Min Fan [b*], and Aynur Kadir[c]

[a]School of Animation and Digital Arts, Communication University of Zhejiang, Hangzhou, Zhejiang, China; [b]School of Animation and Digital Arts, Communication University of China, Chaoyang District, Beijing, China; [c]Department of Communication Arts, University of Waterloo, Waterloo, Ontario, Canada

ABSTRACT

We examine the user experience, learning effectiveness, learning behaviors, and learning motivation of 54 young adults in pairs using an immersive virtual reality (IVR) environment and a multi-touch tabletop (MTT) to gain a cultural-historical understanding of the traditional Chinese paintings, *Spring Morning in the Han Palace* and *The Night Revels of Han Xizai* respectively. We used mixed methods of knowledge assessments, questionnaires, observation, and interviews to collect and analyze data. The results reveal the IVR environment significantly increased the learning effectiveness and motivation compared to the MTT system, particularly in tasks related to recall of details or spatiality. The IVR design incorporates five learning scenes, and as part of this aspect, the qualitative results indicate users in the IVR condition (1) developed a variety of exploratory and embodied learning strategies; (2) quickly switched their attention to the central learning content, and (3) had equitable interactions in collaborative learning. This research contributes to the potential of IVR learning effectiveness and associated design considerations to learn traditional cultural heritage.

1. Introduction

Traditional paintings are an important form of cultural heritage that imparts an understanding of ancient art and historical culture. Traditional Chinese paintings especially impart important aspects of rich history and cultural knowledge that are specific to historical periods (Song, 2017). Fine arts students in Mainland China are required to take courses on the appreciation of traditional paintings, learning both artistic styles and cultural-historical knowledge. Such education is also recommended for the public as such arts are a matter of national pride. But what is the best way to convey this knowledge so that it is retained?

Virtual reality (VR) offers a novel way to recreate cultural heritage in artistic forms and facilitate users' learning of cultural-historical knowledge. In this research, we define *VR* as any experience in which participants are immersed in a virtual environment, and the sub-type *Immersive VR* (IVR) as the experience with high degree of immersiveness provided by head-mounted displays (HMD)[1] or by Cave Automated Virtual Environment (CAVE)[2] (Muhanna, 2015; Slater & Sanchez-Vives, 2016). The unique characteristics of VR (particularly IVR) are its high degree of *immersion*,[3] *interactivity*[4] and *imagination*[5] (Concannon et al., 2019). The high degree of immersiveness and interactivity that VR offers that makes it a suitable option to experience and interact with cultural heritage that is physically inaccessible (Limniou et al., 2008; Sharples et al., 2008). Being able to feel like one is moving through inaccessible three-dimensional (3D) spaces or interacting with objects augmented with audio knowledge interpretation makes users experience the "plausibility illusion," allowing users to imagine and learn aspects of culture that are no longer obtainable (Sanchez-Vives & Slater, 2005). Research also suggests VR may facilitate users' understanding of spatial information (Tost & Economou, 2009) and increase their attention and retention of cultural–historical knowledge (Economou & Tost, 2008; Mortara & Catalano, 2018), and promote their learning motivation by engaging them into the exploratory activity (Mortara & Catalano, 2018).

Many VR applications have been designed for viewing cultural heritage (e.g., Barreau et al., 2014; Fassi et al., 2016; Flint et al., 2018), but few[6] have fully explored the value of reconstructing two-dimensional (2D) paintings with 3D virtual space and its potential influence on young audiences' learning effectiveness, learning behaviors and learning motivations.

In our previous research, we took a traditional Chinese handscroll painting from the Han Dynasty, titled *Spring Morning in the Han Palace* (汉宫春晓图), as a design case to explore the initial design space of a 3D IVR painting experience (Jin et al., 2020). The painting was created by 16th century Ming Dynasty artist, Qiu Ying (仇英) (1494–1552), and is currently exhibited in the National Palace Museum in Taipei. The entire piece is painted on a long scroll which measures 30.6 centimeters in height and 574.1

CONTACT Min Fan mfan@cuc.edu.cn School of Animation and Digital Arts, Communication University of China, No1 Dingfuzhuang East Street, Chaoyang District, Beijing 100024, China

*These authors contributed equally to this work.

centimeters in width, and accommodates 115 figures, three main imperial palaces, trees and other objects in the Han Dynasty (202 BC-AD 220). This long handscroll work is an imaginary representation of various activities in a Han Dynasty palace on a spring morning.

In this study, we refined the initial IVR design through formal and informal user testing. Our current IVR system enables multiple users (2–6) to simultaneously explore a virtual, 3D reconstruction of *Spring Morning in the Han Palace*. Users can explore the space through locomotion and teleportation, watch animated characters and activities, and move through five learning scenes to learn associated cultural-historical knowledge. This system was presented as a means of supporting young adults' use and learning compared to a similar multi-touch tabletop (MTT) design.

We focus on young adults because (1) they are the main target group for using IVR in this way; and (2) we can easily access the user group for ethnic reasons. This study also implemented an interactive MTT system as a comparative control using a traditional Chinese handscroll painting from the Tang Dynasty, *The Night Revels of Han Xizai* (韩熙载夜宴图) by Gu Hongzhong (顾闳中).[7] The MTT system was chosen as a control because this approach has often been used to support collaborative learning in school classrooms (Wise et al., 2015) or enrich the audience's personalized visiting experience in museums (Muntean et al., 2016). The learning material of *The Night Revels of Han Xizai* was selected based on the recommendation of two educational experts who have the research background on traditional Chinese painting art.

In the evaluation study, we focus on user experience,[8] learning effectiveness,[9] learning behaviors,[10] and learning motivation[11] when using the two systems. We are interested in investigating these aspects because (1) system user experience is the fundamental aspect that may influence learning effectiveness; (2) learning effectiveness can demonstrate whether or not the system has achieved its design goal; (3) learning behaviors can reveal details of users' interactional and learning process, which may suggest their underlying learning mechanisms and beneficial design strategies; and (4) learning motivation may partially explain users' learning effectiveness and learning behaviors.

Our research questions are:

(1) How well do young adults use the IVR and MTT systems?
(2) How effectively do young adults learn with the IVR and MTT systems?
(3) How do young adults interact and learn with the IVR and MTT systems?
(4) How motivated are young adults when using the IVR and MTT systems?

By addressing these questions, we may better understand (1) what design strategies work or not for reconstructing IVR learning environments for 2D traditional Chinese artworks; (2) how young adults use IVR systems to experience cultural heritage and learn cultural–historical knowledge; and (3) how intertwined factors such as the VR design strategies and users' interactional and learning behaviors and learning motivation may contribute to the learning effectiveness. This knowledge would benefit researchers and designers who work in the field of IVR design for ancient paintings (or similar cultural heritage) and educators who intend to employ IVR technology in future curriculums, as well as users who are interested in traditional Chinese culture and arts.

2. Background

Previous research has illustrated VR affordances in supporting learning and presented the design and evaluation of a variety of VR cultural heritages (Drossis et al., 2018; Huang et al., 2010; Ibrahim & Ali, 2018). In this section, we discuss the main benefits of IVR environments in supporting learning for cultural heritage, the design of virtual heritage on ancient paintings, and the evaluation of VR experience and learning outcomes on cultural heritage. We also identify the research gap by reviewing the previous research and their inconsistent empirical results.

2.1. *Beneficial features of IVR in supporting learning for cultural heritage*

Various researchers have recognized the advantages of IVR technology in supporting learning for cultural heritage (Bekele & Champion, 2019; Drossis et al., 2018; Kersten et al., 2018). IVR can provide a wider range of synthetic sensorial stimuli, which creates a strong sense of presence in a realistic, sensory-rich digital environment in assisting users to conceptualize imagination and construct meaning and knowledge through direct personal experience (Huang et al., 2010). It can revive the past so users can revisit non-existing historical sites and have a close look at reconstructed cultural objects (Loizides et al., 2014).

In its applications, IVR can present interactive, multisensory information relevant to cultural objects (Loizides et al., 2014). Users can manipulate an authentic virtual object while simultaneously viewing or listening to the enhanced knowledge interpretation. The multiple cues (e.g., visual object, audio instruction and kinesthetic action) may improve users' understanding, attention and retention (Kersten et al., 2018; Mortara & Catalano, 2018) due to the embodied cognition[12] and dual-coding theory.[13] One often discussed feature unique to IVR is its spatial nature. In the context of cultural learning, the spatial characteristics of IVR technology enable users to freely explore through authentic historical sites or interact with realistic 3D replicas (Bekele et al., 2018). Such spatial affordances may improve users' spatial perception of historical places and objects (Mortara & Catalano, 2018). However, previous researchers noted users' challenges in navigating through IVR environments (Concannon et al., 2019). Poor usability of IVR navigation severely limits the learning effectiveness (Sutcliffe, 2003), meaning the navigation mechanism should be carefully considered when designing IVR for cultural learning.

Another important feature of IVR applications is the aspect of collaborative learning for cultural heritage (Bekele & Champion, 2019). The theoretical perspective of social constructivist learning has suggested that learning is largely driven by exchange with peers through in-depth discussion and joint action (De Back et al., 2020; Dalgarno, 2002).

Communication may encourage people's ongoing discourse on a topic and help them to consider multiple perspectives of the topic (Greenwald et al., 2017). Shared virtual spaces that can be accessed simultaneously by multiple users, usually through digital avatars, provide a "safe" social environment and shared learning space for discussion and interaction (De Back et al., 2020). Users in groups are able to build an understanding of a learning topic by exchanging ideas, sharing resource, and completing a task with one another (Zheng et al., 2018). The shared interpretations may help to evoke in-depth discussions on learners' personal values and heritage conservation (Rahaman & Tan, 2011). And such value-laden discussions that exceed beyond the basic level of knowledge acquisition are important in learning of cultural heritage.

2.2. The design of virtual heritage for ancient paintings

The term *virtual heritage* can be broadly defined as the preservation of cultural heritage through replication, visualization, or simulation using computer graphics technology (Ibrahim & Ali, 2018). Most virtual heritage environments have focused on reconstructing historical sites and architecture or displaying artifacts (Barreau et al., 2014; Flint et al., 2018). We have found only a few research projects that explored the digital recreation of ancient paintings, and even fewer using a VR design approach.

Earlier works focused on the digital enhancement of paintings using animation. Inspired by Zhu et al.'s (2004) animated Chinese Dunhuang murals, an animation of the *Along the River During the Qingming Festival* (清明上河图) scroll was exhibited in the Chinese pavilion of the Shanghai World Expo 2010 (Crystal Corporation, 2010). Ma et al. (2012) focused on auditory experience and presented a digital version of the *Along the River During the Qingming Festival* scroll with voice overs and environmental sounds. The system synthesized realistic stereo audio according to the viewer's orientation and position in physical space when viewing a high-quality digital representation of the painting on a touch display device.

Yuan and Yun (2016) presented one of the first IVR reconstructions of traditional Chinese paintings that enabled viewers to enter the virtual world of *Listening to the Qin* (听琴图). The designers carefully observed the spatial arrangement of the composition to recreate a 3D space based on the 2D painting. However, while the audience could view the 3D scene from various perspectives, no further information was provided to help audiences understand the background or significance of the painting.

Hürst et al. (2016) designed and evaluated an IVR museum wherein audiences could walk into each room of the museum to view stylized (e.g., Van Gogh style), animated (e.g., falling leaves), or stylized and animated paintings. The preliminary results of a study with 12 university students showed that the animated paintings triggered more emotional verbalizations than the other two states. However, this project only created a 3D virtual museum to display 2D paintings; no interaction was possible in this virtual environment.

2.3. The evaluation of VR on cultural heritage experience and learning

Although research on virtual heritage of VR is on the rise, recent studies have focused almost exclusively on the design of digital documentation. Little attention has been paid onto the impact of VR technologies on the audience' cultural heritage experience and learning.

Liu (2020) presents a mixed-methods study (combining questionnaires and semi-structured interviews) to evaluate visitors' expectations, acceptance and experience of digital technologies such as VR and augmented reality (AR) at cultural heritage sites in Taiwan. The study indicates that digital display technologies received high acceptance from the audience and had a positive impact on some specific aspects, such as encouraging their exploration of and further learning about the site.

Antonietti and Cantoia (2000) conducted an empirical study to explore how desktop VR environments could help undergraduate students deeply understand historical Western paintings. Forty students were randomly assigned to either reflect upon a high-quality 2D reproduction of the 17th century painting *Saint Jerome* or be guided through a VR tour of it on a desktop computer screen. The results showed that the VR experience prompted students to consider *why* and *how* something was in front of them rather than what they faced; VR also helped the students to conceptualize the experience at an abstract level and stimulated a free and imaginative elaboration.

However, learning effectiveness with VR technologies is not always positive. Champion (2006) evaluated the impact of various interaction methods (i.e., observation, instruction, and activity-based exploration) on users' cultural learning experience in a desktop VR environment. Task performance, knowledge recall, and interactional behaviors were collected and measured. The study suggests game-based interactivity in VR design may help reduce navigation errors, but it may also restrict the meaningful learning experience.

Tost and Economou (2009) present a study to investigate whether IVR is suitable for learning about archeology in cultural heritage settings. This research compared the learning outcomes, perception and use by audiences of two different IVR systems (called "Kivotos" and "Tholos"[14]) and a related traditional non-IVR exhibition. Qualitative and quantitative methods were used to analyze data gathered through in-situ observations, interviews with museum educators and face-to-face questionnaires with visitors. The results revealed a contradiction between visitors' higher-level perceived learning effectiveness and their lower-level actual learning outcomes with IVR technology. The researchers acknowledge the complexity of the causes. One partial reason may be the audio-visual characteristics of IVR is more suitable for obtaining spatial information rather than historical information, and the other may be the lack of the interactivity in their IVR design.

The previous research demonstrates the inconsistent results on the learning effectiveness of VR technologies in cultural heritage contexts, and also suggests the impact of the design on the learning outcomes. It is necessary to consider VR design features and users' experience and learning behaviors while analyzing learning effectiveness, so as to discover the underlying learning mechanism and beneficial design features that may contribute to users' learning outcomes.

3. Reconstructing traditional paintings digitally

3.1. Immersive virtual reality environment

The IVR environment for *Spring Morning in the Han Palace* was developed by Unity 3D and Maya, supported by the HTC Vive, Oculus Rift, or Gear VR (Figure 1). After selecting a personal avatar, viewers can start their free exploration of the virtual reconstruction of the painting.

Once in the world viewers can:

- Visit the world through locomotion or teleportation by pointing to a location using the hand controllers (Figure 1).
- View animated scenes (e.g., playing Chinese music instruments, playing chess, chasing a butterfly, dancing, chatting).
- Interact with and teleport between provided "learning scenes" (Figure 1).
- Interact with a series of 3D cultural objects and learn associated cultural-historical knowledge through audio-visual features (Figure 1).
- Access the original 2D painting with an indication of their current position in 3D space (Figure 1).
- View their teammate's avatar and movement in real-time (Figure 1) to provide an awareness of presence and promote discussion and "emergent dialogues" during the exploration (Antle et al., 2014).

We considered *authenticity* and *fidelity* when designing the 3D visual representation. We also enhanced our visualization with 3D animations and added audio–visual information to allow the user to further interpret the background knowledge of the art, history, and culture. The design rationale and system details have been reported previously (Jin et al., 2020).

In designing the learning scenes, we collaborated with the two educational experts. The five learning scenes in the IVR system (Figure 2) presented artifact styles and details, activities and stories (allusions), or spatial configurations as follows: (1) cultural object: ancient fans (*Uchiwa*; 团扇) from the Han dynasty; (2) musical instrument: the use of *Konghou* (箜篌) in *Qingshangyue* (a music genre) during the Han Dynasty; (3) folk tradition: *Doubaicao* (斗百草), a game from the Dragon Boat Festival (one of China's four traditional festivals); (4) allusion: the story of Mao Yanshou (毛延寿) painting the portrait of Wang Zhaojun (王昭君) (one of the "four beauties" of ancient China); and (5) architecture (spatial configuration): the "reduction of columns" (减柱造) method (an ancient building technique for reducing the number of internal columns in wooden buildings). Based on the experts' suggestions, three assumed difficulty levels were designed, including (i) easy (i.e., the first two tasks); (ii) intermediate (i.e., the third and fourth tasks); and (iii) difficult (i.e., the last task). The levels of difficulty were based on the complexity and familiarity of the information that learners need to understand and memorize in a certain task.

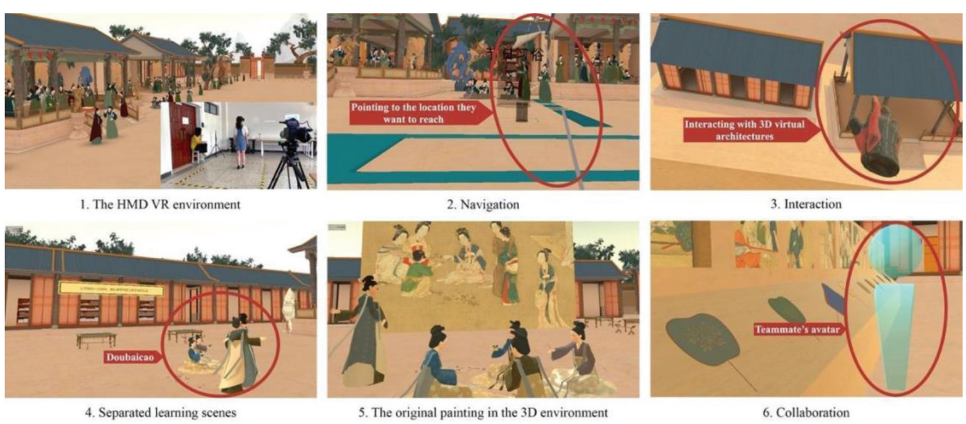

Figure 1. Design of the immersive VR environment.

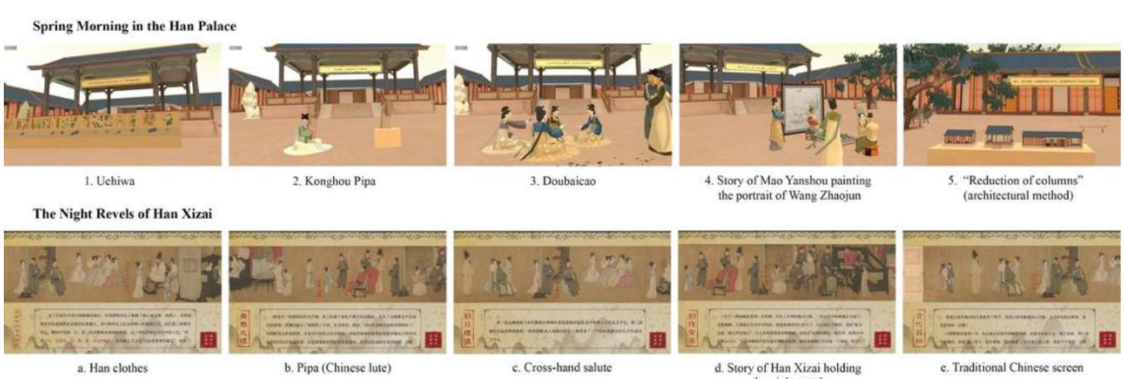

Figure 2. Five learning scenes in the VR and multi-touch tabletop (MT) systems.

3.2. Multi-touch table system

The MTT system was developed by Ventuz. It allows users to scroll through the 2D handscroll painting with similar multimedia materials and interactive functions to the IVR system, but displayed in 2D non-immersive form. Users can interact with five hotspots on the painting to trigger audio-visual contents including audio and text below the original painting (Figure 3).

To control for learning effects (i.e., the same participants used both systems to learn in the within-subjects design), we selected an equivalent handscroll for the MTT system, *The Night Revels of Han Xizai* from the Tang Dynasty. We also consulted with the educational experts to design five equivalent learning scenes, which were (Figure 2): (a) cultural object: traditional Han clothes and their unique color palate (汉服); (b) musical instrument: the use of pipa (琵琶) (a Chinese lute); (c) folk tradition: cross-hand salute (叉手礼) from the Tang Dynasty; (d) allusion: story of Han Xizai holding the night revels; and (e) furniture (spatial configuration): traditional Chinese screen (屏风).

4. Evaluation study

4.1. Study design

We used mixed-methods research (concurrent design) as the methodology for our investigations (Creswell & Clark, 2011). We performed a comparative study to explore how the IVR system compares to the MTT system. In the IVR condition, pairs of participants used the 3D IVR system to learn *Spring Morning in the Han Palace*. In the MTT condition, the same pairs of participants used the MTT system to learn *The Night Revels of Han Xizai*. We asked participants to perform the task in pairs rather than individually because (1) we believe that learning is largely driven by in-depth discussion and joint action in pairs according to the social constructivist learning view (Dalgarno, 2002); (2) both IVR and MTT are suggested to support collaborative learning (De Back et al., 2020; Wise et al., 2015); and (3) such a collaborative learning approach may make the invisible learning process visible. A within-subject design was used because group dynamics can influence collaboration (Meerbeek et al., 2010) and participant's personal characteristics (e.g., spatial ability) and prior technology experience may influence their UX experience and/or learning outcomes (Sagnier et al., 2020; Sun et al., 2019). To control for order effects, the conditions were counterbalanced.

4.2. Apparatus

The IVR environment used in this study was supported by an Oculus Rift CV1 (2016, the Oculus VR Corporation). The Oculus Rift CV1 offers separate displays for each eye with 1080 × 1200 resolution (2160 × 1200 in total) and yields a 100-horizontal field of view with a refresh rate of 90 Hz. The device was connected to an HP Z840 with an Intel Xeon E5-2620V4 × 2 (8 core 2.1 GHz/ 3.0 GHz turbo ×2), 16 GB RAM and an NVIDIA GTX1080ti Graphic card.

This MMT system used in this study based on Gonsion Multi-Touch tabletop computer (2016, Gonsion displays Corporation). This tabletop system is featured with a 3 M 55 inch 4 K UHD multi-touch screen, which uses 3 M projected capacitive (PCAP) multi-touch technology that can track 80 simultaneous touch points on the surface. The device was connected to an Thinkstation with an Intel Core E-2124 G (quad-core 3.3 GHz), 16 GB RAM and an NVIDIA Quadro Q6000 Graphic card.

4.3. Participants

We collected data from 54 participants (27 male, 27 female) who used both systems in pairs. The participants were university students aged 19–22 (mean age = 20.43 ± 0.13) who volunteered to participate in our study, which was advertised through the university's online forum. We tested each participant's prior knowledge about the two paintings and relevant knowledge on traditional culture, and only those who had none or minimal knowledge (i.e., had not heard of the paintings or had only heard their names) were qualified to participate. All participants had used a touchscreen device before. In terms of familiarity with MTT technology, 37.0% (n = 20) reported that they had used it many times, 25.9% (n = 14) had used it a few times, 35.2% (n = 19) had rarely used it, while 1.9% (n = 1) had never used it. As for VR technology, 3.7% (n = 2) of the participants had used it many times, 14.8% (n = 8) had used it a few times, 46.3% (n = 25) had rarely used it, while 35.2% (n = 19) had never used it. None of the participants had used our systems before. The groups were randomly assigned, and the participants knew each other to different degrees: some were good friends (11 of 27), some were classmates (12 of 27), while others did not know each other at all (4 of 27).

4.4. Tasks

The participants were given tasks during and after the learning phase. During the learning phase, the participants were

Figure 3. Design of the multi-touch tabletop (MT) system.

asked to “*Explore the interactive system and complete the learning of five knowledge points*.” After this, participants were asked to complete five written tasks related to the culture, history, and art embodied by the painting. We intentionally told participants the post-learning tasks before the learning phase to encourage them to explore the systems with a clear learning goal. In each system, participants had a maximum of 25 minutes for the learning phase, but they could finish earlier if both participants agreed. There was no fixed approach for the learning phase; that is, the participants were given complete freedom when using the system.

4.5. Procedure

The IVR and MTT systems were set up in a controlled lab space. The session started with a demographic survey. The participants were given a basic tutorial of the system before each trial. The pairs then had 3–5 minutes to familiarize themselves with the system. When they were ready to start, they were given the tasks. Each pair worked together in each system (Figure 4). They changed systems after 25 minutes (or earlier if desired). Each participant was given an individual posttest which included a total of five four-option-multi-choice questions, and each was related to one of the knowledge points in the five learning scenes. Each participant was also given a five-point Likert scale questionnaire[15] that included 33 items related to *user experience* (RQ1), *self-perceived learning effectiveness* (RQ2), and *learning motivation* (RQ4). The questionnaire was developed based on previous research (Huang et al., 2010). That is, most of the items were from the validated questionnaire but with the change of the wording to better fit for the current study context (e.g., changing *I feel that I have developed better understanding of structures and orientations of organs by using this system* to *I feel that I have developed a better understanding of structures and orientations of architectures by using this system*) while a few items were also added based on our research questions (e.g., *using this system has increased my interest in traditional culture*). We then conducted a four-step content validity study[16] to validate our questionnaire with an external expert who specialized in VR evaluation and by a pilot study with five participants. After the two sessions, the participants were interviewed in pairs about their general feedback and experience of two systems (e.g., *what did you find most impressive about the two systems?*). The study lasted a total of 70–90 minutes and each participant was compensated with 150 RMB (around 20 USD USD) for their time.

4.6. Data collection and analysis

We used a mixed-methods approach to collect and analyze our data. To address RQ1 (*user experience*) we used quantitative methods. We collected individual participants’ responses of user experience on interaction and satisfaction when using the two systems and used descriptive analysis (mean and standard deviation (SD)) and the Wilcoxon signed-rank test (a nonparametric alternative to paired-sample t-tests when the data are *not* normally distributed but still *symmetrically* distributed) to compare the differences in the median rated scores between the two conditions. Cronbach’s alpha test was used to measure the scale reliability for all the items of our questionnaire (same for all the questionnaires mentioned as below). We also recorded participants’ self-determined exploration in each system and used paired-sample t-tests to analyze the difference in the mean values for the two conditions. We supplemented these quantitative findings with qualitative interview data. We used thematic analysis to identify common and interesting themes (see details of our coding approach in RQ3).

To answer RQ2 (*learning effectiveness*), we recorded the participants’ raw accuracy scores (out of five) of the cultural–historical knowledge in the posttests for each condition and used the Wilcoxon signed-rank test to determine the differences in learning effectiveness (median raw scores) and efficiency (i.e., how efficiently raw score divided by exploratory duration) between the pretest scores (= 0) and posttest scores. We also collected individual participants’ responses of their perceived learning effectiveness when using the two systems and used descriptive analysis and the Wilcoxon signed-rank test to compare the differences in the median rated scores between the two conditions. We supplemented these quantitative findings with interview data (see details in RQ3).

To answer RQ3 (*learning behaviors*), we collected data about participants’ interactions and learning behaviors using video recordings and observational notes. The observational notes were taken by two researchers. We used a hybrid coding approach to create both pre-set themes and codes and emergent themes and codes. We began with pre-set themes and codes based on previous

Figure 4. Participants interacted in pairs with two systems during the within-subjects design experiment.

research and our research questions. For example, previous research suggests VR environments encourage users' exploration, so one pre-set theme was "exploratory behaviors." Under this theme, several pre-set codes were also created, including users' (1) exploratory strategies, and (2) social interaction during exploration. Other pre-set themes and codes can be found in Table 1. In addition to the pre-set themes and codes, we coded other emergent behaviors observed or video recorded during the studies. All types of the qualitative data were coded separately by two researchers and disagreements between the coders were resolved through discussion.

To answer RQ4 (*learning motivation*), we collected and analyzed participants' individual responses to the questionnaire and used the Wilcoxon signed-rank test to compare the differences. We conducted post-learning interviews with each pair to query their learning motivations and general impression while using each system.

A pilot study with five students was undertaken before the evaluation study to ensure that the learning tasks were appropriately designed and the items of the questionnaire were appropriately adapted to the current study context. The data obtained from the pilot study was analyzed for response competence, accuracy and reliability (internal consistency). The results were general positive while some adjustments were made to the questionnaire items. For example, we highlighted the reserve items to improve the readability.

5. Results

5.1. User experience

The Cronbach's alpha test yielded a high level of internal consistency for all items of User Experience (10 items, IVR-alpha = 0.876, MTT-alpha = 0.847) and each of the two subscales of Interaction (5 items, IVR-alpha = 0.891, MTT-alpha = 0.770) and Satisfaction (5 items, IVR-alpha = 0.881, MTT-alpha = 0.810) in our questionnaire. The questionnaire results showed that participants had a steeper learning curve with the IVR system compared to the MTT system. There was a statistically significant median decrease in ease of following instructions in the IVR system (p = .038, z = −2.070). However, participants were able to easily navigate in the virtual space (p < .0005, z = 3.186) and preferred the way of viewing (p = .002, z = 3.049) in the 3D simulated environment in the IVR system rather than the 2D painting in the MTT system. They described the IVR system as much more immersive (p < .0005, z = 3.676) and were strongly motivated to continue using the IVR system in the future (p = .022, z = 2.286) (Table 2).

We also examined participants' self-determined exploratory time between the two conditions using paired sample t-tests. The pairs spent significantly longer in the IVR condition (mean 19.2 ± 3.25 min) than in the MTT condition (mean 9.2 ± 4.34 min) (p < .0005, t = 13.053).

The interview results were also consistent with the quantitative results. Most participants mentioned it was easier for them to learn and interact with the MTT system, but they would quickly get bored with the simple text-based content and the touch-based interaction. Several participants (n = 14)

Table 1. Pre-set themes and codes in the qualitative analysis process.

Data	Pre-set themes	Pre-set codes	Meanings	Based on
Observation And video data	**Exploration (RQ3)**	Exploratory strategies	Any verbal and non-verbal behaviors indicate users 'decisions on the exploratory orders	Previous research (Hanson & Shelton, 2008; Liu, 2020) and research questions
		Social interaction during exploration	Any verbal and non-verbal behaviors indicate users' social interaction during exploration	
	Interaction (RQ3)	Hands-on actions	Any hands-on behaviors with/without virtual artifacts	Previous research (Rahaman & Tan, 2011; Takács, 2011) and research questions
		Whole-body actions	Any whole-body behaviors with/without virtual artifacts	
		Navigations	Any verbal and non-verbal behaviors related to navigation between scenes	
	Attention (RQ3)	Comments on attention	Any verbal behaviors that indicate that users have focused attention	Previous research (Huang et al., 2010; Mortara & Catalano, 2018) and research questions
		Actions on attention	Any non-verbal behaviors that indicate users have focused attention	
	Collaboration (RQ3)	Discussions	Any verbal behaviors that indicate users make mutual understanding through discussions	Previous research (Mulholland & Collins, 2002) and research questions
		Conflicts	Any verbal behaviors that indicate users have conflicts and/or resolve the conflicts	
Interview	**User Experience (RQ1)**	Easy to learn	Any verbal behaviors related to easy to learn of IVR or MTT systems	Previous research (Huang et al., 2010; Rahaman & Tan, 2011) and research questions
		Interaction and satisfaction	Any verbal behaviors related to interaction and satisfaction of IVR or MTT systems	
	Learning (RQ2 & RQ4)	Perceived learning effectiveness	Any verbal behaviors related to learning effectiveness (i.e., imagination and enhanced learning capability)	
		Learning motivation	Any verbal behaviors related to intrinsic and extrinsic learning motivation	

Table 2. Participants' user experience under IVR and MTT conditions.

Aspects	System interaction and experience	Mean (SD)		Median (IQR)		Wilcoxon signed-rank test (IVR-MTT)	
		IVR	MTT	IVR	MTT	z	p
Interaction	I feel that I can easily learn to use the system.	4.3(0.71)	4.4(0.71)	4(1)	4.5(1)	−0.991	0.321
	I feel that the system instruction is easy to follow (*reversed*).	4.0(0.64)	4.3(0.74)	4(0)	4(1)	−2.070	0.038*
	I feel that the system feedback is clear.	3.8(0.64)	4.0(0.73)	4(1)	4(0)	−1.290	0.197
	By using this system, I can easily navigate my position within this system.	3.9(0.79)	3.7(0.71)	4(1.25)	4(1)	1.741	0.082
	By using this system, I can easily jump to or jump back from the learning scenes/hotspots (*reversed*).	4.1(0.66)	3.7(0.73)	4(1)	4(1)	3.186	0.001*
Satisfaction	I like the way to view the 3D simulated environment/2D painting.	4.2(0.75)	3.7(0.68)	4(1)	4(1)	3.049	0.002*
	I like the way to navigate the system (*reversed*).	3.8(0.78)	3.6(0.71)	4(1)	4(1)	1.823	0.068
	I feel the simulated environment provided by this system is immersive.	4.0(0.82)	3.5(0.63)	4 (1)	4(1)	3.676	0.000*
	I think I am satisfied with the system.	3.7(0.80)	3.5(0.77)	4(1)	4(1)	1.209	0.227
	I am willing to continue using this system in the future (*reversed*).	3.9(0.10)	3.6(0.81)	4(0.25)	4(1)	2.286	0.022*

further presented evidence of their previous experiences with similar interactive MTT systems in museums or public spaces; they often quickly scanned the systems, interacted with one or two points, and then walked away.

5.2. Learning effectiveness

A Wilcoxon signed-rank test with the pre- and posttest results showed a statistically significant improvement in the median test scores in the IVR condition (raw scores = 3.5) compared to the MTT condition (raw scores = 3.0) (p = .006, z = 2.737). We compared the participants' accuracy for each learning task; the descriptive analysis showed that participants had higher accuracy in the IVR condition in four of the five learning tasks (Table 3). However, in the second task that focused on activity, the participants had higher accuracy in the MTT condition. This may be related to the activities presented; many users leveraged the use of an embodied action to imitate the cross-hand salute in the MTT condition. In contrast, the equivalent activity in the IVR condition, the *Doubaicao* game, was complex and difficult to participate in, so the participants simply observed others playing it. The Wilcoxon signed-rank test revealed statistically higher median raw scores for accuracy in the second (p < .0005, z = 4.459) and fourth (p = .002, z = 3.128) and fourth tasks, which focused on story comprehension and details of musical instruments, respectively. The results suggest the IVR system enabled better recall of the details of cultural artifacts, understanding of stories, and performance in tasks related to spatiality.

We also compared the learning efficiency (raw score divided by time used) between the two conditions. The learning efficiency was statistically higher in the MTT condition (p < .0005, t = 4.767). This was considered to be because (1) all pairs self-determined to explore for longer in the IVR environment, and (2) there was a small number of learning tasks offered in the test. Our interviews provided further explanations on the learning efficiency. Around two-thirds of participants (n = 33) mentioned they could quickly scan the text in the MTT system, which facilitated efficient learning. However, most of the 33 participants (n = 29) said there was a greater chance they would not see or read all the knowledge points within the MTT system if they were not instructed to do so in real-life contexts.

The individual questionnaire included questions about the participant's self-perceived learning effectiveness with the two systems (Table 4). Cronbach's alpha for all items of Learning Effectiveness (12 items, IVR-alpha = 0.831, MTT-alpha = 0.858) and each of the two subscales of Imagination (5 items, IVR-alpha = 0.803, MTT-alpha = 0.849) and Enhanced Learning Capability (7 items, IVR-alpha = 0.747, MTT-alpha = 0.709) yielded a high level of internal consistency. There was a significant difference between the two approaches in the median scores of perceived effectiveness at facilitating imagination during learning. That is, participants tended to believe that the 3D IVR learning environment helped them imagine the details (p = .025, z = 2.243; p = .025, z = 2.245,) and spatial qualities (p = .025, z = 2.24) of historical artifacts and understand historical events (p = .008, z = 2.652) more easily than the MTT system. Participants also indicated that the IVR system enriched their self-exploration learning capability (p < .0005, z = 3.860). However, there were no significant differences in the perceived effectiveness of the two approaches in supporting attention and other learning capabilities such as knowledge construction and management. This may be because both systems used multimedia and interactions (hierarchically and logically), which help with visualizing and organizing information.

Table 3. Raw accuracy scores in the posttest of different learning tasks under IVR and MTT conditions.

Type	Learning task		Accuracy (%)		Median accuracy (IQR)		Wilcoxon signed-rank test (IVR–MTT)	
	IVR	MTT	IVR	MTT	IVR	MTT	z	p
Cultural object	Uchiwa shape patterns	Color palettes of Han clothes	80	76.36	1(0)	1(0)	0.5	0.617
Musical instrument	Konghou	Pipa (Chinese lute)	76.36	29.09	1(0)	0(1)	4.459	0.000***
Activity	Doubaicao	Cross-hand salute	36.36	58.18	0(1)	1(1)	−2.191	0.028*
Allusion	Story of Mao Yanshou painting the portrait of Wang Zhaojun	Story of Han Xizai holding the night revels	87.27	60	1(0)	1(1)	3.128	0.002**
Spatiality	"Reduction of columns" (architectural method)	Traditional Chinese screen	43.64	41.82	0(1)	0(1)	0.2	0.841

Table 4. Participants' self-perceived learning effectiveness under IVR and MTT conditions.

Aspects	Learning effectiveness	Mean (SD) IVR	Mean (SD) MTT	Median (IQR) IVR	Median (IQR) MTT	Wilcoxon signed-rank test (IVR-MTT) z	Wilcoxon signed-rank test (IVR-MTT) p
Imagination	I feel that I have developed a better understanding of relative positions of the architectures, objects and people by using this system.	4.2(0.74)	3.4(0.96)	4(1)	4(1)	4.250	0.000***
	I feel that I have developed a better understanding of structures and orientations of architectures by using this system (*reversed*).	4.0(0.78)	3.5(0.99)	4 (0.5)	4(1)	2.652	0.008**
	Using this system has helped me develop a better understanding of the details of historical items.	4.0(0.81)	3.7(0.90)	4 (1.25)	4(1)	2.243	0.025*
	Using this system has helped me develop a better understanding of the details of the clothes (*reversed*).	3.8(0.88)	3.4(0.92)	4(1)	3(1)	2.245	0.025*
	I feel that it is easier to understand the historical event by using this system.	3.9(0.87)	3.6(0.79)	4(2)	4(1)	2.171	0.030*
Enhanced learning capability	I feel that interaction makes me concentrate while learning.	3.8(0.86)	3.5(0.82)	4(1)	3.5(1)	1.871	0.061
	I feel the system can draw my attention during learning (*reversed*).	3.6(0.81)	3.5(0.82)	4(1)	3(1)	1.098	0.272
	The collaborative learning mode of this system can improve my concentration.	3.5(1.04)	3.4(0.82)	3.5(1)	4(1)	0.635	0.526
	The system can enhance my capability of self-exploration (*reversed*).	4.2(0.69)	3.7(0.84)	4(1)	4(1)	3.860	0.000***
	The system can enhance my learning capability.	3.7(0.82)	3.5(0.75)	4(1)	3(1)	1.383	0.167
	The system can enhance my capability of knowledge construction.	3.8(0.78)	3.7(0.76)	4(1)	4(1)	0.344	0.731
	The system can enhance my capability of knowledge management (*reversed*).	3.6(0.88)	3.5(0.79)	4(1)	4(1)	0.807	0.420

5.3. Learning behaviors

The observational data revealed several interesting learning behaviors with the use of the IVR system. First, all pairs (number of pairs (*np*) = 27) in the IVR condition discussed and developed "space exploratory strategies" at the beginning of the session. We identified four main strategies. Many pairs (*np* = 14) decided to explore the space from left to right; around half of the pairs (*np* = 10) used a nearby-to-far (center–around) strategy, while three pairs adopted a global–local strategy (that is, they inspected the entire space first and then entered each of the five scenes). We also found that half of the pairs (*np* = 14) designated one participant as an *explorer* to lead the tour and the other as a *follower* who would follow the other.

Second, all participants were extremely *engaged* in exploring details of the cultural artifacts and space. We observed that all participants *actively* explored the virtual space, even after completing the five learning scenes. They also repeatedly *manipulated* the virtual cultural objects and observed them carefully from various 3D perspectives. Many participants (*n* = 42) even intended to collect and carry their favorite cultural artifacts with throughout the exploratory process. Several participants (*n* = 16) attempted to hold down the controller to bring the cultural artifacts displayed on the counters out from the learning scenes.

Third, most of the participants (*n* = 47) developed a set of *embodied learning strategies* while learning the knowledge points. For example, a few participants (*n* = 6) simultaneously manipulated two cultural objects (e.g., two architectures) in each hand (i.e., controller) to compare them in detail; several participants (*n* = 12) constantly compared the reconstructed 3D cultural objects with those depicted in the 2D painting; most participants (*n* = 47) grasped the virtual objects and rotated them to view them from different perspectives closely, rather than walking around the counter to view the cultural objects from a distance (which could be achieved in our IVR system); several participants (*n* = 16) attempted to use and experience different cultural objects, such as using the fan and playing the *Konghou* instrument; and most of the participants (*n* = 42) used whole-body movements to better facilitate their learning process (e.g., bending over/turning around to watch, making the fanning motion) (Figure 5).

Fourth, most participants (*n* = 47) appeared to quickly switch their attention to the learning mode when entering

Figure 5. Embodied learning strategies when interacting with the immersive VR environment.

the separate learning scenes; they immediately became quiet and listened to the audio instruction. They were also aware that they could grasp the cultural objects on the counters without being told. In addition, around 75% of the participants (n = 41) associated the immersive virtual scenes and experience with their real-life experience, which may help them to memorize the learned knowledge. For example, several participants (n = 10) were afraid to knock down other virtual people or historical objects while in the learning scenes; a few participants (n = 4) expressed their surprise when discovering "unreal" moments in the IVR space (e.g., "*Wow! It's not broken!*" when they saw virtual objects did not break when they fell on the ground).

Finally, around 80% of the pairs in the IVR system (np = 22) tended to have more *equitable* interactions and collaboration compared to the MTT system. That is, pairs tended to make mutual decisions through discussion and negotiation. One main reason for this is that each participant had a personal avatar and the autonomy to control their own movement and interaction in the virtual space. The other reason may be linked to social norms among adults. For instance, we noticed one participant would often prompt questions when s/he had finished learning in a certain scene and wanted to switch to another (e.g., "*Have you finished? Shall we switch to the next section to check out ___?*") rather than having confirmatory statements (e.g., "*Let's switch to the next section.*").

However, we also noticed two main potential issues when learning with the IVR system. First, we found that a few participants (n = 7) were extremely attracted to the IVR technology, to the extent that it obstructed their learning quality. Rather than exploring the space based on the pre-set learning goals, these participants appeared to forget the original learning aim and completely indulged in *playing with* the virtual objects (e.g., treating cultural objects as toys, repeatedly throwing them and picking them up). Second, the pairs in the IVR condition had a poorer awareness of the presence of their partner, even though the system enabled them to view their partner's avatar and movement and the two were physically in the same space. All pairs frequently checked their positions with their partner. This may be because each user could independently jump between the learning scenes (infinite space levels) and explore various locations, so it was easy to lose sight of their partner. This poor awareness may reduce the chance of prompting emergent dialogs around the appreciation of the traditional painting that we had expected to occur during the learning process.

The learning behaviors observed in the MTT condition were as follows. All pairs started to view the painting from right to left (a common way to view traditional Chinese paintings) by scrolling the screen. The participants interacted with the hotspots to check the audio–visual information associated with the painting, but half of the pairs (np = 14) merely scanned the texts quickly and were too impatient to listen to the entire audio piece (e.g., skipped). The pairs had better awareness of presence and shared view than in the IVR condition, as they were standing together in front of a large display. This resulted in more verbal communications such as information sharing (e.g., "*Look here, this man is bigger than others!*") and discussion (e.g., "*Can you check my gesture is correct?*" and "*The colors are simple and elegant.*"), but they related to their personal life-experience. We observed that half of the participants (n = 28) used embodied actions to facilitate their learning, as they did in IVR. For example, they imitated the cross-hand salute gesture when learning that knowledge point. In contrast to the IVR condition, in which both participants had equitable interactions with the system, around 70% of the pairs (np = 19) in the MTT condition had one *dominant* participant who was in charge of the main interaction (e.g., selecting the interactive points and scrolling), while the other acted as a *facilitator* who verbally responded but rarely interacted with the system (e.g., only scrolled the screen).

5.4. Learning motivation

The questionnaire results showed that the participants found the learning materials in both systems useful and that the collaborative learning mode was beneficial for improving learning motivation. This may be because the participants were initially unfamiliar with both paintings and the associated knowledge. However, the participants in the IVR condition had significantly higher median scores regarding intrinsic learning motivation and interests compared to the MTT condition (Table 5). Cronbach's alpha for all items of Learning Motivation (11items, IVR-alpha = 0.873, MTT-alpha = 0.894) and each of the two subscales of Intrinsic Motivation (6 items, IVR-alpha = 0.771, MTT-alpha = 0.855) and Extrinsic Motivation (5 items, IVR-alpha = 0.762, MTT-alpha = 0.763) also demonstrated a high level of internal consistency.

The interviews confirmed that the participants were extremely interested in the IVR approach and provided further explanations. Almost all of the participants (n = 52) mentioned that learning with the IVR system made them forget about time passing and that it was like playing a game. Most of the participants (n = 36) claimed that the 3D immersive form enabled them to explore details of the historical artifacts, visualize abstract knowledge in vivid 3D forms, and freely explore the space as they wished, which enhanced their curiosity, learning motivation, and interest. Some specific comments were as follows:

> **F12-A**: *The VR learning system is very fun. [Using it is] like playing a game … [it is] very entertaining and very attractive.*

> **F21-B**: *It allows us to freely explore every detail of the painting in the scene, which is very interesting!*

> **M31-B**: *The VR learning system transformed the characters, buildings, and objects in the original painting into 3D form. The animated characters and the background music provided an immersive experience.*

> **M11-A**: *The fan could be picked up and used in the simulation, which felt very interesting. It let me observe the fan more closely.*

6. Discussion

When designing the IVR environment, we aimed to produce an authentic representation of the traditional Chinese painting, *Spring Morning in the Han Palace*, while staying loyal to the original painting. The goal was to open up opportunities for viewers to interpret the painting and learn the associated cultural–historical knowledge. By comparing participants'

Table 5. Learning motivation under IVR and MTT conditions.

Aspects	Learning motivation and interests	Mean (SD)		Median (IQR)		Wilcoxon signed-rank test (IVR-MTT)	
		IVR	MTT	IVR	MTT	z	p
Intrinsic motivation	I feel the system can enhance my learning interest.	4.0(0.67)	3.5(0.77)	4(0)	3(1)	3.165	0.002*
	The collaborative learning mode of this system can enhance my interest in learning.	3.8(0.87)	3.4(0.86)	4(1)	3.5(1)	2.439	0.015*
	It is impressive using this system for exploring the historical paintings (*reversed*).	4.1(0.73)	3.5(0.72)	4(1)	3(1)	3.935	0.000*
	The information discovered through the experience stimulated my curiosity (*reversed*).	4.2(0.63)	3.7(0.85)	4(1)	4(1)	3.495	0.000*
	Using this system has increased my interest in traditional Chinese paintings.	3.9(0.76)	3.7(0.82)	4(0.25)	4(1)	2.738	0.006*
	Using this system has increased my interest in traditional culture (*reversed*).	4.0(0.79)	3.7(0.82)	4 (1.25)	4(1)	2.717	0.007*
Extrinsic motivation	I feel the learning materials are useful to me.	3.6(0.71)	3.6(0.78)	4(1)	4(1)	−0.363	0.717
	The collaborative learning mode of this system can improve my learning motivation.	3.9(0.85)	3.6(0.83)	4(1.25)	4(1)	1.906	0.057
	The collaborative learning mode of this system can improve the effect of my learning.	3.6(0.80)	3.5(0.89)	4(1)	4(1)	0.937	0.349
	Overall, I think this system is a good learning tool.	4.1(0.65)	3.7(0.76)	4(1)	4 (0.25)	2.730	0.006**
	I wish that other historical paintings also adopt this system to facilitate my learning.	4.2(0.75)	3.8(0.72)	4(1)	4(1)	3.289	0.001*

learning effectiveness, learning behaviors, and learning motivation when using the IVR and MTT systems, we found that IVR systems are a promising approach for facilitating an appreciation of traditional arts and delivering cultural–historical knowledge. Learning with the IVR system resulted in better learning outcomes and greater learning motivation compared to learning with the MTT system, particularly in tasks related to recall of details or spatiality. This finding is consistent with previous research (Tost & Economou, 2009). Although the MTT system supported better learning efficiency, we expect that this result would only occur in similar contexts (that is, where participants are supervised in learning and/or asked to complete specific tasks).

The observational results suggest six IVR learning opportunities and design implications. First, the simulated 3D environment provides more self-exploratory choices than the 2D interface, so that users can choose how to access the learning materials according to their preferences and learning styles. This finding is consistent with previous research (Hanson & Shelton, 2008; Liu, 2020), but our result further reveals specific exploratory strategies either based on spatial positions (e.g., left-right, center-around) or learning logics (e.g., central learning contents of the five learning scenes, contextual learning information of the environment) and the social interactions between two collaborators (explorer and follower). The exploratory strategies that emerged can visualize the learning process of an individual student; for example, an educator can analyze the student's learning behavior to identify whether they are a *sequential* or *global* learner. These findings provide two implications for designing IVR learning environments. That is, designers could (1) leverage users' exploratory strategies to provide them with personalized guidance in virtual space; and (2) assign users different social roles to encourage their active decision making.

Second, the separate learning scenes create a *formal* learning space inside the larger exploratory (*informal*) space. By using this design strategy, we designed an implicit method of filtering irrelevant contextual information, focusing users' attention on the important learning content, and accelerating their learning efficiency through the relative stable interaction and learning flows (e.g., entered the scene; explored the audio–visual information; interacted with the virtual objects on the counter; compared the objects with those in the original painting; went back to the main scene). As reported in previous research (Champion, 2006; Sutcliffe, 2003), it is difficult to balance users' *cognitive load*[17] between navigation and learning in VR environments. This design strategy enables users the opportunity to self-explore the space while keeping their attention on the central learning task and eliminating extraneous information.

Third, flexible and meaningful interaction with cultural objects (e.g., rotating a certain object, listening to its audio knowledge interpretation, operating its certain functions) can purposefully direct users' attention to details of learning aspects and support a variety of embodied learning strategies through hands-on and whole-body actions. This kind of learning on one hand is owed to the high degree of interactivity in IVR environments (Concannon et al., 2019); on the other hand it is supported by IVR's spatial affordances for representing 3D artifacts in space (Mortara & Catalano, 2018). We observed embodied learning behaviors in both the IVR and MTT conditions.

In the MTT system, users' embodied behaviors were merely hand gestures (without physical objects), while in the IVR system, users tended to interact with and make sense of the virtual objects through whole-body strategies (e.g., comparing similar fans). Although interacting with virtual objects is not the same as interacting with physical artifacts owing to the lack of tactile and haptic senses; interestingly, the virtual nature can also remove the sanctity of the historical artifacts. Once users realized that these cultural artifacts could not be broken, they were confident in manipulating them as they wished. This created an intimate emotional connection between the artifacts and the users, as demonstrated by the desire of several participants to carry their favorite artifacts with them throughout the virtual world. Instead of being distant objects that were displayed but untouchable, as in the majority of museums and other forms of virtual display, the interactable objects took on a meaningful identity to the participants.

Fourth, immersion in the IVR environment stimulated users' empathy and facilitated the unconscious creation of connections with the virtual world. Previous research suggests that IVR spaces may facilitate the conceptualization of ideas through direct personal experience in a sensory-rich digital environment (Herrera et al., 2018). Our results further showed users tended to associate their IVR experience with their real-life experience, which may help to transfer the knowledge they learn in the IVR environment to real-life contexts.

Fifth, autonomous control over an IVR avatar ensures equitable interactions and encourages collaborative learning. Previous research suggests that MTT systems may not support the equitable participations of pairs in collaborative learning processes, unless specific design strategies or social configurations are used to enforce it (Fan et al., 2014). While using the IVR system, the autonomy of independent control ensures each user must interact with the system, which may help to ensure each user makes an equitable contribution.

Finally, learning with the IVR environment stimulated the *flow experience*[18] that is frequently found in playing games (Mirvis, 1991), although our design did not include any gamification or reward system. Previous research shows IVR environments encourage users to learn (Ibrahim & Ali, 2018; Mortara & Catalano, 2018), but our result further indicates the motivation could be attained without any game mechanism. This suggests that the immersive and interactive 3D environment intrinsically stimulated the learning motivation and flow experience, which may turn serious learning into an "edutainment" experience.

7. Limitations

We acknowledge that there are several limitations of this study:

- Short-term use of the system was tested; long-term effects of IVR learning outcomes could be further explored.
- Small number of learning tasks may have impacted learning efficiency results.
- Potential influence of personal characteristics and group dynamics on learning effectiveness and collaborative learning was beyond the limited scope.
- Study was conducted in controlled lab space; further study is necessary to understand the application of such technology in public spaces like museums.
- Participants' learning behaviors were mainly analyzed through the thematic analysis of observational and video data; future study can collect and analyze participants' gaze data and fine-grained interaction data (e.g., clicking and dragging) to provide a more sophisticated analysis.

8. Future work and conclusion

Future work in IVR as a learning tool would require us to expand future testing with a wider range of participants and with a more sophisticated methodology. We intend to improve in three aspects of our design: (1) design a step-by-step tutorial to reduce the initial learning curve of the IVR system; (2) integrate gamification or game-based instruction/feedback and rewards to encourage completion of tasks; and (3) improve the ability for partnered users to move with each other and see what the other is doing to increase collaborative learning efforts. After refining the process, we aim to conduct experiments with learnings in schools and community members in public spaces to improve upon the limitations we identified.

Our study contributed to the design strategies for reconstructing the IVR learning environment and identified potential benefits of using IVR systems to support young adults' collaborative learning of cultural–historical knowledge, compared to a similar MTT system. IVR environments have shown to significantly increase the learning effectiveness and motivation compared to the MTT system. The results suggest that (1) the high-level interactivity in the IVR environment enables users to develop a variety of embodied learning strategies with cultural artifacts to have equitable interactions in collaborative learning; (2) the high-level immersive environment allows users to explore the reviving historical places based on their learning styles and unconsciously relate their virtual experience with real-world experience to make meaning (reflection); and (3) the IVR environment supports users to imagine the details of the past architectures, artifacts and events which have not existed currently. Our unique design strategy of separated learning scenes may help users quickly filter irrelevant information and switch their attention to the central learning content (at the learning scenes) while also keeping the flexibility of their self-exploration (at the outside space). Further refining our IVR system would enable future students to take advantage of the system's affordances to develop a deeper appreciation of artistic, cultural, and historical knowledge.

Notes

1. A binocular head-based fully IVR system where two small screens display the virtual scene to each of the user's eyes (Muhanna, 2015).
2. A room-based fully IVR system composed by large projection screens that enables users to view a 3D effect with stereoscopic glasses (Muhanna, 2015).
3. "The degree a user associates being within a virtual environment." (Concannon et al., 2019, p. 2).
4. "The degree of accuracy and responsiveness a user' s actions represent when using the input hardware." (Concannon et al., 2019, p. 3).
5. "The extent of belief a user feels is within a virtual environment, despite knowing he or she is physically situated in another environment." (Concannon et al., 2019, p. 4).
6. See Hürst et al. (2016) and Yuan and Yun (2016) for more.
7. http://www.chinaonlinemuseum.com/painting-qiu-ying-han-palace.php
8. User experience refers to the quality of users' experience on interaction (i.e., system interaction, feedback and navigation) and their overall satisfaction of interacting with the systems (Huang et al., 2010).
9. Learning effectiveness refers to the users' task performance on cultural-historical knowledge tests and users' perceived learning effectiveness on learning with the systems (Lee et al., 2009).
10. Learning behaviors refer to the users' verbal and non-verbal interactional behavior that may be correlated with learning.
11. Learning motivation refers to the intrinsic motivation (i.e., performing an activity for its inherent interests and satisfactions) and extrinsic motivation (i.e., performing an activity to attain some separable outcomes) (Ryan & Deci, 2000).
12. The cognitive process are deeply rooted in physical body's interactions with the world (Wilson, 2002).
13. Information conveyed by both verbal (e.g., text and audio) and nonverbal (e.g., photographs and animation) modalities is easier to learn (Clark & Paivio, 1991). This theory provides a reasonable explanation for the positive learning outcomes of IVR involving multimedia.

14. Kivotos is a CAVE system composed by 4 back-projection screens in an area of 3 m^3 that enables a maximum of 10 users to view a 3D effect with stereoscopic glasses; Tholos is a CAVE theater made of a semi-spherical screen that contains 132 seats with each equipped with a joystick and four buttons (Tost & Economou, 2009).
15. See the online questionnaire for more: https://wj.qq.com/s2/7581516/643f/.
16. The study included selecting the reviewer, preparing for the validity test, setting up the questionnaire, and analysis whether the questionnaire was a valid instrument.
17. "The load imposed on an individual's working memory by a particular learning task." (Van Gog & Paas, 2012, p. 599).
18. "The holistic experience that people feel when they act with total involvement." (Csikszentmihalyi, 1975, p. 36).

Acknowledgments

We would like to thank Dongxu Yin, Peiyu Wu, Furong Zhang, Yun Wu, Hanyi Wang, Yeming Li, Jian Zhou, Qianchun Xu, and Ziqi Gong for facilitating the experiment. We would like to thank Sid Heeg for helping to edit the paper.

Funding

This work was supported by the National Key Research and Development Program of China under Grant [2020AAA0105200], the Zhejiang Provincial Administration of Cultural Heritage under Grant [202000015]; the National Social Science Fund of China under Grant [19ZD12], and the Fundamental Research Funds for the Central Universities under Grant [CUC210F003].

ORCID

Min Fan http://orcid.org/0000-0001-5323-8168

References

Antle, A. N., Warren, J. L., May, A., Fan, M., & Wise, A. F. (2014). Emergent dialogue: Eliciting values during children's collaboration with a tabletop game for change. *Proceedings of the 2014 Conference on Interaction Design and Children*, 37–46. Aarhus, Denmark. https://doi.org/10.1145/2593968.2593971

Antonietti, A., & Cantoia, M. (2000). To see a painting versus to walk in a painting: An experiment on sense-making through virtual reality. *Computers & Education*, *34*(3–4), 213–223. https://doi.org/10.1016/S0360-1315(99)00046-9

Barreau, J. B., Gaugne, R., Bernard, Y., Le Cloirec, G., & Gouranton, V. (2014). Virtual reality tools for the west digital conservatory of archaeological heritage. *Proceedings of the 2014 Virtual Reality International Conference on - VRIC '14*, 1–4. Laval, France. https://doi.org/10.1145/2617841.2617845

Bekele, M. K., & Champion, E. (2019). A comparison of immersive realities and interaction methods: Cultural learning in virtual heritage. *Frontiers in Robotics and AI*, *6*, 91. https://doi.org/10.3389/frobt.2019.00091

Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S., & Gain, J. (2018). A survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH)*, *11*(2), 1–36. https://doi.org/10.1145/3145534

Champion, E. (2006). Evaluating cultural learning in an online virtual environment. *International Journal of Continuing Engineering Education and Life-Long Learning*, *16*(3/4), 173–182. https://doi.org/10.1504/IJCEELL.2006.009197

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF01320076

Concannon, B. J., Esmail, S., & Roduta Roberts, M. (2019). Head-mounted display virtual reality in post-secondary education and skill training. *Frontiers in Education*, *4*, 80. https://doi.org/10.3389/feduc.2019.00080

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage.

Crystal Corporation. (2010) . *Who animate the Chinese symphonic picture riverside scene at Qingming festival*. Ming Bao Publishers.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety: The experience of play in work and games*. Jossey-Bass.

Dalgarno, B. (2002). The potential of 3D virtual learning environments: A constructivist analysis. *Electronic Journal of Instructional Science and Technology*, *5*(2), 3–6. https://www.researchgate.net/publication/216458729_The_Potential_of_3D_Virtual_Learning_Environments_A_Constructivist_Analysis

De Back, T. T., Tinga, A. M., Nguyen, P., & Louwerse, M. M. (2020). Benefits of immersive collaborative learning in CAVE-based virtual reality. *International Journal of Educational Technology in Higher Education*, *17*(1), 51. https://doi.org/10.1186/s41239-020-00228-9

Drossis, G., Birliraki, C., & Stephanidis, C. (2018). Interaction with immersive cultural heritage environments using virtual reality technologies. In C. Stephanidis (Ed.), *HCI international 2018 – posters' extended abstracts* (pp. 177–183). Springer International Publishing. https://doi.org/10.1007/978-3-319-92285-0_25

Economou, M., & Tost, L. P. (2008). Educational tool or expensive toy? Evaluating VR evaluation and its relevance for virtual heritage. In Y. E. Kalay, T. Kvan, & J. Affleck (Eds..), *New heritage: New media and cultural heritage* (pp. 242–260). Routledge. https://www.routledge.com/New-Heritage-New-Media-and-Cultural-Heritage/Kalay-Kvan-Affleck/p/book/9780415773560

Fan, M., Antle, A. N., Neustaedter, C., & Wise, A. F. (2014). Exploring how a co-dependent tangible tool design supports collaboration in a tabletop activity. *Proceedings of the 18th International Conference on Supporting Group Work*, 81–90. Sanibel Island, FL. https://doi.org/10.1145/2660398.2660402

Fassi, F., Mandelli, A., Teruggi, S., Rechichi, F., Fiorillo, F., & Achille, C. (2016). VR for cultural heritage–A VR-WEB-BIM for the future maintenance of Milan's Cathedral. In L. De Paolis & A. Mongelli (Eds.), *Augmented reality, virtual reality, and computer graphics* (pp. 139–157). Springer. https://doi.org/10.1007/978-3-319-40651-0_12

Flint, T., Hall, L., Stewart, F., & Hagan, D. (2018). Virtualizing the real: A virtual reality contemporary sculpture park for children. *Digital Creativity*, *29*(2–3), 191–207. https://doi.org/10.1080/14626268.2018.1511601

Greenwald, S. W., Kulik, A., Kunert, A., Beck, S., Fröhlich, B., Cobb, S., Parsons, S., Newbutt, N., Gouveia, C., Cook, C., Snyder, A., Payne, S., Holland, J., Buessing, S., Fields, G., Corning, W., Lee, V., Xia, L., & Maes, P. (2017). Technology and applications for collaborative learning in virtual reality. *Proceedings of the International Conference on Computer-Supported Collaborative Learning (CSCL'17)* (pp. 719–726). Philadelphia, PA. https://repository.isls.org/handle/1/210

Hanson, K., & Shelton, B. E. (2008). Design and development of virtual reality: Analysis of challenges faced by educators. *Journal of Educational Technology & Society*, *11*(1), 118–131. https://doi.org/10.1177/07356331177733995

Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zaki, J. (2018). Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PLOS ONE*, *13*(10), e0204494. https://doi.org/10.1371/journal.pone.0204494

Huang, H. M., Rauch, U., & Liaw, S. S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education*, *55*(3), 1171–1182. https://doi.org/10.1016/j.compedu.2010.05.014

Hürst, W., De Coninck, F., & Tan, X. J. (2016). Complementing artworks to create immersive VR museum experiences. *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*, 34. Osaka, Japan. https://doi.org/10.1145/3001773.3001806

Ibrahim, N., & Ali, N. M. (2018). A conceptual framework for designing virtual heritage environment for cultural learning. *Journal on*

Computing and Cultural Heritage (JOCCH)*, *11*(2), 11. https://doi.org/10.1145/3117801

Jin, S., Fan, M., Wang, Y., & Liu, Q. (2020). Reconstructing traditional Chinese paintings with immersive virtual reality. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–8. Honolulu, HI. https://doi.org/10.1145/3334480.3382934

Kersten, T. P., Tschirschwitz, F., Deggim, S., & Lindstaedt, M. (2018). Virtual reality for cultural heritage monuments – From 3D data recording to immersive visualisation. In M. Ioannides, E. Fink, R. Brumana, P. Patias, A. Doulamis, J. Martins, & M. Wallace (Eds.), *Digital heritage. progress in cultural heritage: Documentation, preservation, and protection* (pp. 74–83). Springer International Publishing. https://doi.org/10.1007/978-3-030-01765-1_9

Lee, E. L., Wong, K. W., & Fung, C. C. (2009). Learning effectiveness in a desktop virtual reality-based learning environment. *Proceedings of the 17th International Conference on Computers in Education*, 832–839. Hong Kong, China. https://core.ac.uk/download/pdf/11237833.pdf

Limniou, M., Roberts, D., & Papadopoulos, N. (2008). Full immersive virtual environment CAVETM in chemistry education. *Computers & Education*, *51*(2), 584–593. https://doi.org/10.1016/j.compedu.2007.06.014

Liu, Y. (2020). Evaluating visitor experience of digital interpretation and presentation technologies at cultural heritage sites: A case study of the old town, Zuoying. *Built Heritage*, *4*(1), 14. https://doi.org/10.1186/s43238-020-00016-4

Loizides, F., El Kater, A., Terlikas, C., Lanitis, A., & Michael, D. (2014). Presenting cypriot cultural heritage in virtual reality: A user evaluation. In M. Ioannides, N. Magnenat-Thalmann, E. Fink, R. Žarnić, A. Y. Yen, & E. Quak (Eds.), *Digital heritage. Progress in cultural heritage: Documentation, preservation, and protection* (pp. 572–579). Springer International. Publishing.https://doi.org/10.1007/978-3-319-13695-0_57

Ma, W., Wang, Y., Xu, Y. Q., Li, Q., Ma, X., & Gao, W. (2012). Annotating traditional Chinese paintings for immersive virtual exhibition. *Journal on Computing and Cultural Heritage (JOCCH)*, *5*(2), 6. https://doi.org/10.1145/2307723.2307725

Meerbeek, B., Bingley, P., Rijnen, W., & Van Den Hoven, E. (2010). Pipet: A design concept supporting photo sharing. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, 335–342. Reykjavik, Iceland. https://doi.org/10.1145/1868914.1868954.

Mirvis, P. H. (1991). *Flow: The psychology of optimal experience*. JSTOR.

Mortara, M., & Catalano, C. (2018). 3D Virtual environments as effective learning contexts for cultural heritage. *Italian Journal of Educational Technology*, *26*(2), 5–21. https://doi.org/10.17471/2499-4324/1026

Muhanna, M. A. (2015). Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions. *Journal of King Saud University - Computer and Information Sciences*, *27*(3), 344–361. https://doi.org/10.1016/j.jksuci.2014.03.023

Mulholland, P., & Collins, T. (2002). Using digital narratives to support the collaborative learning and exploration of cultural heritage. In *Proceedings. 13th International Workshop on Database and Expert Systems Applications* (pp. 527–531). Aix-en-Provence, France. https://doi.org/10.1109/DEXA.2002.1045951

Muntean, R., Hennessy, K., Antle, A., Matkin, B., Rowley, S., & Wilson, J. (2016). Design interactions in? Eləwkw: Belongings. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, 582–594. Brisbane, QLD, Australia. https://doi.org/10.1145/2901790.2901912

Rahaman, H., & Tan, B.-K. (2011). Interpreting digital heritage: A conceptual model with end-users' perspective. *International Journal of Architectural Computing*, *9*(1), 99–113. https://doi.org/10.1260/1478-0771.9.1.99

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Sagnier, C., Loup-Escande, E., & Valléry, G. (2020). Effects of gender and prior experience in immersive user experience with virtual reality. In T. Ahram & C. Falcão (Eds.), *Advances in usability and user experience* (pp. 305–314). Springer International Publishing. https://doi.org/10.1007/978-3-030-19135-1_30

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality | Nat Rev Neurosci. *Nat Rev Neurosci*, *6*(4), 332–339. https://doi.org/https://doi.org/10.1038/nrn1651

Sharples, S., Cobb, S., Moody, A., & Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): Comparision of Head-mounted display (HMD), desktop and projection display systems. *Displays*, *29*(2), 58–69. https://doi.org/10.1016/j.displa.2007.09.005

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, *3*, 74. https://doi.org/10.3389/frobt.2016.00074

song, D. (2017). *How to learn an ancient Chinese painting (怎样读一幅古代中国画)*. Heilongjiang Educational Publisher (黑龙江教育出版社).

Sun, R., Wu, Y. J., & Cai, Q. (2019). The effect of a virtual reality learning environment on learners' spatial ability. *Virtual Reality*, *23*(4), 385–398. https://doi.org/10.1007/s10055-018-0355-2

Sutcliffe, A. (2003). *Multimedia and virtual reality: Designing multisensory user interfaces*. Psychology Press.

Takács, B. (2011). Immersive interactive reality: Internet-based on-demand VR for cultural presentation. *Virtual Reality*, *15*(4), 267–278. https://doi.org/10.1007/s10055-010-0157-7

Tost, L. P., & Economou, M. (2009). Worth a thousand words? The usefulness of immersive virtual reality for learning in cultural heritage settings. *International Journal of Architectural Computing*, *7*(1), 157–176. https://doi.org/10.1260/147807709788549367

Van Gog, T., & Paas, F. (2012). Cognitive load measurement. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 599–601). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_412

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Wise, A. F., Antle, A. N., Warren, J. L., May, A., Fan, M., & Macaranas, A. (2015). What kind of world do you want to live in?: Positive interdependence and collaborative processes in the land-use planning game YouTopia. *Proceedings of Conference on Computer Supported Collaborative Learning (CSCL '15)* (pp. 236–243). Gothenburg, Sweden: ISLS Press. https://repository.isls.org/handle/1/413

Yuan, C., & Yun, Z. (2016). Tunable, a VR reconstruction of listening to a Guqin from emperor Zhao Ji. *SIGGRAPH ASIA 2016 VR Showcase*, Article No: 6, 2. Macau, China. https://doi.org/10.1145/2996376.2996379

Zheng, L., Xie, T., & Liu, G. (2018). Affordances of virtual reality for collaborative learning. *2018 International Joint Conference on Information, Media and Engineering (ICIME)*, 6–10. Osaka, Japan. https://doi.org/10.1109/ICIME.2018.00011

Zhu, Y. B., Li, C. J., & Shen, I. F. (2004). A new style of ancient culture: Animated Chinese dunhuang murals. *ACM SIGGRAPH 2004 Sketches*, *130*. Los Angeles, CA. https://doi.org/10.1145/1186223.1186386

About the Authors

Sheng Jin is an associate professor in the School of Animation and Digital Arts at the Communication University of Zhejiang. His research interest is interactive design for traditional culture and arts.

Min Fan is an assistant professor in the School of Animation and Digital Arts at the Communication University of China. Her research focuses on the design and evaluation of innovative technologies for supporting learning in various contexts.

Aynur Kadir is an assistant professor at the faculty of Communication Arts at the University of Waterloo, Canada. Her research conceptualizes the poetics and politics of interactive media in the representation of traditional knowledge, memory and cultural heritage, and contributes to the ethical use of new media through collaboration.

Appendix

The details of the test results can be downloaded at the https://drive.google.com/file/d/1pRpspKDuqMn503jGFkmZfUqSd55rq-tU/view?usp=sharing.