“Capture Your Experience at This Moment”: Collecting Concurrent User Experience Data in Immersive Virtual Environment

Mingming Wang
School of Design
The Hong Kong Polytechnic University
Hong Kong, China
mingming.wang@connect.polyu.hk

Henry Been-Lirn Duh
School of Design
The Hong Kong Polytechnic University
Hong Kong, China
henry.duh@polyu.edu.hk

Xiaoyu Zhang
Computer Science & Information Technology
La Trobe University
Melbourne, Australia
gysjzxy@gmail.com

Yihan Mei
School of Design
The Hong Kong Polytechnic University
Hong Kong, China
cherry-doppel.mei@connect.polyu.hk

Haohan Wang
School of Design
The Hong Kong Polytechnic University
Hong Kong, China
haohanwang0026@163.com

Haiyan Jiang
School of Design
The Hong Kong Polytechnic University
Hong Kong, China
jianghybit@163.com

Zhibin Zhou*
School of Design
The Hong Kong Polytechnic University
Hong Kong, China
zhibin.zhou@polyu.edu.hk

Abstract

Environmental User Experience (UX) data collection is essential for user research, enabling evidence-based design decisions. However, traditional retrospective methods like micro-phenomenological interviews suffer from recall inaccuracies and memory distortions. Concurrent UX data collection methods with environmental contexts are promising but lack in-depth investigation. To examine this potential, we conducted a formative study with 34 participants, identifying design goals such as natural interaction, in-situ annotation, and spatial-temporal coupling. We developed JourneyCapturer, an interactive tool that fulfills these goals to integrate concurrent annotation within Immersive Virtual Environment (IVE), enabling real-time UX data capture within contextual scenarios. Using a mixed-method design, we comparatively evaluated concurrent IVE annotations, retrospective interviews, and the combined method with 20 participants, demonstrating how JourneyCapturer improves UX collection processes and outcomes. Our findings suggest that a consciously proactive concurrent IVE method with a first-person perspective advances UX research, offering implications for expert collaboration, multi-modal analytics, and IVE-based field studies.

*Corresponding author.

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790362

CCS Concepts

• **Human-centered computing → Systems and tools for interaction design**.

Keywords

User Experience, Concurrent Collection, Immersive Virtual Environment

ACM Reference Format:
Mingming Wang, Henry Been-Lirn Duh, Xiaoyu Zhang, Yihan Mei, Haohan Wang, Haiyan Jiang, and Zhibin Zhou. 2026. “Capture Your Experience at This Moment”: Collecting Concurrent User Experience Data in Immersive Virtual Environment. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3772318.3790362

1 Introduction

User Experience (UX) data collection plays a critical component of the design process, directly facilitating the effectiveness of design decisions. Traditionally, designers rely primarily on retrospective methods such as micro-phenomenological interviews, where users attempt to recall and verbalize past experiences [13, 73, 77]. While these retrospective methods have been widely adopted in HCI research to collect lived experiences of interactive systems [21, 26] and VR environments [54], they face significant limitations, including memory loss and post-experience passive rationalization that compromises data authenticity, leading to flawed design decisions and iterative validation cycles [83, 85]. For example, users may forget critical pain points during interaction or unconsciously modify their recollections to create consistent narratives, resulting in design solutions that are difficult to address actual user needs. These limitations become more pronounced when collecting UX data in spatial environments, where retrospective methods may struggle to accurately capture nuanced insights [13]. In the scope of this research, we focus on environmental UX[1] data gathered from spatial environments, rather than general UX data derived from interactions with systems or digital services (e.g., apps, websites). Additionally, "concurrent" refers to capturing real-time UX data while users experience their journeys, rather than through post-experience recall.

Immersive Virtual Environment (IVE) technology offers promising solutions to address the challenges in retrospective methods. By creating immersive and authentic environments, IVE enables users to immerse themselves in simulated past journey scenarios, which can activate situational memory and enhance UX data accuracy through environmental cues [56]. More importantly, IVE supports concurrent in-situ annotation capabilities, allowing users to document their feelings, feedback, and decisions during the real-time experience [31]. This "experience-while-recording" approach can significantly reduce memory loss and post-experience rationalization issues inherent in traditional retrospective methods [46, 80]. Recent research demonstrates that VR technology can effectively elicit UX data comparable to that collected from the real world, across multiple dimensions, including emotion [60], cognition [52], sensation [51], and social perception [55]. Meanwhile, digital storytelling platforms [29] in VR can support reflection and documentation of travel memories.

However, critical aspects remain unexplored regarding how to effectively integrate the concurrent method with IVE into the UX data collection. Existing research has primarily focused on utilizing IVE itself to simulate an immersive environment for experiencing journeys, but still leverages traditional retrospective interviews to collect data [5, 48], rather than exploring proactively concurrent data collection within IVE-based capability. These gaps highlight the need for further investigation of IVE's capabilities in optimizing the concurrent UX collection process and outputs. In this scope, we investigate the following research questions (RQs). ***RQ1****: How do designers expect the concurrent method to facilitate UX data collection within IVE?* ***RQ2****: How does the concurrent IVE method impact the collected user data, in comparison to the retrospective IVE method?* ***RQ3****: How do users perceive the concurrent IVE method for UX data collection process, in comparison to the retrospective IVE method?*

To answer these questions, we followed the Research through Design paradigm [93]. In particular, we conducted a three-stage approach for incorporating the insights from participants to develop and evaluate an IVE-empowered tool named JourneyCapturer that assists participants in concurrently and proactively collecting UX data. First, we conducted semi-structured interviews as the formative study with 34 participants (9 design practitioners, 8 researchers, 17 journey experiencers) to gather their demands in the UX data collection process and their expectations for IVE-empowered concurrent collection tools. Based on these insights, we summarized five design goals for a proactively concurrent method and developed the initial prototype. Second, six experts participated in a co-design workshop to conduct heuristic evaluations and iterations with specific functional requirements, which resulted in the final JourneyCapturer prototype that incorporates a progressive four-step UX collection process. Third, to evaluate JourneyCapturer, we conducted a comparative study between the traditional retrospective method (micro-phenomenological interview), concurrent method (JourneyCapturer's concurrent annotation), and combined method (concurrent method and then retrospective method), where 20 participants completed three UX data collection tasks separately.

Our research on integrating concurrent IVE method with UX data collection makes three significant contributions to the field of design, organized according to Wobbrock's HCI taxonomy [84]: (1) ***Artifact***: This research presents JourneyCapturer as a novel design exemplar, a concurrent collection tool for UX data within IVE, enabling spatial-temporal coupled UX data capture through environmental anchoring and immediate structured annotation capabilities, and demonstrating consciously proactive collection with first-person experiential access. (2) ***Empirical***: Through the mixed-method investigation using JourneyCapturer as a research probe, we offer empirical insights revealing trade-offs between data authenticity and collection efficiency in proactive concurrent IVE methods. These insights inspire future directions for complete UX data analytics, including enhanced gestural annotation, multi-modal data integration, etc. (3) ***Methodological***: The study incorporates the "consciously proactive concurrent UX collection" method in IVE, and further proposes UX collection guidelines, shifting passive data collection into immediate context-aware practice. The guideline provides three hierarchical levels for collecting UX data: establishes a data collection strategy at the macro-level, segments experiences into meaningful units at the meso-level, and executes proactively concurrent data capture at the micro-level.

2 RELATED WORK

2.1 Retrospective Methods and Concurrent Methods for UX Data Collection

Traditional UX data collection primarily relies on retrospective approaches that require users to recall and articulate their past experiences after completing their journeys or experience sessions. Interview-based methods serve as a primary retrospective approach, encompassing semi-structured interviews [1] and micro-phenomenological interviews [77] that guide users in verbalizing detailed experiential contents [13]. These interview approaches focus on helping users access specific experiential moments rather than making analytical judgments [50]. Survey-based methods provide a quantitative complement through standardized questionnaires that measure predetermined UX dimensions [9, 36], though they constrain the expression of experiential richness and nuance [38, 68]. Longitudinal approaches include diary studies, where users document experiences through written or digital logs over extended periods [15], requiring substantial time investment while still relying on the retrospective reconstruction of past events [30]. Additionally, journey mapping has emerged as a popular UX documentation method, where users reconstruct their experiences through structured visualization of touchpoints, emotions, and pain points across temporal sequences [41, 78].

[1]Throughout this paper, we refer to "environmental UX" as "UX" to avoid repetition.

Concurrent data collection methods aim to capture UX data concurrently during interaction with systems or environments. For example, Hertzum's meta-analytic review [25] demonstrates that concurrent thinking aloud enables concurrent access to users' thoughts during task performance, though facing reactivity challenges where continuous verbalization may alter natural user behavior. Comparative studies reveal that while concurrent methods detect more usability problems than retrospective methods, concurrent methods may impose cognitive burden by requiring simultaneous task performance and continuous verbalization [2]. Additionally, the "memory-experience gap" shows that concurrent emotion ratings differ significantly from retrospective assessments, particularly in the case of unpleasant experiences [11]. Observation-based methods provide another concurrent approach, where researchers, from a third-person perspective, directly document user behaviors through participant or non-participant observation [14, 17]. Similarly, the shadowing method [19] can capture naturalistic interactions as they unfold. However, the shadowing method typically still requires post-event elicitation to access meanings [18, 40]. Physiological monitoring represents a technology-enhanced concurrent method that captures unconscious responses during interaction through sensors measuring heart rate, galvanic skin response, and eye movements, though they primarily focus on behavioral indicators rather than subjective experiential dimensions [42].

Despite widespread adoption, both retrospective and concurrent methods present inherent limitations that compromise UX data authenticity. Retrospective methods suffer from memory loss and post-experience rationalization. Meanwhile, concurrent think-aloud introduces cognitive burden through passive verbalization that may generate excessive irrelevant information and disrupt natural interaction flows. However, existing concurrent methods for UX data collection during original experiences (e.g., observational methods) remain constrained by third-person perspectives, missing crucial internal experiential processes. These limitations highlight the need for novel approaches that capture experiential data consciously and proactively from a first-person perspective. Our approach addresses this by enabling concurrent documentation of users' journeys during IVE experiences, eliminating reliance on post-experience recall or disruptive verbalization during original experiences. This gap drives our investigation into RQ1 to understand expectations and requirements for proactive concurrent IVE methods.

2.2 IVE Systems for visual annotation and experiential documentation

Virtual visual annotation systems enable concurrent experience capture through diverse spatial, visual, and interactive modalities that support immediate user feedback and contextual documentation. Body-anchored annotations position information relative to the user's embodiment, creating persistent spatial relationships [4, 62, 81]. Textual annotations, including speech-to-text conversion, provide detailed contextual capture for concurrent experience documentation [49]. Iconic annotation systems utilize symbolic representations to enable rapid input during active VR engagement without disrupting immersive experiences [59, 90]. Static annotations preserve temporal snapshots of user states and environmental contexts, creating persistent records for retrospective analysis [12, 65]. These annotation functions establish foundational methods for capturing UX data during active VR engagement rather than through post-hoc reconstruction methods.

IVE systems leverage VR's embodiment and presence to enable immersive construction of personal journey narratives through spatial storytelling. VR's core functions of embodiment and spatial presence transform how individuals construct narratives by enabling physical movement through IVE as a storytelling medium [61]. The medium supports "storyliving" experiences where users actively navigate spatial representations rather than consuming pre-constructed narratives [75]. Besides, VR creates immersive, diegetic spaces like "democratic surrounds" that facilitate learning through personal perspectives and spatial exploration [23, 74]. Additionally, the spatial abstraction inherent in VR environments supports sequences and temporal flows that guide narrative structure, enabling environmental storytelling approaches where space itself becomes a vehicle for expressing temporal progression and emotional journeys [3, 72, 79].

Despite VR's established capabilities for concurrent annotation and immersive narrative construction, existing personal journey documentation systems primarily rely on retrospective reconstruction rather than integrating concurrent methods. This gap represents an opportunity to combine VR's annotation functions with spatial storytelling capabilities to achieve authentic documentation of immediate emotional responses and contextual nuances during personal experiences. The integration of concurrent annotation methods into personal journey documentation could enable more complete capture of experiential data as it emerges, prompting our investigation of RQ2 to evaluate the effectiveness of proactively concurrent data collection in IVE.

2.3 IVE Applications in UX research and Data Collection

IVE technology primarily functions as experience prototypes or narrative galleries in the UX data collection process, creating controlled environments that simulate real-world scenarios with high fidelity [5, 16] to enable the collection and analysis of UX data in reproducible conditions [86, 94]. VR prototypes can effectively elicit UX responses comparable to real environments across multiple dimensions, including emotion, cognition, and social perception [57]. For example, high-fidelity architectural prototypes for experiencing patient journeys in healthcare settings, digital storytelling platforms supporting travel memory reflection [87], and multi-agent conversational systems for reminiscence experiences among older adults [66]. This approach allows researchers to study user responses in reproducible conditions through immersive simulation.

Despite theoretical frameworks recognizing VR's potential for enhancing UX evaluation, current IVE applications rely on retrospective data collection that creates temporal separation between experience and evaluation [39, 76]. For example, Rebelo et al. [55] proposed a VR-based UX evaluation framework, identifying VR's advantages in providing controlled experimental conditions while maintaining ecological validity and enabling manipulation of contextual factors that significantly influence UX. However, contemporary IVE studies predominantly employ retrospective evaluation protocols by administering standardized questionnaires and interviews after immersive sessions [35, 39, 76]. Additionally, advanced multi-modal approaches that integrate physiological signals with subjective measures require retrospective interviews to enable conscious experiential interpretation, despite their ability to detect emotional responses during VR engagement [47].

Although IVE has proven effective for conceptual validation of VR-based UX evaluation frameworks, current applications only utilize IVE as simulated environments, followed by retrospective interviews. These applications miss opportunities to capture authentic experiential data through conscious documentation during the immersive experience itself. This represents a critical methodological gap where VR's empirical potential for concurrent UX data capture compared to traditional retrospective approaches remains underexplored, particularly in enabling consciously proactive experiential documentation with a first-person perspective. The gap motivates our RQ3 to understand user perceptions of concurrent methods compared to retrospective methods in IVE.

3 FORMATIVE STUDY

3.1 Method

To address RQ1, our formative study identified the five design goals that participants expect from the concurrent IVE approaches for UX data collection (Section 3.2). Based on these findings, we developed a design framework that integrates participant expectations throughout a four-stage UX data collection workflow (Section 4.2), which subsequently guided the implementation of our JourneyCapturer system.

Participants. We recruited 34 participants (9 design practitioners, labeled D1 - D9; 8 researchers, labeled R1 - R8; 17 participants with experience in target environments, labeled E1 - E17; aged 22 - 38) through campus networks and online public platforms, with 24 females and 10 males. Participants' demographic information is presented in Appendix A Table 5. All design practitioners and researchers had at least 4 years of professional experience in UX data collection. Additionally, all journey participants were required to have experienced both hospital and subway environments within the past three months. Each interview lasted between 40 and 80 minutes, with participants receiving an incentive of CNY 100.

Setup. We implemented the IVE using Unity (6000.0.26f1) and Meta Quest 3 hardware, featuring scenario simulations and core interaction capabilities such as movement systems and dedicated UX data capture interfaces. Throughout the sessions, we documented participants' operations through both audio and video recordings to support subsequent data analysis.

Procedure. We conducted individual online sessions by Zoom, commencing with an overview of our research objectives and procedures. Participants initially engaged in semi-structured interviews discussing their challenges, insights, and expectations regarding current UX data collection methods. Subsequently, we presented video demonstrations of our VR prototype to participants while using the think-aloud protocol to capture their cognitive processes and reflections about the insights of the concurrent IVE method and its tool.

Data Analysis. All interview sessions were recorded and fully transcribed to enable complete analysis. We employed an inductive thematic analysis to analyze participants' interview transcripts about their expectations, identified challenges, and desired functionality for incorporating concurrent IVE approaches into the UX data collection. These insights were then synthesized into '*Design Goals*' relevant to RQ1 to guide system development. For example, participants emphasized expectations for realistic scenarios with ambient sounds and expressed challenges with environmental distractions and desired minimized interfaces to enhance immersive focus, which were thematically synthesized into design goal - G1.

To ensure analytical rigor, two researchers independently coded the data. Following the initial thematic analysis, they engaged in collaborative discussions to inductively develop the coding framework (i.e., design goals), ensuring both consistency and completeness. To ensure reliability, a third researcher subsequently reviewed the coding framework and integrated divergent perspectives before confirming the emerging codes. This collaborative process ultimately yielded agreement on five key design goals that would inform the development of our concurrent IVE method through JourneyCapturer, as detailed in Section 3.2.

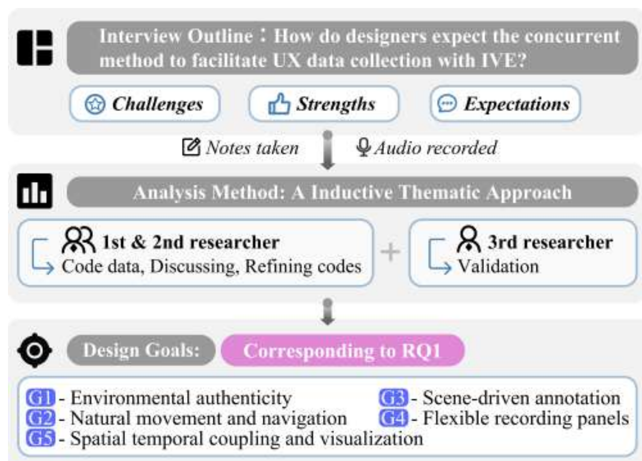

Figure 1: The interview procedure and data analysis methods of formative study.

3.2 Design Goals

This section synthesizes insights to address ***RQ1*** by presenting design goals derived from interview scripts, exploring designers' expectations regarding how concurrent and retrospective methods enhance UX data collection within IVE. The following design **G**oals (***G1 – G5***) provide support for concurrent IVE method (see Fig. 1).

G1 - Environmental authenticity. This goal aims to enhance sustained focus by minimizing external distractions and interruptions concurrently. Participants emphasized the importance of realistic scenarios with adjustable ambient sounds that don't disturb the collecting and thinking processes. As participants noted, "*concurrent collection with visuals and sounds is more immersive*" (E7, E17) and "*the real journey's environment helps simulate and recall my memory*" (E3, E11). The concurrent system should provide immersive focus modes with "*virtual pathways*" (R8) and "*time panels*" (R3) while "*minimizing interface panels*" (D2, D9) to strengthen environmental presence and reduce cognitive load during UX data collection process.

G2 - Natural movement and navigation. This goal facilitates an authentic journey experience while maintaining comfort and orientation in concurrent UX collection process. The concurrent system should provide movement functions, as participants noted that "*I want to move in the IVE to different locations*" (E4, E12). Additionally, teleportation-based movement could "*avoid motion sickness and accommodate diverse mobility preferences without speed adaptation requirements*" (R4). Some participants also expressed concerns regarding navigation issues (E6, E9, E16). Navigation features include "*ground-based pathway indicators showing available spaces*" (E8) and "*progress visualization from start to endpoint*" (R5) while preserving participants' freedom to explore autonomously.

G3 - Scene-driven in-situ annotation. This goal focuses on improving the quantity and quality of UX data by enabling "*direct recording*" (E2) through "*point-and-click scene object selection*" (E5, E13) in the specific IVE's locations. Features include "*concurrent voice transcription*" (D6), "*annotation tools*" (D2), and "*drawing capabilities*" (D1, D5) while minimizing keyboard input and menu operations to maintain seamless data collection processes. R6 mentioned the importance of "*transcription quality*" to ensure the accuracy of recorded UX data, suggesting that "*integrating LLMs can enhance this quality*" (D3).

G4 - Structured and adaptable organization of UX data. This goal ensures structured and usable UX data through adaptable recording interfaces within IVE. The concurrent system should provide virtual floating panels for UX collection dimensions, including Phases, Scenarios, Actions, Touch Points, Emotions, Pain Points, and Opportunities (Suggestions). E1 and E10 emphasized that "*It is better to edit the virtual panels' content*". Participants should be able to adjust recorded content through "*multiple-time recording*" (E15), "*editing*" (D6), and "*deletion*" (D4, D8) functions, ensuring flexibility while maintaining the integrity and organizational structure of collected UX data throughout the concurrent collection process.

G5 - Spatial temporal coupling and visualization. This goal enables "*ready-to-use*" data by providing inherent structure, shortening the process from UX data collection to organization. The concurrent system should automatically bind "*timestamps*" (R7) and "*spatial anchors*" (R1) for journey phases' area identification, creating dual indexing through "*temporal trajectories and spatial distribution*" (D8). For example, semi-automatic phase categorization allows hospitals to define consultation phases, with recorded UX data automatically categorized accordingly in concurrent, exporting tables of structured collected UX data with preserved spatial-temporal evidence chains. That is to say, timestamps and spatial anchors are automatically bound to create dual indexing through temporal sequences and spatial distribution.

4 DESIGN OF JOURNEYCAPTURER

4.1 Co-design Workshop of JourneyCapturer

To fulfill the design goals (***G1 - G5***) and meet the specific functional expectations for a concurrent IVE tool, we organized a design workshop with six participants (two UX researchers, two design practitioners, and two design educators), see Table. 1. All participants possessed a minimum of six years of professional expertise with UX data collection. The workshop focused on a VR prototype of a subway station environment, featuring fundamental VR navigation capabilities and a template panel for UX collection. During the session, participants verbalized their thoughts and observations using the think-aloud method while reviewing video demonstrations of the prototype. Subsequently, the participants engaged in a group discussion to elicit critique and suggestions regarding the prototype's design and functionality. Throughout the session, participants drew upon their field experience to evaluate specific feature requirements and conceptualize novel interaction paradigms for concurrent UX data capture within IVE. We documented all feedback, including suggestions for interface simplification to reduce cognitive load, enhanced recording indicators for better user awareness, and streamlined interaction workflows to improve overall usability. For example, experts emphasized the need for minimalistic interface elements to maintain immersion, refining G1, which led us to design the floating panels with reduced visual complexity and implement the *"BACK"* function for scalable viewing. These professional insights were incorporated into the iterative refinements of JourneyCapturer, maintaining consistency with both practitioner expectations and our research objectives. The details of how we summarize and implement the collected feedback are provided in Fig. 2.

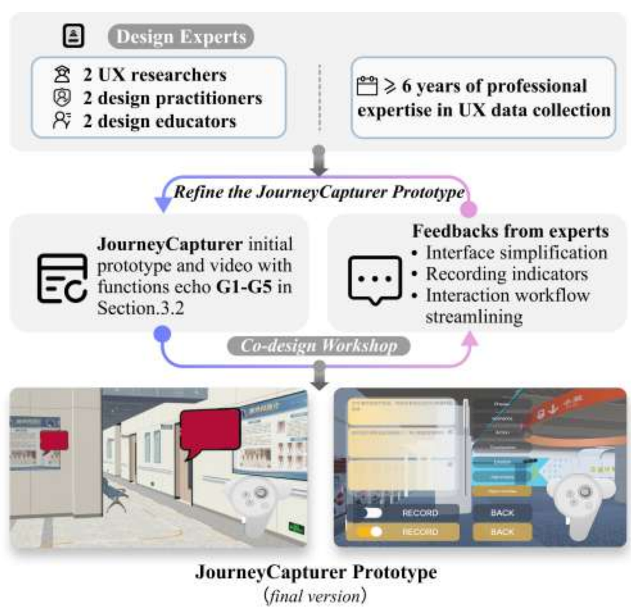

Figure 2: The co-design process of JourneyCapturer.

Table 1: The demographic information of the experts who participated in the JourneyCapturer co-design session.

Participant	Age	Gender	UX Collection Exp./Years	VR Exp./Years	Frequency of Using VR
UX Researcher 1	32	F	5	11	Almost every project used
UX Researcher 2	27	M	5	9	Almost every project used
Design Practitioners 1	30	F	7	7	Most projects used
Design Practitioners 2	29	M	8	6	Most projects used
Design Educator 1	31	F	12	10	Almost every project used
Design Educator 2	29	F	10	8	Most projects used

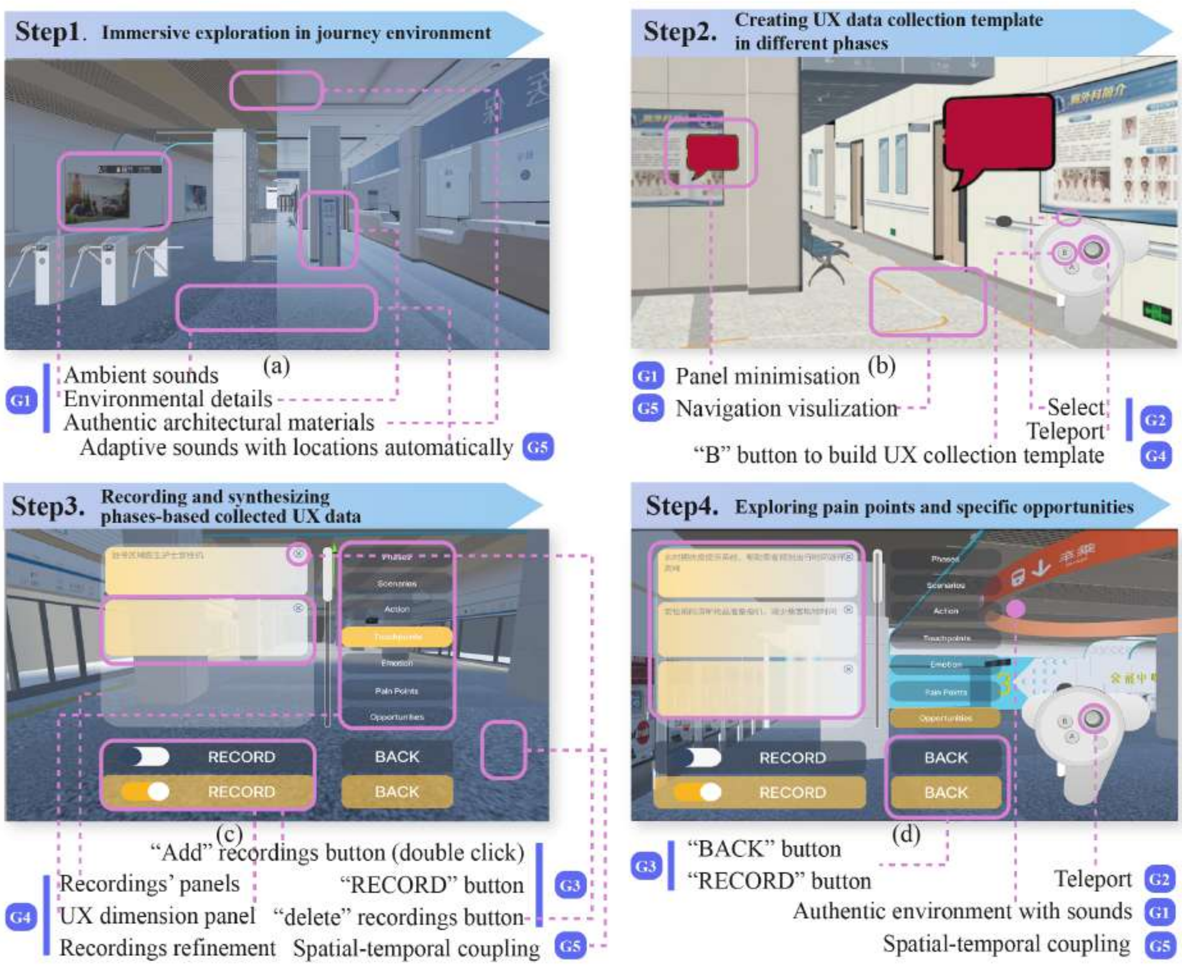

Figure 3: The detailed use scenario of the four UX collection steps with JourneyCapturer.

4.2 Description of JourneyCapturer

JourneyCapturer was developed using Unity Version 6000.0.26f1 and deployed as a standalone APK application on Meta Quest 3 headsets for the user study. We tested the system on a Windows computer equipped with an ASUS TUF RTX5090 32GB graphics card. At last, JourneyCapturer ran locally without requiring external connectivity during testing.

For *the development of immersive environments*, we created the JourneyCapturer system utilizing Unity and Meta Quest 3 to construct authentic immersive scenarios addressed G1. For *the implementation of virtual interaction*, we addressed G2 - G4 by developing interaction and navigation capabilities, incorporating UX collection templates derived from journey mapping dimensions with interactive virtual panels for collecting UX data. The panel includes voice recording and content deletion capabilities, with the interfaces providing scalable viewing through a *"BACK"* function to maintain immersion while ensuring efficient information access. The *backend server* addresses G5 by managing UX collection template management and handling interaction data, incorporating OpenAI's Whisper for speech-to-text transformation with GPT-4o mini performing content optimization and contextual validation. The FastAPI database framework supports cross-platform data management, archiving essential information including timestamps, content, coordinates, and user identifiers to ensure spatial-temporal coupling.

JourneyCapturer provides a progressive four-step UX collection process, designed to fulfill the expected design goals (G1 – G5) through the system features.

Step1. Immersive exploration in scenario environment (see Fig. 3. a). This step establishes an authentic foundation for UX data collection through immersive environmental recreation. JourneyCapturer enables participants to navigate through a realistic IVE constructed with authentic architectural materials and environmental details (echoes G1), such as the signage indicating the direction of the underground railway within the subway station. JourneyCapturer provides contextually appropriate ambient sounds that automatically adapt based on participants' spatial movement within the IVE (echoes G5), such as crowd noise at subway security checkpoints transitioning to train announcements at waiting platforms. This environmental authenticity enhances subway journey immersion (echoes G1) while minimizing external distractions, creating optimal conditions for subsequent UX data collection activities within spatially contextualized journey experiences.

Step2. Creating UX data collection template in different phases (see Fig. 3. b). This step enables participants to establish structured data collection frameworks at relevant journey phases through spatial exploration. JourneyCapturer facilitates movement through teleportation using the right controller's joystick (echoes G2) with ground-based visual pathways indicating start and end points (echoes G5). When participants identify significant locations during their journey experience, they can direct the controller's ray toward desired scene objects and activate the *"B"* button to trigger UX collection template creation (echoes G3, G4). JourneyCapturer supports minimising the template panel as a button to maintain environmental immersion (echoes G1), enabling participants to establish organized collection points while preserving the authentic spatial context essential for effective UX data capture.

Step3. Recording and synthesizing phase-based collected UX data (see Fig. 3. c). This step facilitates UX data capture through structured recording interfaces positioned at established collection points. JourneyCapturer provides UX collection templates featuring dimension panels on the right side, enabling participants to record individual dimension information (echoes G4) through *"RECORD"* button activation (echoes G3). Double-clicking *"RECORD"* allows additional recording creation (echoes G3). The left-side recordings panel displays transcribed content for each UX collection dimension, with *"delete"* button functionality for data management (echoes G3, G4). This structured approach ensures UX data collection while maintaining spatial-temporal coupling (echoes G5), enabling participants to efficiently document journey experiences with organized, contextually anchored UX insights across multiple dimensions.

Step 4. Exploring pain points and specific opportunities (see Fig. 3. d). This step promotes analysis by connecting collected UX data with environmental contexts to identify potential improvements to the overall journey. JourneyCapturer maintains teleportation movement and visual pathway guidance (echoes G2, G5), enabling participants to navigate between collection points while reviewing their documented UX data. JourneyCapturer preserves authentic IVE and immersive background audio (echoes G1, G5) to maintain contextual awareness during analysis. Participants utilize existing UX collection templates to record pain points and opportunities through *"RECORD"* button and manage content through *"delete"* functionality (echoes G3, G4), facilitating identification of self-insights rooted in spatial-temporal journey contexts.

5 USER STUDY

The purpose of this research is to investigate the potential of applying IVE for consciously proactive collecting concurrent environmental UX data with a first-person perspective, primarily by comparing concurrent methods to retrospective ones. We selected a micro-phenomenological interview, a retrospective method focused on recalling experiential memory, as our baseline to evaluate the concurrent IVE method, JourneyCapturer, ensuring a fair comparison since both methods involve conscious, proactive information sharing rather than passive verbalization. We then conducted a within-subjects study comparing JourneyCapturer with the retrospective method to explore the advantages and concerns introduced by concurrent annotation with spatial context in the UX data collection process, using Journey mapping content as the task.

5.1 Setup

Participants used VR to experience their journeys through realistic environmental simulations, which triggered their authentic memories and allowed them to document these experiences in context, rather than relying on hypothetical scenarios to recall. We selected two representative scenarios from daily life for our UX collection tasks: subway journeys and hospital examination processes. For both conditions, we provided identical UX collection templates with consistent recording dimensions, including Phases, Scenarios, Action, Touch Points, Emotion, Pain Points, and Opportunities. The same UX collection content was implemented as a virtual floating panel version in the concurrent condition and a micro-phenomenological interview version in the retrospective condition. In both scenarios, participants employed a think-aloud method while recalling their journey experiences. However, there was a key difference between conditions. In the concurrent condition, participants received preliminary VR training and independently recorded UX data using JourneyCapturer. In the retrospective condition, designers collected UX data by conducting micro-phenomenological interviews with participants.

In the ***concurrent condition***, we utilized the Meta Quest 3 headset with controller-activated functions in IVE, maintaining visual route guidance and audio transcription in the UX collection template panel. In the ***retrospective condition***, to balance the spatial information across conditions, participants first explored the IVE using the Meta Quest 3 headset. They then orally described their experiences in a micro-phenomenological interview conducted by the designers. In both conditions, participants worked within a 2.5 × 2.5 meter tracking space, preserving a consistent balance among materials, experimental setting, and the collection process.

5.2 Participants

We recruited twenty participants (P1 - P20) comprising 8 males and 12 females, with ages ranging from 18 to 30 years (M = 24.75, SD = 2.57; detailed demographics in Appendix B Table 6) by institutional announcements and social media platforms. All participants had full cognitive and sensorimotor capacities. Additionally, participants had no medical conditions contraindicating VR use and no severe vision problems that would make the VR experience impossible or excessively challenging. We conducted the study in person and video recorded each session. We also captured screen recordings of tool usage during the sessions. Each experimental session lasted approximately 2.5 - 3 hours, and participants received vouchers valued at HKD 200 upon completion.

5.3 Procedure

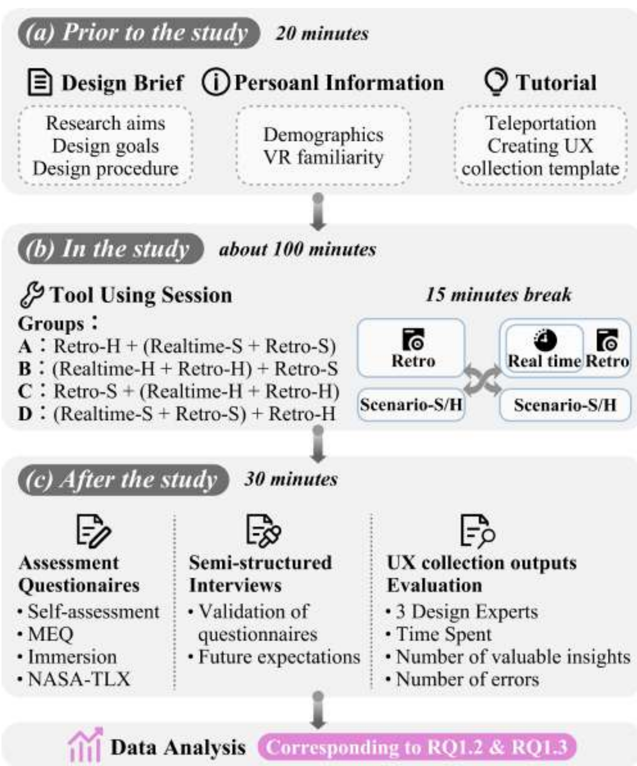

Figure 4: The procedure of user study.

The study incorporates three consecutive stages (see Fig. 4) to assess both the concurrent IVE method and the retrospective method.

We conducted the user study using a within-subjects design. Prior to the study, we conducted 10-minute preliminary interviews with each participant for background information (see Fig. 4. a). Participants completed a pre-experiment questionnaire about their demographics and VR familiarity. Then, the session began with VR equipment calibration and basic operational training in JourneyCapturer. Participants practiced essential tasks, including content creation, environment navigation, and interactive element manipulation. Participants explored specific functionalities of JourneyCapturer, rigorously testing each feature 1 - 2 times while navigating guided pathways, typically requiring 5 - 10 minutes. Consequently, the entire tutorial lasted approximately 10 minutes. Additionally, for the UX data collection template panel in three conditions, which organizes Phases horizontally, with elements such as Scenarios, Action, Touch Points, Emotion, Pain Points, and Opportunities arranged vertically, making it similar to three methods.

In the study, to evaluate the effectiveness of different UX data collection approaches and capture nuanced differences in data quality and completeness, we designed three distinct methods (see Fig. 5) that enabled precise comparison of concurrent with retrospective collection methods, identifying specific strengths and limitations of each approach and insights for our subsequent findings analysis. (1) ***Retrospective IVE method***: participants first experienced the IVE in JourneyCapturer without any UX data being captured during the session. Afterward, a designer conducted a micro-phenomenological interview to collect UX data retrospectively by guiding participants to recall and articulate their journeys' experiences. (2) ***Concurrent IVE method***: participants used JourneyCapturer to collect and record UX data concurrently within the IVE using UX collection templates. Then, the collected data were exported from JourneyCapturer. (3) ***Combined method***: building on the concurrent IVE method's procedures, the designer conducted a micro-phenomenological interview on the same journey scenario to elicit additional UX data retrospectively. For all three methods separately, a second designer oversaw the entire process and mapped the collected UX data into a simple Journey mapping in Figma for further evaluation of the collected errors and valuable insights within three methods. Only UX data mapping was performed without any design activities. We selected Figma because it is widely used by UX designers, and the adopted official template has been widely employed in numerous UX design projects.

We randomly assigned the participants into four equal groups (A, B, C, D; 5 participants per group), according to two representative scenarios (hospital and subway station) (see Fig. 4. b). We used a Latin square design to randomize the order of methods and scenarios used across groups and reduce sequence effects. For example, Group A completed the retrospective method with hospital journey, then the concurrent IVE method, and followed the combined method with subway journey, while Group B started the concurrent IVE method and then the combined method with hospital journey, and then the retrospective method with subway journey. This design ensured that each method-scenario combination was tested under comparable conditions and that all participants engaged with both methods and scenarios. To prevent fatigue effects, participants received 15-minute breaks between each condition, allowing sufficient recovery time before proceeding to the next experimental condition. For every three methods, we did not impose time limits, and each task ended when participants had no further experience data to provide. For the second designer's journey mapping task in Figma, the subway scenario required a journey mapping focused on metro wayfinding systems, and the hospital scenario required a journey mapping focused on hospital navigation systems.

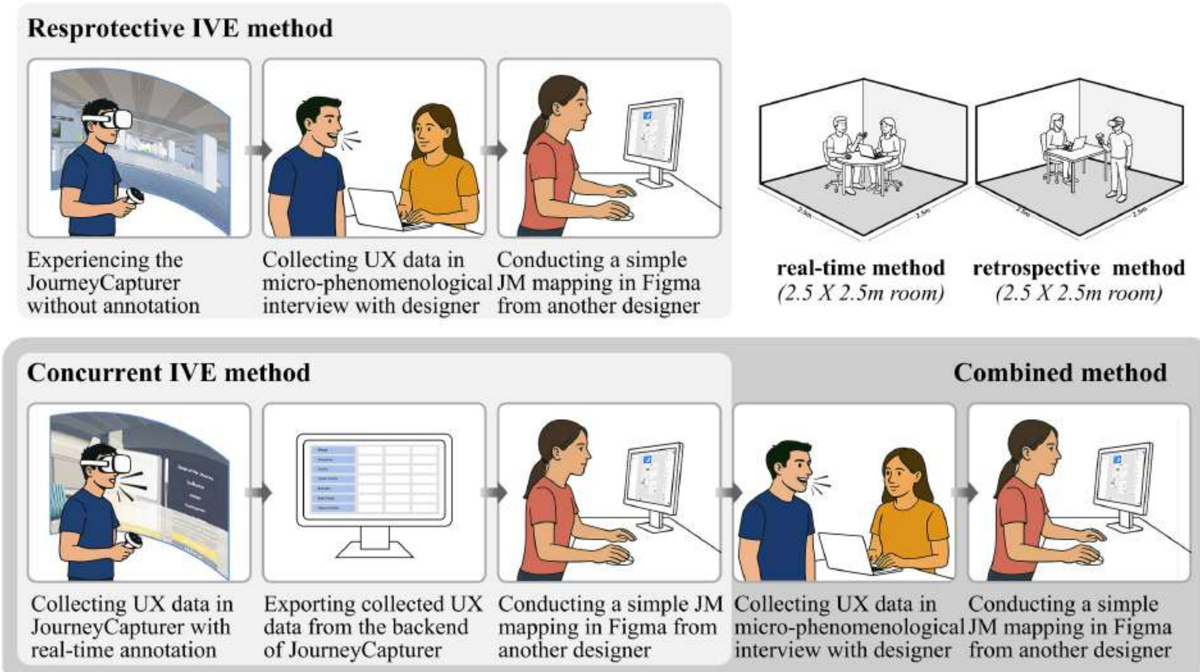

Figure 5: The user study procedure of three methods.

After the study, participants filled out a set of questionnaires to gather immediate feedback on their experiences and a post-study interview to understand the reasons behind any rating differences (see Fig. 4. c). For the quantitative evaluation, participants completed five standardized questionnaires to assess both tools: the Memory Experiences Questionnaire (MEQ) [67] assesses the phenomenological experience of autobiographical memories, the NASA Task Load Index (NASA-TLX) [24] to assess cognitive workload, the System Usability Scale (SUS) [9] measure the system usability. Additionally, Task Immersion, adapted from the Presence Questionnaire [82], is employed to assess immersion across various comparative experimental conditions. Inspired by established approaches for defining evaluation dimensions from prior UX research [41, 92], we developed specific self-assessment questionnaires to measure UX data collection performance in terms of satisfaction and goal achievement, see Table. 2.

5.4 Data analysis

We implemented data cleaning procedures to ensure the accuracy and reliability, including screening questionnaire outliers and recalibrating reverse-coded items for consistency across all data points.

For quantitative analysis, we first used Cronbach's alpha coefficient [34] to confirm the reliability of Immersion (0.838, indicating "good") and Self-assessment of UX Collection questionnaires (0.853, indicating "good"). Subsequently, a Shapiro-Wilk test confirmed that the ***Time Spent*** data were normally distributed while other variables were not. Accordingly, we conducted Repeated Measures ANOVA to compare Time Spent and performed the Friedman test to compare ***the Number of collected valuable insights***, ***self-assessment of UX collection questionnaires***, ***MEQ***, ***Immersion questionnaires***, and ***NASA-TLX*** across the three UX collection methods. To assess within-method consistency in the combined method, we quantified discrepancies between the concurrent IVE and subsequent retrospective reports as ***the collected UX data's errors***. We then performed a one-sample Wilcoxon signed-rank test to test whether the median error count differed from zero, where zero indicates perfect agreement. Additionally, ***SUS*** total scores were calculated following the standard SUS scoring procedure.

For qualitative analysis, we conducted a mixed deductive-inductive method. The deductive coding framework was derived from our five standardized questionnaires' dimensions. As participants justified their questionnaire ratings in the post-study interview, we applied inductive thematic coding to extract underlying reasons and emergent themes from these explanations. Then, the emerged codes were analyzed independently and mapped onto the 5-dimensional deductive framework when appropriate. For example, our qualitative theme "*Thematic relevance organization through spatial anchoring*" (Section 6.1) demonstrates a direct correspondence with the Relevance dimension of our self-assessment UX collection questionnaire, and the findings "*Enhanced memory vividness through multi-sensory stimulation*" (Section 6.4) directly correspond to the Vividness dimension of MEQ. To ensure analytical rigor, we implemented a dual independent-coding protocol: one researcher performed the initial coding, followed by a second researcher who independently reviewed and refined the codes. A third researcher then reviewed both coding sets, addressed any discrepancies, and discussed with

Table 2: The questionnaire of the self-assessment of UX collection and Immersion.

Self-assessment of UX collection	Completeness	I provided complete UX data that covers all UX collection dimensions when using concurrent IVE method/retrospective interview.
	Accuracy	The UX data I provided accurately reflects my actual experiences when using concurrent IVE method/retrospective interview.
	Clarity	My UX data expression is clear and easy to understand when using concurrent IVE method/retrospective interview.
	Consistency	My emotional responses align logically with the situations and experiences I described when using concurrent IVE method/retrospective interview.
	Relevance	I can effectively provide corresponding UX data based on the specific dimensions provided in the concurrent IVE method/retrospective interview.
	Insight	I provided valuable insights that reveal hard-to-detect observations with concurrent IVE method/retrospective interview.
Immersion	Expression Naturalness	It feels natural to express my UX data in the concurrent IVE method/retrospective interview.
	Exploration Freedom	I can autonomously explore and express UX data from different perspectives in concurrent IVE method/retrospective interview.
	Engaged Level	I am highly engaged during concurrent IVE method/retrospective interview.
	Overall Immersion	I feel fully immersed when capturing UX data during concurrent IVE method/retrospective interview.
	Cognitive Fluency	I can express UX data smoothly without frequent pauses to think about how to articulate during concurrent IVE method/retrospective interview.
	Interference Factors	External factors in concurrent IVE method/retrospective interview do not interfere with my UX data collection.
	Focus	I can focus on sharing UX rather than dealing with technical interactions in concurrent IVE method/retrospective interview.

two researchers to reach consensus. We established the final thematic framework through iterative consensus-building discussions. The finalized codes can be found in the Table. 3.

To address ***RQ2*** regarding the impacts of implementing the concurrent IVE method on the collected user data, we specifically analyzed: (1) quantitative evaluations from Time Spent, Number of collected valuable insights, Number of collected errors, and self-assessment of UX collection questionnaires (see Table. 3) to assess the ***UX data collection process and outputs*** (Section 6.1); (2) qualitative feedback from participants' interviews regarding their perceptions of UX collection performance. This approach correlated observed scoring variations with underlying qualitative insights, revealing key themes in the Table. 3, such as "*Objectivity enhancement with environmental information*" and "*Thematic relevance organization through spatial anchoring*".

To answer ***RQ3*** regarding how participants perceive the concurrent IVE method for UX data collection, we specifically analyzed: (1) quantitative evaluations from MEQ, Immersion (see Table. 3), SUS, and NASA-TLX; (2) qualitative feedback from participants' interviews regarding perception of both the concurrent IVE method and retrospective IVE method. Our integrated analysis identified statistically significant differences in participants' perceptions across three dimensions: ***Usability*** (Section 6.2), ***Immersion*** (Section 6.3), and ***Memory Support*** (Section 6.4), see Table. 3.

6 RESULTS

6.1 UX Data Collection Process and Outputs

6.1.1 Objective UX Data Results. The system-collected UX data first revealed differences among the three conditions in terms of task time, error, and valuable insights. The Repeated Measure ANOVA (Appendix C Table 7) showed a significant main effect on task ***time spent*** across three conditions ($p < .001$). The pairwise comparisons with Bonferroni correction (Appendix C Table 8) showed that the retrospective method required significantly less time than both concurrent IVE method and combined method (both ps $< .05$), whereas the combined method required significantly more time than concurrent IVE method ($p < .001$), see Fig. 6. a.

Three researchers collaboratively analyzed collected UX data with three methods through a two-step evaluation process. First, two researchers evaluated the number of valuable insights in each UX collection dimension by identifying opportunities that provided

Table 3: The outline of the qualitative codes and findings from the user study.

Section	Codes
6.1 UX Data Collection Process and Outputs	Sequential cognitive engagement through environmental immersion.
	Objectivity enhancement with environmental anchoring.
	Thematic relevance organization through spatial anchoring.
	Contextual insight generation through environmental triggers.
	Cognitive load from scene-memory alignment verification.
	Reduced confidence in opportunity generation without designer guidance.
6.2 Perception of Usability	Enhanced engagement through gamified exploration experience.
	Complete system capability validation.
	Physical strain from hardware limitations.
	Initial learning barriers from unfamiliar VR interactions.
6.3 Perception of Immersion	Enhanced autonomous expression through reduced social pressure.
	Increased task motivation through visual progress feedback.
	Sustained attention through distraction-free immersion.
	Impaired verbal fluency due to concurrent recording pressure.
	Attention fragmentation from excessive environmental stimuli.
6.4 Perception of Memory Support	Enhanced memory vividness through multi-sensory stimulation.
	Logical narrative coherence through spatial progression.
	Memory accessibility improvement through environmental cue support.
	Reduced temporal awareness due to spatial immersion focus.

inspiration for designing solutions. Second, they assessed the number of errors compared with the UX data collected from concurrent method and interview conducted in the combined method by examining missing dimensional information, content omissions or inaccuracies, and deviations from user descriptions. Following initial coding and analysis by two researchers, the third researcher (i.e., the designer who completed the journey mapping using three methods) conducted verification to ensure analytical reliability. Both researchers reached a consensus on all evaluation outcomes.

The Wilcoxon signed-rank test revealed that median ***error*** counts differed significantly from zero across six dimensions (Appendix C Table 9): scenarios, actions, touch points, emotions, pain points, and opportunities. This indicates that the retrospective interview component in the combined method yielded significantly different responses after the concurrent UX collection. In contrast, there was no significant in phase (all $ps < .05$), see Fig. 6. b.

The Friedman test revealed significant main effects (Appendix C Table 10) in all seven dimensions on the number of ***valuable insights*** (all $ps < .05$) . The post-hoc tests with Bonferroni correction (Appendix C Table 11) showed that, across six dimensions: phases, scenarios, actions, touch points, emotions, and pain points, the combined method was significantly superior to the retrospective method (all $ps < .05$). In scenarios, actions, touch points, and emotions, the concurrent IVE method was also significantly superior to the retrospective method (all $ps < .05$). No significant differences were found in opportunities between three methods, see Fig. 6. c.

6.1.2 Subjective UX Data Collection Assessment. The results of the Friedman test on ***self-assessment of UX collection*** indicated that the three conditions had main effects (Appendix C Table 12) on Completeness, Consistency, Relevance, and Insight (all $ps <$ .001). There were no main effects on Accuracy and Clarity (both $ps >$ .05), see Fig. 7. The following sections present participant interview rationales that explain these rating differences, revealing key strengths of the concurrent method with JourneyCapturer in both the UX collection process and its outputs:

Sequential cognitive engagement through environmental immersion. Regarding the Completeness subscale, post-hoc pairwise comparisons with Bonferroni correction (Appendix C Table 13) showed that the concurrent IVE method and the combined method were significantly superior to the retrospective method (both ps $<$ .01). The concurrent IVE method through JourneyCapturer's rich environmental details facilitated participants in generating more UX content by encouraging deeper thinking at each stage of their journey. P2 emphasized the IVE's capacity to stimulate situational thinking: "*think about abnormal situations like subway breakdowns or rush hour crowds within each phase*". This cognitive stimulation extended beyond simple recall to analytical thinking, enabling participants to "*observe the environment and generate insights*" (P7) and "*think about different cases*" (P15) at each phase. The environmental richness also promoted exploration, allowing them to "*describe more details such as actions*" (P4) and "*analysis of each interaction point*" (P9).

Objectivity enhancement with environmental anchoring. Compared with the retrospective method, the post-hoc pairwise comparisons with Bonferroni correction (Appendix C Table 13) showed that the concurrent IVE method and the combined method exhibited significantly higher scores in terms of Consistency (both

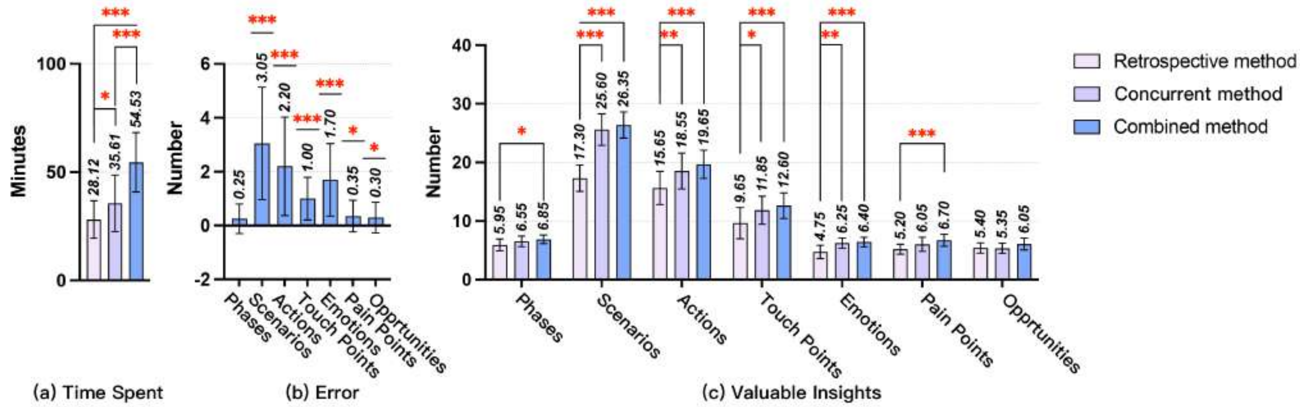

Figure 6: The results of (a) Time Spent, (b) Error Count, and (c) the number of Valuable Insights.

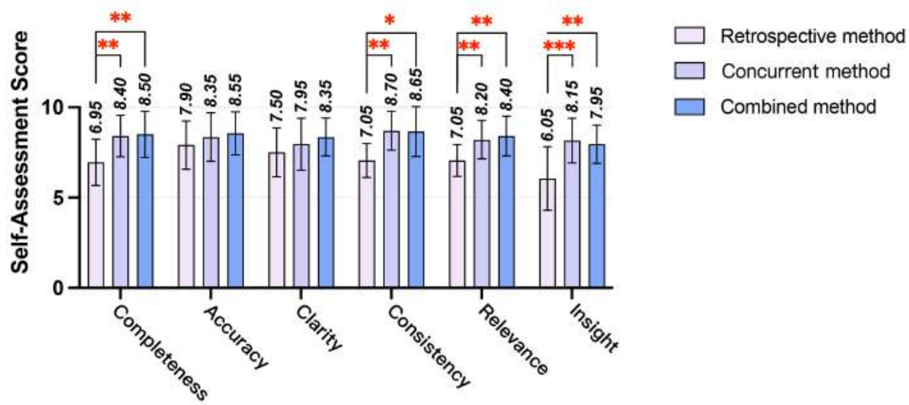

Figure 7: The results of the self-assessment of the UX collection.

$ps < .05$). Environmental anchoring in the concurrent IVE method enhanced data objectivity and authenticity, with participants noting VR emotions were "*more genuine and spontaneous*" (P6, P11). The environmental anchoring prevented biased responses, as in interviews "*details might not be accurate*" (P1) or even can be "*fake*" (P18) as they cater to the designer and lose independent thinking. P11 exemplified this: "*In VR, seeing the hospital window made me feel frustrated about queuing, but in the interview, I reported feeling calm because of the calm state I recalled*".

Thematic relevance organization through spatial anchoring. In the subscale of Relevance, the post-hoc pairwise comparisons with Bonferroni correction (Appendix C Table 13) showed that the concurrent IVE method and the combined method scored significantly higher than the retrospective method (both $ps < .01$). Concurrent IVE method achieved a strong correlation between content and dimensions through JourneyCapturer's autonomous exploration and spatially triggered recording, preventing topic drift. P17 noted: "*I record content near relevant objects, with panels for review*". The environmental anchoring maintained "*thematic focus*" (P2) and "*obvious structural continuity*" (P4), while interviews were fragmented and "*suffered from memory confusion across phases*" (P13). P16 illustrated: "*When discussing registration in hospital, I mentioned queuing, mixing different phases*".

Contextual insight generation through environmental triggers. The post-hoc pairwise comparisons with Bonferroni correction (Appendix C Table 13) showed that both the concurrent IVE method and the combined method significantly enhanced Insight generation, compared with the retrospective method (both $ps < .01$). Concurrent IVE method through JourneyCapturer's IVE stimulated deeper thinking, generating valuable insights about pain points and opportunities (P2, P19), including environmental and accessibility considerations. Participants emphasized JourneyCapturer's depth advantage, such as "*recall surrounding people's pain points in specific environments*" (P14) and "*suggesting user-friendly designs for disabilities when seeing turnstiles*" (P3). In contrast, participants focused more on "*personal feelings*" (P4) in interviews, finding it difficult to think deeply without surrounding environments.

Two notable challenges were mentioned: the concurrent IVE method required participants to verify differences between JourneyCapturer's virtual environment and their personal memories, and reduced confidence due to a lack of concurrent designer feedback.

Cognitive load from scene-memory alignment verification. The concurrent IVE method using JourneyCapturer required participants to validate scene-memory alignment, necessitating time spent checking the IVE against real memory scenarios (P16). P5 acknowledged limited environmental reference: "*I didn't use the VR environment much for reference, mainly relying on real experience recall*". Despite scene differences, some participants emphasized maintained effectiveness, such as "*checking against wrong scenarios makes me more rigorous*" (P9) and "*scenes are realistic, and walking nearby lets me think more relevant details*" (P14).

Reduced confidence in opportunity generation without designer guidance. The concurrent IVE method's self-recording diminished participants' confidence in generating opportunities (P3) due to a lack of concurrent designer feedback, despite maintaining the collection of rich detail. P15 emphasized the necessity of guidance: "*Designers can guide me to think more precisely about professional design opportunities that designers want*". The absence of interactive validation particularly affected opportunity identification, as P10 observed: "*I don't know what to say, which parts are important, but designers serve as reminders when I go off-topic*".

6.2 Perception of Usability

We used SUS, behavior pattern, and NASA-TLX to evaluate the users' perception of usability. The ***SUS*** evaluation revealed high usability for both IVE systems, with JourneyCapturer scoring 77.60 (*SD* = 10.57, grade B+) and the IVE tool without annotation that we provided before conducting the interview method, achieving 80.15 (*SD* = 13.26, grade A-). Both exceeded the established SUS benchmark of 68.

To complement participants' subjective usability perceptions, we analyzed ***the patterns of interaction behavior*** within JourneyCapturer through system logs capturing the numbers of the operations (i.e., *annotation creation*, *deletion*, and *edit*), see Table 4. The total number of *annotations* (M = 36.90) appears to be associated with annotation creation of journey Phases (M = 6.80). This is because when participants created UX collection template panels-Phase (i.e., interactive panels for recording UX dimensions in IVE), they would consistently record UX collection dimensions of Scenarios (M = 6.40), Actions (M = 5.65), and Emotions (M = 6.10), and therefore accumulated total *annotation* numbers. However, participants showed reduced engagement with dimensions of Touch Points (M = 3.40) and Opportunities (M = 3.55), which required professional opinions. Additionally, *Deletion* (M = 3.70) operations primarily addressed speech-to-text errors, while *edits* (M = 1.45) occurred when participants needed to categorize recorded content or provide contextual supplements to previous recordings throughout the whole journey. As a result, the numbers of *edits* and *deletions* were noticeably lower compared to other operations.

The Friedman test results showed significant differences of ***NASA-TLX*** among three conditions in Physical Demand, Temporal Demand, and Performance (all ps < .01), as shown in Appendix C, Table 14. Post-hoc analyses revealed that both the concurrent method and the combined method imposed significantly higher Physical Demand compared with the retrospective method (both ps < .05). Both concurrent and combined methods imposed higher Performance than retrospective method (both ps < .05), while concurrent method yielded lower Temporal Demand than combined method (p < .01), see Fig. 8 and Appendix C, Table 15. Participants indicated that moderate physical demand did not undermine usability, with JourneyCapturer demonstrating advantages in two ways:

Table 4: The number of separate annotations for seven dimensions, total annotations, total deletions, and total edits.

Operations	Total Numbers	Mean (SD)
Separate Annotations for seven dimensions with Template	Phases	6.80 (11.56)
	Scenarios	6.40 (6.64)
	Actions	5.65 (6.83)
	Touch Point	3.40 (10.34)
	Emotions	6.10 (7.79)
	Pain Points	5.00 (6.10)
	Opportunities	3.55 (4.35)
Total Annotations		36.90 (13.14)
Total Deletions		3.70 (2.91)
Total Edits		1.45 (1.75)

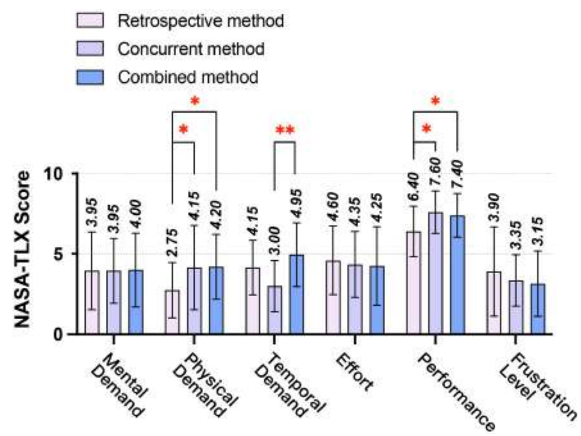

Figure 8: The results of NASA-TLX.

Enhanced engagement through gamified exploration experience. Gamified interaction elements fostered an engaging sense of achievement in the concurrent IVE method. P11 noted "*moving through IVE felt like playing a game*", while P6 emphasized that the interactive panels created discovery-based engagement. The visualization of recorded annotations through progress indicators motivated continued participation, with "*seeing my journey's completion encouraged thorough exploration*" (P14). This intrinsic motivation reduced research anxiety while maintaining the effectiveness of the data collection process.

Complete system capability validation. JourneyCapturer fulfilled design expectations across five design goals. Participants confirmed the achievement of environmental authenticity (G1) through "*distraction-free focus*" (P8). Navigation comfort optimization (G2) by teleportation, preventing motion sickness while maintaining exploration freedom (P2, P12). Scene-driven annotation (G3) enabled efficient data collection with "*seamless voice recording and* *object selection*" (P15). Flexible template panel (G4) supported structured documentation through editing and deletion functions (P3, P6, P18). Spatial-temporal coupling (G5) provided organized output corresponding to specific scenarios (P7, P12).

Nevertheless, participants also identified limitations in the concurrent IVE method, including JourneyCapturer's physiological discomfort and challenges with unfamiliar operations.

Physical strain from hardware limitations. Mental and physical demands created a significant burden for users through the device's weight and motion discomfort. P9 reported "*VR device's weight caused neck strain after extended use*", while P13 noted "*fast movement transitions triggered mild dizziness*" despite teleportation implementation. These physiological barriers potentially limited effectiveness in the last phases of UX data collection and users' willingness for repeated participation.

Initial learning barriers from unfamiliar VR interactions. The specific interaction functions of JourneyCapturer created learning curve barriers affecting initial performance. P4 struggled with "*understanding teleportation and button functions*", requiring "*more practice time*" (P16). Voice recording activation initially confused participants, with P10 noting, "*when recording started or stopped wasn't clear at first*". However, participants typically achieved complete independence within 10 minutes without affecting overall effectiveness.

6.3 Perception of Immersion

The Friedman test results for ***Immersion*** (Appendix C, Table 16) indicated that the three conditions had significant main effects on Expression Naturalness, Exploration Freedom, Engagement Level, Overall Immersion, and Cognitive Fluency (all $ps < .01$), but not on Interference Factors and Focus (both $ps > .05$). The post-hoc tests with Bonferroni correction (Appendix C, Table 17) showed that, for Expression Naturalness, both the retrospective method and the combined method scored significantly higher than the concurrent method (both $ps < .01$). In contrast, for Exploration Freedom, Engagement Level, Overall Immersion, and Cognitive Fluency, the concurrent method yielded significantly higher scores than the retrospective method (all $ps < .01$), so as combined method (all $ps < .05$), see Fig. 9. We present participants' perspectives on task immersion for the different dimensions in the following sections. We first present the identified strengths of the concurrent method with JourneyCapturer as follows:

Enhanced autonomous expression through reduced social pressure. Concurrent IVE method through JourneyCapturer's self-directed recording environment reduced social pressure and enhanced participants' willingness to express authentic experiences. Participants noted that an immersive self-paced environment allowed them to "*record more relaxedly and freely*" (P7) and "*feel safer and more comfortable for speaking*" (P19), enabling deeper reflection for collecting data. The autonomous nature eliminated the demands for immediate responses that characterize traditional interviews. Participants explained that retrospective interviews require "*immediate feedback*" (P14) and "*interacting with people*" (P12) under pressure, but JourneyCapturer allowed them to"*think first then speak naturally*" (P3).

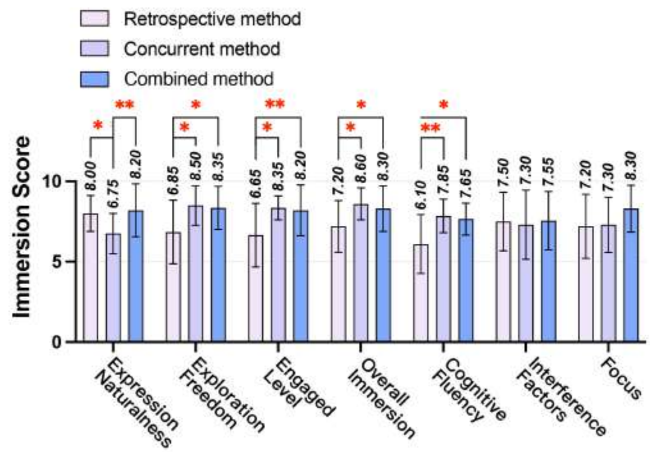

Figure 9: The results of Immersion.

Increased task motivation through visual progress feedback. Concurrent IVE method enhanced participants' task completion satisfaction through the visual feedback system with tangible progress panels in the virtual environment. P5 noted that "*Looking at the panel text feels like completing a task in progress, which gives me a sense of accomplishment*", while P6 observed "*The record panel creates motivation to complete the task*". The visible accumulation of transcribed content provided continuous validation of collecting experiential data, as P13 explained: "*I can see my thoughts being organized logically on the panel, which makes me feel productive*".

Sustained attention through distraction-free immersion. Concurrent IVE method enhanced participants' concentration by eliminating external distractions and designer interference. P1 noted that "*VR creates environmental immersion where I can follow my own thoughts completely*", while P9 observed being "*easily attracted by interactions and environment in VR*". This protected cognitive space enabled sustained attention and deeper engagement, with participants feeling "*safer with higher security*" (P18). The isolation from social pressure proved particularly beneficial, contrasting with retrospective interviews where participants noted feeling "*nervous without safety*" (P4).

Although the concurrent IVE method effectively enhanced concurrent UX collection motivation through three key strengths regarding facilitated autonomous expression, improved achievement cognition, and enhanced focus, two challenges emerged from participants' feedback:

Impaired verbal fluency due to concurrent recording pressure. Concurrent IVE method created temporal pressure that disrupted participants' natural expression flow through JourneyCapturer's concurrent voice recording mechanism. Participants noted that they needed to "*think well before expressing in the record panel*" (P6) and "*requires time to organize language, because I want to speak frequently for each experiential dimension*" (P8, P13). The concurrent recording format demanded pre-planning and sometimes disrupted participants' expression flow, as P14 noted "*I feel a little nervous when I am recording and believe that I just have one chance to record, so sometimes I get disrupted when speaking*", while participants

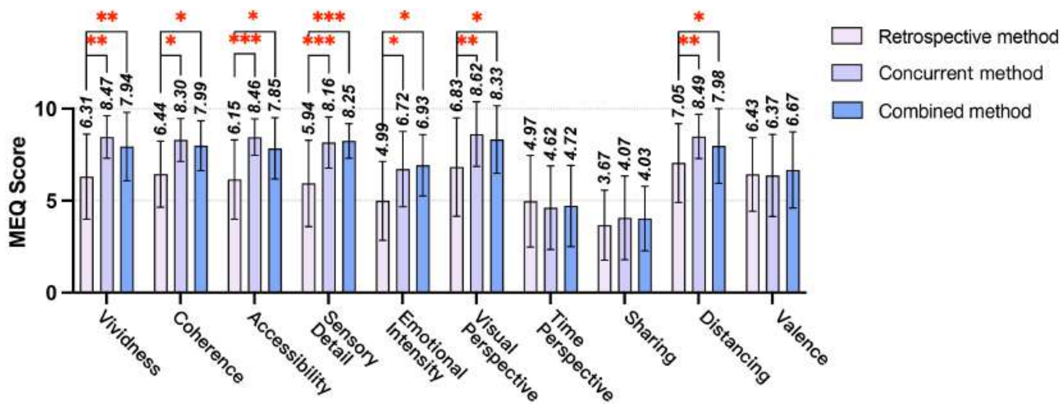

Figure 10: The results of MEQ.

could "*describe more fluently and quickly*" (P16) in retrospective interviews.

Attention fragmentation from excessive environmental stimuli. Concurrent IVE method through JourneyCapturer's rich visual environment occasionally diverted participants' attention from core UX reflection to irrelevant environmental elements. P8 observed that "*VR's multi-sensory environment can output more information, but visuals also distract attention*", while P9 noted "*VR is creating extra information, and I am attracted to interactions, making it difficult to focus on my own sensations*". This cognitive fragmentation contrasted with retrospective interviews where P20 emphasized "*VR's concurrent recording has too much freedom, which may result in some irrelevant collection, as I want to include everything I see*".

6.4 Perception of Memory Support

The ***MEQ*** results (Appendix C, Table 18) revealed that the three conditions had significant main effects on Vividness, Coherence, Accessibility, Sensory Detail, Emotional Intensity, Visual Perspective, and Distancing, but not on Time Perspective, Sharing, or Valence (all $ps < .05$). The combined method scored significantly higher than the retrospective method on Vividness, Coherence, Accessibility, Sensory Detail, Emotional Intensity, Visual Perspective, and Distancing (all $ps < .05$). Concurrent method also scored significantly higher than the retrospective method on Vividness, Coherence, Accessibility, Sensory Detail, Emotional Intensity, Visual Perspective, and Distancing (all $ps < .05$), see Fig. 10 and Appendix C, Table 19. However, no significant differences were found in time perspective, sharing, and valence between the three methods. We present the identified strengths from the reflections of the participants:

Enhanced memory vividness through multi-sensory stimulation. Concurrent IVE method through JourneyCapturer's audio-visual integration effectively intensified memory vividness through coordinated sensory stimulation, as "*Sounds help me recall many subway audio details, making it more immersive*" (P11). P6 noted cross-sensory activation: "*Seeing medicine in the environment makes me think of its taste*". P20 highlighted environmental memory triggers: "*Environmental factors provide memory prompts for vividness*". The multi-modal stimulation created richer concurrent recall experiences, as P7 observed: "*Visual elements stimulate other senses, which is difficult to achieve through pure retrospective memory recall*".

Logical narrative coherence through spatial progression. Concurrent IVE method achieved enhanced logical coherence by leveraging JourneyCapturer's spatial embodiment and sequential environmental navigation, in contrast to fragmented interview recalls. P4 and P14 emphasized: "*I am moving into an environment to record my experience, so the whole content has logic and organization*". The spatial progression maintained ordered recall phases, through "*recording content near corresponding objects with one-to-one correspondence*" (P17) and "*reviewing previous recordings through panels*" (P5). In contrast, P19 noted that interviews were "*more like segmented memory summaries*", lacking the structured narrative flow without contextual anchoring.

Memory accessibility improvement through environmental cue support. Concurrent IVE method through JourneyCapturer's realistic scenes reduced cognitive burden and enhanced memory retrieval through environmental cues. P3 and P11 emphasized the reduction of cognitive load: "*VR immersion eliminates the need to imagine scenarios, reducing mental workload as I don't need to visualize scenes in my mind*". Participants noted procedural support: "*the walking flow with environments reduces recall burden, such as phases and actions*" (P12), and "*lightens cognitive task load*" (P18). Additionally, JourneyCapturer enhanced the retrieval of UX experiences, such as "*seeing scenes' details triggers associations for recording*" (P7) and even "*thinking of things when seeing the self-service machine in hospital that I couldn't recall through retrospective method*" (P2).

However, participants revealed one primary concern with the concurrent IVE method: reduced attention to duration estimation due to the concurrent spatial immersion while prioritizing event sequence and spatial details.

Reduced temporal awareness due to spatial immersion focus. Concurrent IVE method diminished participants' attention to temporal details through JourneyCapturer's spatial focus, prioritizing event sequence over the estimation of duration. P9 noted the redirection of attention: "*VR makes me focus more on current phases' details rather than time consciousness and thinking about duration*". P16 emphasized temporal immersion: "*VR makes time*

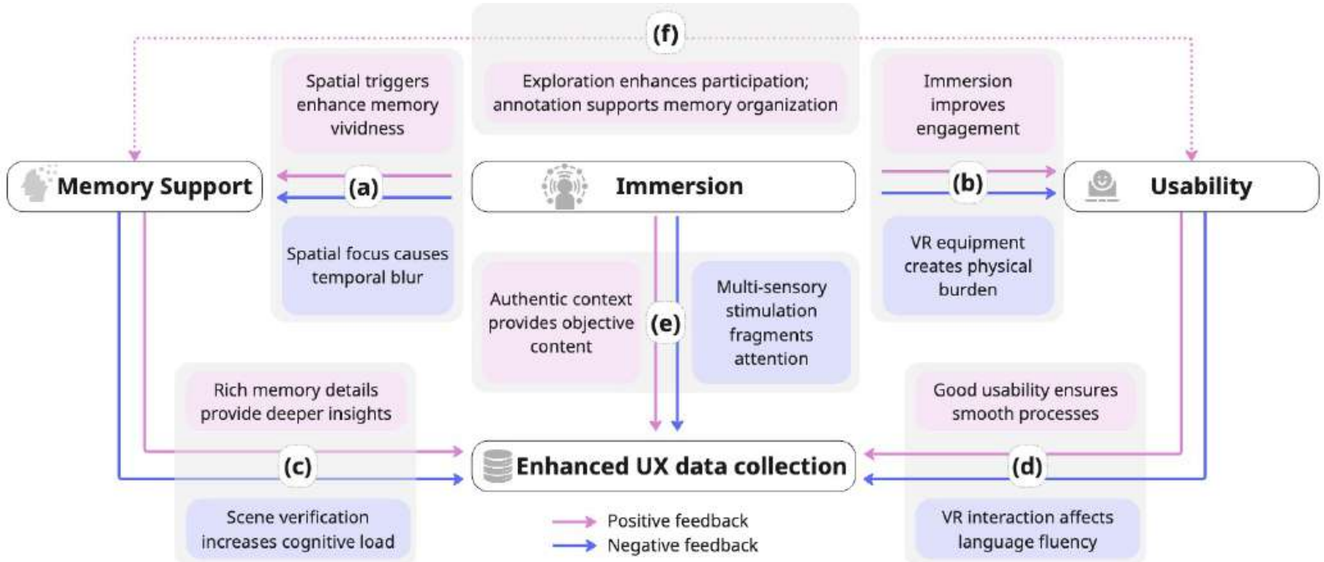

Figure 11: The Overview of Interdependent Relationships across the three dimensions of findings.

feel slower and more relaxed, which helps with deep thinking but reduces time awareness". Conversely, interviews naturally elicited temporal estimates, as P15 observed: "*Interviews help me provide time estimates like 'passing through gates takes 5-10 minutes', while VR focuses more on spatial positioning and location clarity*".

7 DISCUSSION

7.1 Comparative Trade-offs in Concurrent IVE Method and Retrospective Method

Our comparative evaluation revealed distinct advantages and considerations for each UX data collection method, highlighting the complex trade-offs among efficiency, authenticity, and completeness in UX data collection. Each method demonstrated unique strengths, specific challenges drawn from participants' perceptions, and the insights to address ***RQ2*** and ***RQ3*** (*Section 6.1 - 6.4*).

Concurrent IVE Method: Consciously Proactive Collection with Environmental Anchoring. The concurrent IVE method represents a departure from traditional passive concurrent approaches [2, 21] by enabling consciously proactive data collection triggered by environmental contexts rather than reactive verbalization during task performance. This approach demonstrated advantages through environmental anchoring, enabling UX content generation beyond simple recall [28, 63] while maintaining thematic relevance through spatial progression. Environmental contexts enhanced objectivity by preventing biased responses [7] common in retrospective interviews. However, the requirements for cognitive scenario validation emerged when IVE differed from personal memories, and the disruption of expression fluency occurred due to the concurrent recording pressure. Additionally, reduced confidence in opportunity generation resulted from a lack of designer guidance [27].

Retrospective method: Professional Guidance with Memory Limitations. The retrospective method provided valuable facilitation by the designer [22, 53] through guided divergent thinking and professional interpretation [71], particularly effective for pain points and opportunities requiring design insight. Expression fluency remained high, allowing participants to describe experiences more naturally compared to concurrent conditions. However, memory distortion and post-experience rationalization [58] compromised data authenticity, with participants acknowledging that interview details might be inaccurate as they cater to designers' expectations. Thematic drift and fragmented recall across journey phases created structural inconsistencies, while social pressure from designers affected the sharing of authentic responses.

Combined method: Complete Coverage with Efficiency Concerns. The combined method offered cross-validation [6, 10] between insights generated by participants and facilitated interpretations by designers [44] while ensuring coverage of UX data through the use of dual approaches. However, the extended time commitment led to redundancy and participant fatigue, thereby affecting data quality. Our error analysis revealed that participants were no longer immersed during post-IVE interviews, creating conflicts in memory and reliance on general knowledge rather than authentic experiential recall. This method reveals evolving the dynamics between participants and designers, where traditional relationships between interviewer and interviewee shift toward collaborative exploration, suggesting the potential for optimizing strategic collaboration between users and designers within the environmental context.

While each method presents distinct trade-offs, the consciously proactive concurrent IVE method demonstrates complex ***interdependent relationships***, see Fig. 11, among three core dimensions of findings that collectively influence the effectiveness of collecting UX data. Our qualitative findings revealed a potential pattern in

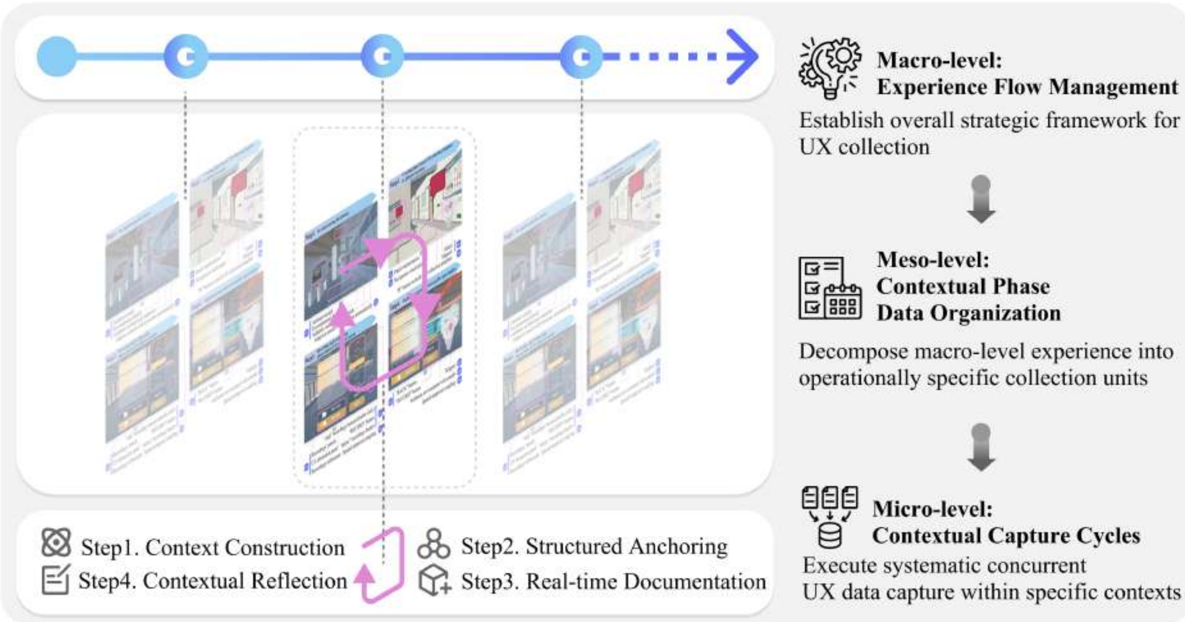

Figure 12: Guidelines for Implementing Proactive Concurrent UX Collection.

how immersion, memory support, and usability mutually influence one another and collectively contribute to the collection of UX data. First, ***Immersion*** creates dual effects on memory and usability: for the memory support, spatial environmental triggers activate situational experience, enhancing vividness of journey articulation while excessive spatial focus leads to temporal perception blur (a). For usability, IVE improves user concentration and engagement, yet introduces VR equipment burden and technical barriers (b). These immersive qualities directly enhance the collection of data by integrating authentic environmental context for objectivity, but create attention fragmentation that distracts from core UX reflection (e). Second, ***Memory Support*** provides environment-triggered rich details for complete insights into user experiences, though scene-experience verification increases cognitive load that requires mental effort to align scenarios (c). Third, ***Usability*** ensures smooth processes for collection and improves content relevance through seamless interaction, yet VR voice interaction affects language fluency and lacks professional guidance, leading to reduced confidence (d). Memory support and usability mutually reinforce each other, with richer contextual details enhancing engagement while smoother interactions facilitate better memory organization (f). Ultimately, these six relationship pathways converge on an ***Enhanced UX Data Collection Process and Output***, where the dynamic interplay between immersive engagement, memory retrieval support, and system usability determines both the procedural efficiency and qualitative depth of UX insights within the consciously proactive concurrent IVE method.

7.2 Guidelines for Implementing Proactive Concurrent UX Collection

Our study revealed that participants naturally exhibited consciously proactive behavior patterns to collect UX data within situational contexts across multiple dimensions. At the Micro-level, participants systematically executed iterative processes for capture, including context construction, structured anchoring, concurrent documentation, and contextual reflection within each phase. At the Meso-level, they deliberately segmented experiences by breaking journeys into self-defined phases with distinct contextual characteristics and collection requirements. At the Macro-level, participants established collection boundaries, identified key experiential milestones, and strategically allocated attention across entire journeys.

Building on these empirical observations, we propose the ***Proactive Concurrent UX Collection Guidelines*** that transform participants' consciously proactive behaviors for collection into a structured framework for future UX research, see Fig. 12. We formalize this approach through ***Context-Data Coupling Loops***, iterative cycles where context and experiential data become mutually reinforcing. These loops operate across three hierarchical levels: 1) ***Micro-level: Contextual Capture Cycles***. This level executes concurrent data collection within defined phases through four iterative steps (see Section 4.2): a) Context Construction to establish situational immersion; b) Structured Anchoring to bind UX dimensions to specific contextual elements; c) Concurrent Documentation to capture concurrent experiential data; d). Contextual Reflection to generate situation-triggered insights. Users actively select recording moments based on immediately authentic feelings, directly utilizing contextual elements as capture triggers and analytical anchors. For instance, participants anchored feelings of anxiety to X-ray machines during security checks, then reflected on accessibility implications for elderly passengers. 2) ***Meso-level: Contextual Phase Data Organization***. This level segments experiences into contextually meaningful collection units, where users actively define specific phase boundaries based on contextual transitions, establish specialized data collection templates for each phase, determine situational characteristics and collection priorities, and organize logical relationships between phases. In our study, participants naturally segmented subway journeys into several phases, such as entry orientation, security ticketing, navigation wayfinding, and so on, each with phase-specific collection templates. 3) ***Macro-level: Experience Flow Management***. This level establishes a complete data collection strategy with contexts through holistic experience planning, where users determine overall experience boundaries, establish collection strategies with priority hierarchies, and plan resource allocation across the entire experience. Users achieve conscious perspectives by actively determining personal collection boundaries to ensure the UX collection's completeness and coherence. For example, in our subway study, participants defined their journey boundary from station entrance to exit, prioritizing wayfinding challenges over routine interactions.

These guidelines provide methodological direction for implementing proactive UX documentation through first-person experiencing across diverse research contexts, extending to a general collection of UX data beyond the journey mapping context. This guideline transforms traditional passive data collection into conscious, Context-Data Coupling UX data capture.

7.3 Research Implications for Collecting UX Data

Our findings drive proposed research implications by addressing past methodological limitations and offering innovative UX data collection approaches. These insights inform the development of future UX data collection systems and frameworks, enhancing the rigor, precision, and accessibility of UX research.

Addressing Past Limitations: Concurrent Expert Collaboration for Enhanced Guidance. Our findings revealed that participants experienced reduced confidence in opportunity generation without guidance from expert designers. Future systems could incorporate experts' collaboration to simultaneously observe and analyze data during IVE collection sessions. Experts could monitor users' first-person perspectives, providing immediate guidance and conducting preliminary analysis [32] while users engage with JourneyCapturer. This approach would enable concurrent assurance of quality [33], professional data interpretation [8, 43], and structured preprocessing [64, 73] during collection, enhancing methodological rigor and users' confidence through guidance from designers.

Enhancing IVE Capabilities: Flexible Gestural Annotation for Spatial Precision. While our concurrent IVE method showed participants successfully anchored UX dimensions to spatial elements, the first-person perspective sometimes limited precise reference specification. Future implementations should incorporate gestural annotation capabilities [70], allowing users to draw, point, or highlight [89] specific environmental details within VR spaces, enabling more precise spatial-experiential coupling and addressing the attention fragmentation issues we observed with rich environmental stimuli.

Expanding Future Accessibility: Scalable Methods for Authentic Field Research. Our IVE approach demonstrated the ability to maintain experimental rigor while enabling authentic field-based research, expanding possibilities for broader accessibility to field-based applications. The modular design of our Context-Data Coupling Loops framework transcends spatial contexts to support general UX data documentation, from traditional usability labs to field-based authentic user journeys. Future implementations should expand to Augmented Reality annotation systems [20, 45] overlaying data collection interfaces onto real-world contexts, enabling spatial-temporal coupling in authentic user journeys while preserving our guidelines' structured reflection capabilities.

Advancing Future Applications: Multi-modal Data Integration for deeper Analytics. The environmental grounding we identified acts as a cognitive catalyst, enhancing memory vividness and narrative coherence. Future concurrent IVE data collection could integrate eye-tracking [69], biometric monitoring [37, 88], and behavioral analytics [37, 91] with voice-based annotations, creating rich multi-dimensional datasets that capture both explicit user feedback and implicit behavioral indicators for deeper environmental trigger analysis.

7.4 Limitations

Several limitations constrain the generalizability of our findings. First, the demographic profile was primarily drawn from Hong Kong and mainland China, potentially limiting generalizability to different cultural contexts where UX research practices vary. Second, the relatively small sample size may limit the robustness of our findings, restricting our ability to draw more generalizable conclusions. Third, our reliance on IVE introduces ecological validity concerns, as standardized IVE scenarios represent generic interpretations rather than participants' personal memory landscapes. This limitation manifested in challenges of cognitive scenario validation, where participants noted discrepancies between virtual environments and authentic experiences, requiring additional cognitive effort for alignment between scene and memory. Fourth, the study's scope was specifically constrained to environmental UX contexts. While we demonstrated effectiveness in location-based scenarios (hospital and subway), our findings may not generalize to other UX domains, such as digital interface interactions or online service experiences. Moreover, our study did not compare the conscious spatial-experiential coupling approach against unconscious data collection methods (e.g., think-aloud protocols), limiting our understanding of whether deliberate reflection produces fundamentally different insights than spontaneous verbalization.

8 CONCLUSION

This paper introduced a consciously proactive concurrent IVE method with first-person perspective through JourneyCapturer, a tool designed to enhance UX data collection processes and outputs by integrating concurrent structured reflection and in-situ annotation capabilities within IVE. We identified five design goals that addressed participants' expectations for incorporating concurrent IVE methods into UX data collection through a formative study with 34 participants (***RQ1***). These design goals informed JourneyCapturer's development with six experts' co-design workshop. Subsequently, our user study with 20 participants revealed valuable insights into the advantages and challenges of using three UX collection methods: concurrent IVE annotation, retrospective interview, and combined method. Participants found that concurrent IVE method substantially improved UX data collection through sequential cognitive engagement and environmental anchoring, facilitating a more complete capture of experiential insights, including accessibility considerations and environmental design opportunities (***RQ2***). Regarding the participants' perception of the concurrent IVE method with JourneyCapturer, they valued its memory support capabilities and immersive engagement features, such as the development of narrative cohesion and the promotion of achievement recognition (***RQ3***). However, participants also raised concerns about disruption in expression fluency, and the requirements of cognitive scenario validation when virtual environments diverged from personal memories. Despite these limitations, our results demonstrated concurrent IVE method's potential to transform UX research practices by enabling concurrent, context-aware data collection that addresses memory distortion and post-experience rationalization common in traditional retrospective method. Building on these insights, JourneyCapturer demonstrates system capabilities for consciously proactive concurrent UX data collection within IVE through environmental anchoring and spatial-temporal coupling. We propose future implications for expert collaboration, multi-modal analytics, scalability for AR-based field studies, and flexible gestural annotation.

Acknowledgments

The research presented in this article was funded by grants from Hong Kong Research Grants Council (Project No. 15604825) and the Hong Kong Polytechnic University (Project No. P0046486, No. P0046226).

References

[1] Omolola A Adeoye-Olatunde and Nicole L Olenik. 2021. Research and scholarly methods: Semi-structured interviews. *Journal of the american college of clinical pharmacy* 4, 10 (2021), 1358–1367. doi:10.1002/jac5.1441
[2] Obead Alhadreti and Pam Mayhew. 2018. Rethinking thinking aloud: A comparison of three think-aloud protocols. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12. doi:10.1145/3173574.3173618
[3] Michael J Almeida. 2012. Time in narratives. In *Deixis in narrative*. Psychology Press, 159–189. doi:10.4324/9780203052907-9
[4] Roghayeh Barmaki and Charles E Hughes. 2018. Embodiment analytics of practicing teachers in a virtual immersive environment. *Journal of Computer Assisted Learning* 34, 4 (2018), 387–396. doi:10.1111/jcal.12268
[5] Leif P Berg and Judy M Vance. 2017. Industry use of virtual reality in product design and manufacturing: a survey. *Virtual reality* 21 (2017), 1–17. doi:10.1007/s10055-016-0293-9
[6] Daniel Berrar et al. 2019. Cross-validation. doi:10.1016/B978-0-12-809633-8.20349-X
[7] Christine Blome and Matthias Augustin. 2015. Measuring change in quality of life: bias in prospective and retrospective evaluation. *Value in health* 18, 1 (2015), 110–115. doi:10.1016/j.jval.2014.10.007
[8] Callum Bradley, Luis Oliveira, Stewart Birrell, and Rebecca Cain. 2021. A new perspective on personas and customer journey maps: Proposing systemic UX. *International Journal of Human-Computer Studies* 148 (2021), 102583. doi:10.1016/j.ijhcs.2021.102583
[9] John Brooke. 2013. SUS: a retrospective. *Journal of usability studies* 8, 2 (2013).
[10] Michael W Browne. 2000. Cross-validation methods. *Journal of mathematical psychology* 44, 1 (2000), 108–132. doi:10.1006/jmps.1999.1279
[11] Anders Bruun and Simon Ahm. 2015. Mind the gap! Comparing retrospective and concurrent ratings of emotion in user experience evaluation. In *IFIP Conference on Human-Computer Interaction*. Springer, 237–254. doi:10.1007/978-3-319-22701-6_17
[12] Chris Bryan, Kwan-Liu Ma, and Jonathan Woodring. 2016. Temporal summary images: An approach to narrative visualization via interactive annotation generation and placement. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 511–520. doi:10.1109/tvcg.2016.2598876
[13] Béatrice Cahour, Pascal Salembier, and Moustafa Zouinar. 2016. Analyzing lived experience of activity. *Le travail humain* 79, 3 (2016), 259–284. doi:10.3917/th.793.0259
[14] Kay Caldwell and Anita Atwal. 2005. Non-participant observation: using video tapes to collect data in nursing research. *Nurse researcher* 13, 2 (2005). doi:10.7748/nr2005.10.13.2.42.c5967
[15] Scott Carter and Jennifer Mankoff. 2005. When participants do the capturing: the role of media in diary studies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 899–908.
[16] Y Chen, Z Cui, and LJLR Hao. 2019. Virtual reality in lighting research: Comparing physical and virtual lighting environments. *Lighting Research & Technology* 51, 6 (2019), 820–837. doi:10.1177/1477153518825387
[17] Janet Cooper, Rachael Lewis, and Christine Urquhart. 2004. Using participant or non-participant observation to explain information behaviour. *Information Research* 9, 4 (2004), 9–4.
[18] Andrew Crabtree, Mark Rouncefield, and Peter Tolmie. 2012. *Doing design ethnography*. Springer Science & Business Media.
[19] Barbara Czarniawska-Joerges. 2007. *Shadowing: And other techniques for doing fieldwork in modern societies*. Copenhagen Business School Press DK.
[20] Francesco Riccardo Di Gioia, Eugenie Brasier, Emmanuel Pietriga, and Caroline Appert. 2022. Investigating the Use of AR Glasses for Content Annotation on Mobile Devices. *Proceedings of the ACM on Human-Computer Interaction* 6, ISS (2022), 430–447. doi:10.1145/3567727
[21] Jules Françoise, Yves Candau, Sarah Fdili Alaoui, and Thecla Schiphorst. 2017. Designing for kinesthetic awareness: Revealing user experiences through second-person inquiry. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5171–5183. doi:10.1145/3025453.3025714
[22] Jill Grant and Frank Fox. 1992. Understanding the role of the designer in society. *Journal of Art & Design Education* 11, 1 (1992), 77–87.
[23] Malte Hagener. 2015. Beautiful Data/The Democratic Surround. *NECSUS. European Journal of Media Studies* 4, 2 (2015), 223–227. doi:10.5117/necsus2015.2.hage
[24] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183. doi:10.1016/S0166-4115(08)62386-9
[25] Morten Hertzum. 2024. Concurrent or Retrospective Thinking Aloud in Usability Tests: A Meta-Analytic Review. *ACM Transactions on Computer-Human Interaction* 31, 3 (2024), 1–29. doi:10.1145/3665327
[26] Trevor Hogan, Uta Hinrichs, and Eva Hornecker. 2015. The elicitation interview technique: Capturing people's experiences of data representations. *IEEE transactions on visualization and computer graphics* 22, 12 (2015), 2579–2593. doi:10.1109/tvcg.2015.2511718
[27] Robin M Hogarth. 2006. Is confidence in decisions related to feedback? Evidence–and lack of evidence–from random samples of real-world behavior. *In the beginning there is a sample: Information sampling as a key to understand adaptive cognition* (2006), 456–484.
[28] Alisha C Holland and Elizabeth A Kensinger. 2010. Emotion and autobiographical memory. *Physics of life reviews* 7, 1 (2010), 88–131. doi:10.1016/j.plrev.2010.01.006
[29] Radhika Pankaj Jain, Adam Drogemuller, Kadek Ananta Satriadi, Ross Smith, and Andrew Cunningham. 2025. Strollytelling: Coupling Animation with Physical Locomotion to Explore Immersive Data Stories. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17. doi:10.1145/3706598.3713132
[30] Sangsu Jang, Nari Kim, Nanum Kim, Jin-young Moon, Choong-Wan Woo, and Young-Woo Park. 2025. Journey to My Past: Exploring and Journaling Past Memories Evoked by Questions Framed as Proud Moments. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19. doi:10.1145/3706598.3713215
[31] Jason Jerald. 2015. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool. doi:10.2200/S00691ED1V01Y201507ICR043
[32] Marlene Kadar and Susanna Egan. 2006. *Tracing the autobiographical*. Wilfrid Laurier Univ. Press. doi:10.51644/9780889209077
[33] James Kalbach. 2020. *Mapping experiences*. O'Reilly Media.
[34] Terry K Koo and Mae Y Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine* 15, 2 (2016), 155–163. doi:10.1016/j.jcm.2016.02.012
[35] Jérémy Lacoche, Eric Villain, and Anthony Foulonneau. 2022. Evaluating usability and user experience of AR applications in VR simulation. *Frontiers in Virtual Reality* 3 (2022), 881318. doi:10.3389/frvir.2022.881318
[36] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*. Springer, 63–76. doi:10.1007/978-3-540-89350-9_6

[37] Junjie Li, Yichun Jin, Shuai Lu, Wei Wu, and Pengfei Wang. 2020. Building environment information and human perceptual feedback collected through a combined virtual reality (VR) and electroencephalogram (EEG) method. *Energy and Buildings* 224 (2020), 110259. doi:10.1016/j.enbuild.2020.110259

[38] Ann Light, Béatrice Cahour, and Nuno Otero. 2010. Reflections on reflection: how critical thinking relates to collecting accounts of experience using explicitation techniques. (2010).

[39] Camila Loiola Brito Maia and Elizabeth Sucupira Furtado. 2016. A systematic review about user experience evaluation. In *International conference of design, user experience, and usability*. Springer, 445–455. doi:10.1007/978-3-319-40409-7_42

[40] Seonaidh McDonald. 2005. Studying actions in context: a qualitative shadowing method for organizational research. *Qualitative research* 5, 4 (2005), 455–473. doi:10.1177/1468794105056923

[41] Yihan Mei, Zhao Wu, Junnan Yu, Wenan Li, and Zhibin Zhou. 2025. GeneyMAP: Exploring the Potential of GenAI to Facilitate Mapping User Journeys for UX Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22. doi:10.1145/3706598.3713479

[42] Clara Moge, Katherine Wang, and Youngjun Cho. 2022. Shared user interfaces of physiological data: Systematic review of social biofeedback systems and contexts in hci. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16. doi:10.1145/3491102.3517495

[43] Débora Monteiro Moretti, Chad M Baum, Michael Wustmans, and Stefanie Bröring. 2022. Application of journey maps to the development of emergent sustainability-oriented technologies: Lessons for user involvement in agriculture. *Business Strategy & Development* 5, 3 (2022), 209–221. doi:10.1002/bsd2.192

[44] Gjoko Muratovski. 2021. Research for designers: A guide to methods and practice. (2021).

[45] Alaeddin Nassani, Huidong Bai, Gun Lee, and Mark Billinghurst. 2015. Tag it! ar annotation using wearable sensors. In *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*. 1–4. doi:10.1145/2818427.2818438

[46] Richard E Nisbett and Timothy D Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological review* 84, 3 (1977), 231. doi:10.1037/0033-295X.84.3.231

[47] Yunfang Niu, Danli Wang, Ziwei Wang, Fan Sun, Kang Yue, and Nan Zheng. 2019. User experience evaluation in virtual reality based on subjective feelings and physiological signals. *Electronic Imaging* 32 (2019), 1–11. doi:10.2352/J.ImagingSci.Technol.2019.63.6.060413

[48] Alfonsina Pagano, Augusto Palombini, Guido Bozzelli, Maurizio De Nino, Ivana Cerato, and Stefano Ricciardi. 2020. ArkaeVision VR game: User experience research between real and virtual paestum. *Applied Sciences* 10, 9 (2020), 3182. doi:10.3390/app10093182

[49] Sara Papi, Peter Polak, Dominik Macháček, and Ondřej Bojar. 2025. How "Real" is Your Real-Time Simultaneous Speech-to-Text Translation System? *Transactions of the Association for Computational Linguistics* 13 (2025), 281–313.

[50] Claire Petitmengin, Anne Remillieux, and Camila Valenzuela-Moguillansky. 2019. Discovering the structures of lived experience: Towards a micro-phenomenological analysis method. *Phenomenology and the Cognitive Sciences* 18, 4 (2019), 691–730. doi:10.1007/s11097-018-9597-4

[51] Eva Pietroni, Alfonsina Pagano, and Bruno Fanini. 2018. UX Designer and Software Developer at the mirror: Assessing sensory immersion and emotional involvement in Virtual Museums. *Studies in Digital Heritage* 2, 1 (2018), 13–41. doi:10.14434/sdh.v2i1.24634

[52] Marta Pizzolante, Sabrina Bartolotta, Eleonora Diletta Sarcinella, Alice Chirico, and Andrea Gaggioli. 2024. Virtual vs. real: exploring perceptual, cognitive and affective dimensions in design product experiences. *BMC psychology* 12, 1 (2024), 10. doi:10.1186/s40359-023-01497-5

[53] Mike Press and Rachel Cooper. 2017. *The design experience: the role of design and designers in the twenty-first century*. Routledge.

[54] Mirjana Prpa, Sarah Fdili-Alaoui, Thecla Schiphorst, and Philippe Pasquier. 2020. Articulating experience: Reflections from experts applying micro-phenomenology to design research in HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14. doi:10.1145/3313831.3376664

[55] Francisco Rebelo, Paulo Noriega, Emília Duarte, and Marcelo Soares. 2012. Using virtual reality to assess user experience. *Human factors* 54, 6 (2012), 964–982. doi:10.1177/0018720812465006

[56] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. 2007. Affective interactions using virtual reality: the link between presence and emotions. *Cyberpsychology & behavior* 10, 1 (2007), 45–56. doi:10.1089/cpb.2006.9993

[57] Jean-Philippe Rivière, Louis Vinet, and Yannick Prié. 2024. Towards the use of virtual reality prototypes in architecture to collect user experiences: An assessment of the comparability of patient experiences in a virtual and a real ambulatory pathway. *International Journal of Human-Computer Studies* 192 (2024), 103342. doi:10.1016/j.ijhcs.2024.103342

[58] Daniel L Schacter, Scott A Guerin, and Peggy L St Jacques. 2011. Memory distortion: An adaptive perspective. *Trends in cognitive sciences* 15, 10 (2011), 467–474. doi:10.1016/j.tics.2011.08.004

[59] Luyao Shen, Xian Wang, Sijia Li, Lik-Hang Lee, Mingming Fan, and Pan Hui. 2025. EmojiChat: Toward designing emoji-driven social interaction in VR museums. *International Journal of Human–Computer Interaction* 41, 11 (2025), 6987–7003. doi:10.1080/10447318.2024.2387902

[60] Donghee Shin. 2018. Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in human behavior* 78 (2018), 64–73. doi:10.1016/j.chb.2017.09.012

[61] Mel Slater and Maria V Sanchez-Vives. 2016. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI* 3 (2016), 74. doi:10.3389/frobt.2016.00074

[62] Harrison Jesse Smith and Michael Neff. 2018. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12. doi:10.1145/3173574.3173863

[63] Bernhard P Staresina and Maria Wimber. 2019. A neural chronometry of memory recall. *Trends in cognitive sciences* 23, 12 (2019), 1071–1085. doi:10.1016/j.tics.2019.09.011

[64] Marc Stickdorn and Jakob Schneider. 2012. *This is service design thinking: Basics, tools, cases*. John Wiley & Sons.

[65] Jaclynn V Sullivan. 2018. Learning and embodied cognition: A review and proposal. *Psychology Learning & Teaching* 17, 2 (2018), 128–143. doi:10.1177/1475725717752550

[66] Jingwei Sun, Zhongyue Zhang, Mengyang Wang, Nianlong Li, Zhangwei Lu, Yan Xiang, Liuxin Zhang, Yu Zhang, Qianying Wang, and Mingming Fan. 2025. Chorus of the Past: Toward Designing a Multi-agent Conversational Reminiscence System with Digital Artifacts for Older Adults. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22. doi:10.1145/3706598.3713810

[67] Angelina R Sutin and Richard W Robins. 2007. Phenomenology of autobiographical memories: The memory experiences questionnaire. *Memory* 15, 4 (2007), 390–411. doi:10.1080/09658210701256654

[68] David Swallow, Mark Blythe, and Peter Wright. 2005. Grounding experience: relating theory and method to evaluate the user experience of smartphones. In *ACM International Conference Proceeding Series*, Vol. 132. 91–98.

[69] Luma Tabbaa, Ryan Searle, Saber Mirzaee Bafti, Md Moinul Hossain, Jittrapol Intarasisrisawat, Maxine Glancy, and Chee Siang Ang. 2021. Vreed: Virtual reality emotion recognition dataset using eye tracking & physiological measures. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 5, 4 (2021), 1–20. doi:10.1145/34995002

[70] Katy Tcha-Tokey, Olivier Christmann, Emilie Loup-Escande, and Simon Richir. 2016. Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *International Journal of Virtual Reality* 16, 1 (2016), 33–48. doi:10.20870/IJVR.2016.16.1.2880

[71] Cyril Tjahja and Joyce Yee. 2022. Being a sociable designer: Reimagining the role of designers in social innovation. *CoDesign* 18, 1 (2022), 135–150. doi:10.1080/15710882.2021.2021244

[72] Christopher W Totten. 2019. *Architectural approach to level design*. CRC Press. doi:10.1201/b21989

[73] Jakob Trischler, Anita Zehrer, and Jessica Westman. 2018. A designerly way of analyzing the customer experience. *Journal of Services Marketing* 32, 7 (2018), 805–819. doi:10.1108/JSM-04-2017-0138

[74] Fred Turner. 2013. *The democratic surround: Multimedia & American liberalism from World War II to the psychedelic Sixties*. University of Chicago Press.

[75] Michael Vallance and Phillip A Towndrow. 2022. Perspective: Narrative storyliving in virtual reality design. *Frontiers in virtual reality* 3 (2022), 779148. doi:10.3389/frvir.2022.779148

[76] Arnold POS Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. 2010. User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries*. 521–530. doi:10.1145/1868914.1868973

[77] Pierre Vermersch. 1994. The explicitation interview. *French original ESF* (1994).

[78] Mingming Wang, Xiaoyu Zhang, Weiyue Gao, Yihan Mei, Haiyan Jiang, Yusheng Guo, Henry Been-Lirn Duh, and Zhibin Zhou. 2025. ImmerJM: A 3D Design Tool for Creating User Journey Maps Based on Immersive Virtual Environments. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 1224–1234. doi:10.1109/ISMAR67309.2025.00128

[79] Huaxin Wei, Jim Bizzocchi, and Tom Calvert. 2010. Time and space in digital game storytelling. *International Journal of Computer Games Technology* 2010, 1 (2010), 897217. doi:10.1155/2010/897217

[80] Christopher J Wilson and Alessandro Soranzo. 2015. The use of virtual reality in psychology: A case study in visual perception. *Computational and mathematical methods in medicine* 2015, 1 (2015), 151702. doi:10.1155/2015/151702

[81] Margaret Wilson. 2002. Six views of embodied cognition. *Psychonomic bulletin & review* 9, 4 (2002), 625–636. doi:10.3758/BF03196322

[82] Singer Witmer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* (1998). doi:10.1162/105474698565686

[83] Holly O Witteman, Selma Chipenda Dansokho, Heather Colquhoun, Angela Coulter, Michèle Dugas, Angela Fagerlin, Anik MC Giguere, Sholom Glouberman, Lynne Haslett, Aubri Hoffman, et al. 2015. User-centered design and the

development of patient decision aids: protocol for a systematic review. *Systematic reviews* 4, 1 (2015), 11. doi:10.1186/2046-4053-4-11

[84] Jacob O Wobbrock and Julie A Kientz. 2016. Research contributions in human-computer interaction. *interactions* 23, 3 (2016), 38–44. doi:10.1145/2907069

[85] David C Wynn and Claudia M Eckert. 2017. Perspectives on iteration in design and development. *Research in Engineering Design* 28, 2 (2017), 153–184. doi:10.1007/s00163-016-0226-3

[86] Chengyan Xu, Yasemin Demir-Kaymaz, Christina Hartmann, Marino Menozzi, and Michael Siegrist. 2021. The comparability of consumers' behavior in virtual reality and real life: a validation study of virtual reality based on a ranking task. *Food quality and preference* 87 (2021), 104071. doi:10.1016/j.foodqual.2020.104071

[87] Michael Yin and Robert Xiao. 2025. TravelGalleria: Supporting Remembrance and Reflection of Travel Experiences through Digital Storytelling in Virtual Reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–14. doi:10.1145/3706598.3713398

[88] Rongrong Yu, Gabrielle Schubert, and Ning Gu. 2023. Biometric analysis in design cognition studies: A systematic literature review. *Buildings* 13, 3 (2023), 630. doi:10.3390/buildings13030630

[89] Xue Yu, Stephen DiVerdi, Akshay Sharma, and Yotam Gingold. 2021. Scaffoldsketch: Accurate industrial design drawing in vr. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 372–384. doi:10.1145/3472749.3474756

[90] Xiaoxue Zang, Ying Xu, and Jindong Chen. 2021. Multimodal icon annotation for mobile applications. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. 1–11. doi:10.1145/3447526.3472064

[91] Chi Zhang, Wei Zeng, and Ligang Liu. 2021. UrbanVR: An immersive analytics system for context-aware urban design. *Computers & Graphics* 99 (2021), 128–138. doi:10.1016/j.cag.2021.07.006

[92] Zhibin Zhou, Lingyun Sun, Yuyang Zhang, Xuanhui Liu, and Qing Gong. 2020. ML lifecycle canvas: designing machine learning-empowered UX with material lifecycle thinking. *Human–Computer Interaction* 35, 5-6 (2020), 362–386. doi:10.1080/07370024.2020.1736075

[93] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 493–502. doi:10.1145/1240624.1240704

[94] Peter Zimmermann. 2008. Virtual reality aided design. A survey of the use of VR in automotive industry. In *Product engineering: tools and methods based on virtual reality*. Springer, 277–296. doi:10.1007/978-1-4020-8200-9_13

A APPENDIX - A

Table 5. The demographic information of the participants recruited for the formative study.

ID	Age	Gender	Design exp. (years)	UX collection exp. (years)	Education background	VR use frequency
D1	27	F	9	7	UX Design, Product Design	Seldom
D2	29	M	10	8	UX Design	Occasionally
D3	30	F	12	6	UX Design	Occasionally
D4	31	F	13	7	Virtual Reality	Always
D5	32	F	15	8	Product Design	Occasionally
D6	33	F	10	5	Service Design	Seldom
D7	28	F	8	6	UX Design	Occasionally
D8	32	F	8	8	Product Design	Occasionally
D9	28	F	7	5	Industrial Design	Frequently
R1	26	F	9	6	Service Design	Seldom
R2	38	M	18	9	Virtual Reality	Always
R3	29	M	11	7	Augmented Reality	Always
R4	27	F	9	7	UX Design	Seldom
R5	27	F	9	6	UX Design	Seldom
R6	25	F	8	6	Industrial Design	Occasionally
R7	31	F	6	3	Virtual Reality	Always
R8	29	F	7	7	UX Design	Occasionally
E1	24	F	6	0	Media Art Design	Seldom
E2	25	M	6	0	Media Art Design	Seldom
E3	22	F	4	0	Media Art Design	Seldom
E4	24	M	6	0.5	Media Art Design	Seldom
E5	24	M	6	0	Architecture Design	Seldom
E6	23	F	5	3	Product Design	Occasionally
E7	24	F	6	1	Media Art Design	Occasionally
E8	25	F	7	3	Product Design	Frequently
E9	24	F	6	1	Media Art Design	Seldom
E10	22	M	4	1	Interaction Design	Frequently
E11	24	M	4	2	Industrial Design	Occasionally
E12	25	M	5	1.5	Industrial Design	Occasionally
E13	28	M	9	0.5	Media Art Design	Seldom
E14	22	F	4	2	Intelligent Product Design	Occasionally
E15	22	F	4	1	Intelligent Product Design	Seldom
E16	28	F	8	0	Industrial Design	Occasionally
E17	24	F	6	0	Graphic Design	Seldom

Note. 1. Always: Participant utilizes VR in every project. 2. Frequently: Participant utilizes VR in the majority of projects. 3. Occasionally: Participant utilizes VR in about half of their projects. 4. Seldom: Participant utilizes VR only in several projects.

B APPENDIX - B

Table 6. The demographic information of the participants recruited for the user study.

ID	Age	Gender	Education background	VR use frequency
P1	30	F	Optometry	Seldom
P2	23	M	Media Art	Occasionally
P3	28	F	Optometry	Occasionally
P4	28	M	Mechanical Engineering	Seldom
P5	24	F	Media Art	Frequently
P6	22	F	Business Intelligence	Seldom
P7	25	F	Computer Science	Always
P8	24	F	Finance	Occasionally
P9	24	F	Civil Engineering	Seldom
P10	25	F	Media Art	Seldom
P11	23	F	Computer Science	Occasionally
P12	25	F	Multimedia Entertainment	Always
P13	18	F	Civil Engineering	Seldom
P14	25	F	Civil Engineering	Seldom
P15	23	M	Sustainability	Frequently
P16	26	M	Multimedia Entertainment	Occasionally
P17	24	F	Statistics	Seldom
P18	25	M	Civil Engineering	Seldom
P19	25	M	Biotechnology	Seldom
P20	28	M	Physics	Seldom

Note. 1. Always: Participant utilizes VR in every project. 2. Frequently: Participant utilizes VR in the majority of projects. 3. Occasionally: Participant utilizes VR in about half of their projects. 4. Seldom: Participant utilizes VR only in several projects.

C APPENDIX - C

Table 7. The Mean, SD and main effect of ANOVA test results of time spent

Dimensions	Mean (SD)			$F_{2,40}$	η_p^2	p
	1	2	3			
Time Spent	28.12 (8.67)	35.61 (13.03)	54.53 (13.68)	83.995	0.816	< **.001**

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 8. The interaction effects of Bonferroni corrected Post-hoc test results of time spent

Dimensions	p		
	1 vs 2	1 vs 3	2 vs 3
Time Spent	< **.05**	< **.001**	< **.001**

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 9. The mean, SD, and one-sample Wilcoxon signed-rank test results of errors

Dimensions	Combined method (Mean ± SD)	Z	p
Phases	0.25 (0.55)	1.890	.059
Scenarios	3.05 (2.09)	3.740	< **.001**
Actions	2.20 (1.82)	3.655	< **.001**
Touch Points	1.00 (0.79)	3.573	< **.001**
Emotions	1.70 (1.34)	3.436	< **.001**
Pain Points	0.35 (0.59)	2.333	< **.05**
Opportunities	0.30 (0.57)	2.121	< **.05**

Table 10. The mean, SD, and main effects of Friedman test results of valuable insights

Dimensions	Mean (SD)			$\chi^2(2)$	p
	1	2	3		
Phases	5.95 (1.00)	6.55 (0.89)	6.85 (0.75)	10.720	< **.01**
Scenarios	17.30 (2.25)	25.60 (2.68)	26.35 (2.23)	37.130	< **.001**
Actions	15.65 (2.83)	18.55 (3.07)	19.65 (2.41)	29.304	< **.001**
Touch Points	9.65 (2.68)	11.85 (2.41)	12.60 (2.19)	22.116	< **.001**
Emotions	4.75 (1.12)	6.25 (0.85)	6.40 (0.82)	18.091	< **.001**
Pain Points	5.20 (0.83)	6.05 (1.19)	6.70 (1.03)	16.889	< **.001**
Opportunities	5.40 (0.88)	5.35 (0.88)	6.05 (1.00)	6.559	< **.05**

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 11. The interaction effects of Bonferroni corrected Post-hoc test results of valuable insights

Dimensions	$\|Z\|(p)$		
	1 vs 2	1 vs 3	2 vs 3
Phases	1.739 (.246)	2.530 (< **.05**)	0.791 (1.00)
Scenarios	4.032 (< **.001**)	5.455 (< **.001**)	1.423 (.155)
Actions	3.083 (< **.01**)	4.981 (< **.001**)	1.8897 (.058)
Touch Point	2.530 (< **.05**)	4.348 (< **.001**)	1.18 (.069)
Emotions	3.083 (< **.01**)	3.558 (< **.001**)	0.474 (1.00)
Pain Points	1.581 (.342)	3.637 (< **.001**)	2.055 (.119)
Opportunities	0.395 (1.00)	2.214 (.081)	1.818 (.207)

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 12. The mean, SD, and main effects of Friedman test results of self-assessment

Dimensions	Mean (SD)			$\chi^2(2)$	p
	1	2	3		
Completeness	6.95 (1.28)	8.40 (1.14)	8.50 (1.28)	21.460	< **.001**
Accuracy	7.90 (1.33)	8.35 (1.35)	8.55 (1.19)	0.750	.687
Clarity	7.50 (1.36)	7.95 (1.43)	8.35 (1.04)	4.3344	.114
Consistency	7.05 (0.94)	8.70 (1.08)	8.65 (1.39)	15.031	< **.001**
Relevance	7.05 (0.89)	8.20 (1.06)	8.40 (1.10)	19.574	< **.001**
Insight	6.05 (1.76)	8.15 (1.23)	7.95 (1.05)	20.000	< **.001**

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 13. The interaction effects of Bonferroni corrected Post-hoc test results of self-assessment

Dimensions	$\lvert Z \rvert (p)$		
	1 vs 2	1 vs 3	2 vs 3
Completeness	3.637 (< **.01**)	3.479 (< **.01**)	0.158 (1.00)
Accuracy	0.237 (1.00)	0.712 (1.00)	0.474 (1.00)
Clarity	1.265 (.618)	1.818 (.207)	0.553 (1.00)
Consistency	3.241 (< **.01**)	2.688 (< **.05**)	0.553 (1.00)
Relevance	3.083 (< **.01**)	3.558 (< **.01**)	0.474 (1.00)
Insight	3.953 (< **.001**)	3.162 (< **.01**)	0.791 (1.00)

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 14. The mean, SD, and main effects of Friedman test results of NASA-TLX

Dimensions	Mean (SD)			$\chi^2(2)$	p
	1	2	3		
Mental Demand	7.05 (2.42)	7.05 (2.01)	7.00 (2.29)	0.086	.958
Physical Demand	8.25 (1.71)	6.85 (2.62)	6.80 (2.02)	10.400	**< .01**
Temporal Demand	7.35 (1.98)	7.70 (1.92)	6.70 (2.34)	10.886	**< .01**
Effort	6.40 (2.14)	6.65 (2.06)	6.75 (2.43)	0.241	.886
Performance	7.00 (1.92)	7.35 (1.39)	7.20 (1.79)	13.968	**< .001**
Frustration Level	7.10 (2.77)	7.65 (1.60)	7.85 (2.03)	0.984	.611

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 15. The interaction effects of Bonferroni corrected Post-hoc test results of NASA-TLX

Dimensions	$\lvert Z \rvert (p)$		
	1 vs 2	1 vs 3	2 vs 3
Mental Demand	0.237 (1.00)	0.000 (1.00)	0.237 (1.00)
Physical Demand	2.530 (< **.05**)	2.688 (< **.05**)	0.158 (1.00)
Temporal Demand	1.423 (.464)	3.083 (< **.01**)	1.660 (.291)
Effort	0.079 (1.00)	0.395 (1.00)	0.316 (1.00)
Performance	2.925 (< **.05**)	2.767 (< **.05**)	0.158 (1.00)
Frustration Level	0.553 (1.00)	0.870 (1.00)	0.316 (1.00)

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 16. The mean, SD, and main effects of Friedman test results of Immersion

Dimensions	Mean (SD)			$\chi^2(2)$	p
	1	2	3		
Expression Naturalness	8.00 (1.12)	6.75 (1.25)	8.20 (1.64)	13.164	**< .001**
Exploration Freedom	6.85 (1.98)	8.50 (1.24)	8.35 (1.35)	10.764	**< .01**
Engaged Level	6.65 (1.98)	8.35 (0.75)	8.20 (1.58)	13.583	**< .001**
Overall Immersion	7.20 (1.61)	8.60 (0.99)	8.30 (1.42)	10.866	**< .01**
Cognitive Fluency	6.10 (1.83)	7.85 (1.04)	7.65 (0.99)	13.441	**< .001**
Interference Factors	7.50 (1.82)	7.30 (2.15)	7.55 (1.82)	0.471	.790
Focus	7.20 (1.99)	7.30 (1.72)	8.30 (1.45)	2.818	.244

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 17. The interaction effects of Bonferroni corrected Post-hoc test results of Immersion

Dimensions	$\lvert Z \rvert (p)$		
	1 vs 2	1 vs 3	2 vs 3
Expression Naturalness	2.609 (< **.05**)	0.474 (1.00)	3.083 (< **.01**)
Exploration Freedom	2.767 (< **.05**)	2.451 (< **.05**)	0.316 (1.00)
Engaged Level	2.609 (< **.05**)	3.320 (< **.01**)	0.712 (1.00)
Overall Immersion	2.688 (< **.05**)	2.530 (< **.05**)	0.158 (1.00)
Cognitive Fluency	3.241 (< **.01**)	2.451 (< **.05**)	0.791 (1.00)
Interference Factors	0.474 (1.00)	0.000 (1.00)	0.474 (1.00)
Focus	0.237 (1.00)	1.423 (.464)	1.186 (.707)

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 18. The mean, SD, and main effects of Friedman test results of MEQ

Dimensions	Mean (SD)			$\chi^2(2)$	p
	1	2	3		
Vividness	6.31 (2.31)	8.47 (1.16)	7.94 (1.86)	12.182	< **.01**
Coherence	6.44 (1.80)	8.30 (1.17)	7.99 (1.36)	9.455	< **.01**
Accessibility	6.15 (2.15)	8.46 (1.00)	7.85 (1.67)	16.421	< **.001**
Sensory Detail	5.94 (2.34)	8.16 (1.39)	8.25 (0.94)	18.487	< **.001**
Emotional Intensity	4.99 (2.14)	6.72 (2.05)	6.93 (1.67)	9.026	< **.05**
Visual Perspective	6.83 (2.67)	8.63 (1.75)	8.33 (1.84)	12.108	< **.01**
Time Perspective	4.97 (2.49)	4.62 (2.28)	4.72 (2.21)	0.085	.959
Sharing	3.67 (1.90)	4.07 (2.27)	4.03 (1.75)	0.899	.638
Distancing	7.05 (2.14)	8.49 (1.20)	7.98 (2.03)	10.338	< **.01**
Valence	6.43 (2.00)	6.38 (2.23)	6.68 (2.07)	2.054	.358

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.

Table 19. The interaction effects of Bonferroni corrected Post-hoc test results of MEQ

Dimensions	$\|Z\|(p)$		
	1 vs 2	1 vs 3	2 vs 3
Vividness	2.925 (< **.01**)	3.004 (< **.01**)	0.079 (1.00)
Coherence	2.688 (< **.05**)	2.530 (< **.05**)	0.158 (1.00)
Accessibility	3.795 (< **.001**)	2.846 (< **.05**)	0.343 (1.00)
Sensory Detail	3.716 (< **.001**)	3.637 (< **.001**)	0.079 (1.00)
Emotional Intensity	2.214 (< **.05**)	2.767 (< **.05**)	0.553 (1.00)
Visual Perspective	3.162 (< **.01**)	2.530 (< **.05**)	0.632 (1.00)
Time Perspective	0.237 (1.00)	0.237 (1.00)	0.000 (1.00)
Sharing	0.385 (1.00)	0.580 (1.00)	0.752 (1.00)
Distancing	2.767 (< **.05**)	2.451 (< **.05**)	0.316 (1.00)
Valence	1.265 (.618)	1.107 (.805)	0.158 (1.00)

Note. 1: Retrospective method. 2: Concurrent method. 3: Combined method.