SpatialSlides: Integrating Presentation Software with Extended Reality for Contextualized 3D Exploration

Jung Who Nam
Kunsan National University
Gunsan, South Korea
jungwhonam@kunsan.ac.kr

Vuthea Chheang
San José State University
San José, USA
vuthea.chheang@sjsu.edu

Haichao Miao
Lawrence Livermore National Laboratory
Livermore, USA
miao1@llnl.gov

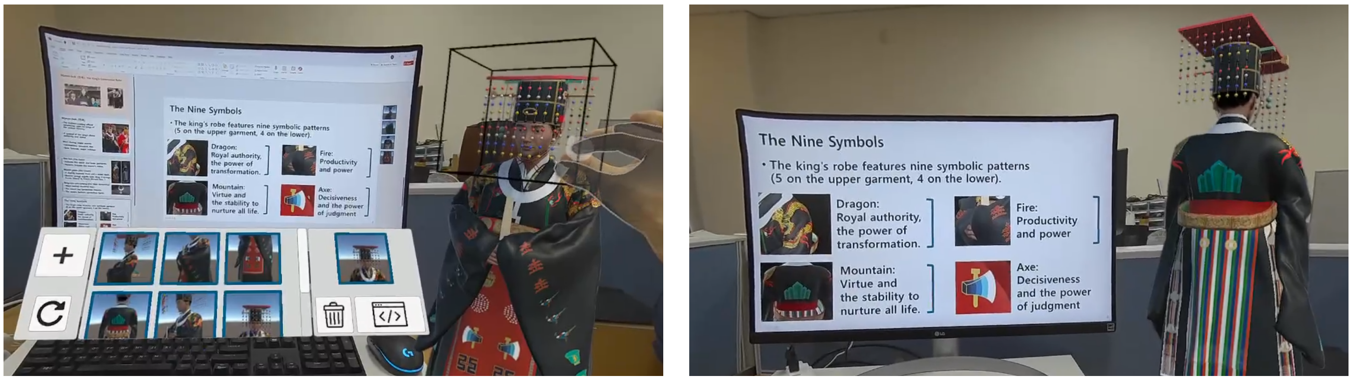

Figure 1: SpatialSlides enables instructors to signal to learners how textual concepts in the slide deck translate into spatial reality. (Left) Authoring Mode: The instructor captures the 3D model's current state and embed it into the active slide, generating a spatial keyframe. (Right) Viewing Mode: As the presentation advances, the XR client automatically animates the 3D model to the corresponding viewpoint, synchronizing the spatial visual with the instructional text.

Abstract

We present SpatialSlides, a system that connects standard presentation software with an XR client to guide learners from traditional 2D lecture materials to specific viewpoints within 3D content. While XR offers significant pedagogical value, it typically remains fragmented from the primary lecture medium, treated as a disconnected, standalone activity. SpatialSlides addresses this gap by synchronizing 2D slides in a presentation file with specific 3D spatial viewpoints. The system features two core modes: an authoring mode that enables users to capture spatial views and link them to specific 2D slides, and a presentation mode that automatically guides learners through these pre-authored spatial perspectives as the slides advance. Results from an exploratory study indicate that this integrated approach provides useful spatial guidance for learners and yields further design considerations for future hybrid workflows. To support future research, we also provide the underlying open-source framework, bridging XR clients directly with Microsoft PowerPoint.

VRST '26, Sendai, Japan

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/XXXXXXX.XXXXXXX

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**.

Keywords

Extended Reality, Authoring Tools, Presentation Interfaces, Educational Technology

ACM Reference Format:
Jung Who Nam, Vuthea Chheang, and Haichao Miao. 2026. SpatialSlides: Integrating Presentation Software with Extended Reality for Contextualized 3D Exploration. In *Proceedings of The 32nd ACM Symposium on Virtual Reality Software and Technology (VRST '26).* ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Extended Reality (XR), including Virtual and Augmented Reality (VR/AR), is transforming educational landscapes by enabling the exploration of complex 3D phenomena — from cellular structures to safety scenarios — with a level of engagement and spatial comprehension that is difficult to achieve with traditional static media [1, 19]. However, integrating XR into standard classroom settings introduces workflow challenges. While the slide deck remains the dominant medium for presentation and archival record [22, 31], immersive content typically resides in standalone applications or game engines. This fragmentation treats the instructional narrative (the slides) and the spatial visualization (the XR app) as separate entities. Consequently, instructors bear the burden of manually aligning their existing materials with 3D content and independently author narrative guidance for these unstructured immersive environments.

To address these workflow challenges, we explore a complementary approach. Rather than assuming open-ended exploration is the only mode for classroom XR, we investigate how immersive tools can function as a directly integrated spatial extension of existing linear course materials. Conceptually, this approach builds upon foundational AR research that embeds 3D content within physical artifacts like books and documents [8, 24], as well as cross-device systems that extend desktop applications into immersive space [21, 55]. We apply this paradigm to the delivery of linear educational narratives: slide-based presentations.

To achieve this, we introduce *SpatialSlides*, a system that enables instructors to explicitly signal to learners how textual concepts in traditional course materials translate into spatial reality (see Fig. 1). By implementing this concept as an extension to a standard presentation application, e.g., Microsoft PowerPoint, we leverage desktop tools that are already ubiquitous in education. Within this familiar environment, instructors capture specific 3D perspectives in an XR viewer and embed them directly into individual 2D slides. When viewed in presentation mode, these embedded perspectives transform the slide deck into a sequence of keyframes: as the presentation advances, the linked XR client automatically animates the 3D model to match the viewpoint intended by the current slide. However, creating a slide deck is rarely a one-time process; instructors constantly adapt their materials after generating an initial draft — frequently reordering topics, changing content length for specific audiences, or merging separate lectures into a new presentation.

To support this iterative process, the underlying architecture of SpatialSlides embeds 3D viewpoint parameters directly within standard presentation image components (see Fig. 3d). By treating this augmented image as a self-contained spatial token, the XR content inherits the editing affordances of traditional slide software — allowing instructors to copy, paste, or reorder the 3D timeline using familiar desktop editing practices. Crucially, because this data does not interfere with native presentation features, a single presentation file can serve as a synchronized 3D scaffold when connected to the XR viewer while remaining immediately usable as a standard 2D lecture in a traditional classroom. This allows educators to adapt their delivery method based on their current pedagogical needs or available hardware.

In summary, our specific contributions are:

- **(primary) the design of a hybrid presentation system** that bridges the gap between immersive 3D environments and standard 2D lecture slides to align spatial content with the narrative flow,
- **the development of an open-source framework**[1] that connects XR clients directly with ubiquitous desktop productivity applications (Microsoft PowerPoint), and
- **user feedback and insights from an exploratory study (**N = 16**)** confirming the system's baseline usability and providing suggestions for future hybrid presentation workflows.

[1] https://github.com/JungWhoNam/SpatialSlides

2 Related Work

The benefits of using XR in education are well-established in prior work, with studies demonstrating its potential to enhance engagement, comprehension, and recall through the affordances of immersion and interactivity [1, 13, 19, 27, 41]. However, the literature also highlights significant challenges, such as the risk of cognitive overload and the importance of applying sound instructional principles [30, 39, 47]. The focus of this paper is not to re-evaluate these established benefits and drawbacks. Rather, we accept the consensus that different modalities are not replacements for one another but are complementary tools to be used interchangeably [12, 57]. We therefore focus on the design challenge of architecting systems that facilitate this seamless integration of media. The following sections review the state of the art in authoring tools that attempt to bridge these workflows.

2.1 XR Authoring Paradigms and Barriers

The creation of interactive XR content has traditionally been the domain of skilled developers using game engines such as Unity and Unreal. While powerful, these platforms present a high barrier to entry for non-programmers [29, 45]. A central challenge for accessible authoring tools is balancing a low "threshold" (ease of use for beginners) with a high "ceiling" (the power to create complex experiences) [43]. Beyond the initial learning curve, tools that require users to constantly switch between a desktop interface for authoring and a VR headset for testing make debugging difficult and frustrating [3].

To address these challenges, researchers have explored several authoring paradigms that move beyond traditional coding. One common approach is the use of Visual Programming Languages (VPLs), like the immersive dataflow editors found in [14, 60, 63], to make authoring more intuitive [62]. Other approaches use Programming by Demonstration (PbD) [44], where users define interactions by performing them directly in XR, as seen in tools like Mucho for multimodal interactions [33] and Teachable Reality for tangible interactions with real-world objects [42].

Research in pedagogical and training contexts has further refined the requirements for these tools. There is a demand for minimal coding solutions [29, 45] and the ability to import and reuse assets is consistently cited as a critical feature for usability [16, 58]. However, several critical gaps remain. As the effectiveness of a learning experience is highly dependent on the learner's prior knowledge [20, 52] and the specific learning content [9, 32, 51, 54], many tools fail to support diverse learners by producing content for only a single level of immersion [53]. To address this, some tools like XRAuthor have been developed to generate content for multiple levels of immersion — such as 2D video, 360° interactive video, and full VR — from a single authoring process [53]. However, even with these advancements, these tools still force instructors to adopt new, standalone workflows rather than building upon the familiar linear presentation materials they already use.

2.2 Extending Traditional Media into XR

To bypass the friction of adopting new workflows, another paradigm focuses on using familiar, traditional media as a foundation for authoring immersive content. This approach leverages users' existing skills and artifacts while increasing compatibility between VR and non-VR working modes [6].

One thread of research focuses on augmenting physical media to create hybrid interactive experiences. A primary focus has been on paper-based documents to retain their advantages, such as robustness, portability, and flexibility [37]. By overlaying virtual content, AR books, e.g., [8, 25, 56], can convey information difficult to present in static pages, such as temporal changes and 3D scenes. Furthermore, AR enables interactivity that supports reader-driven storytelling through features like providing details on demand [4, 24] and dynamic changes to storylines [50]. Systems like Paper Trail [48], Living Paper [15], and HoloDoc [34] exemplify this by providing tools for instructors and children to augment physical documents with AR content.

Another thread of research explores using XR HMDs to enhance interaction with familiar desktop and mobile applications. By integrating these immersive technologies, developers can establish significantly larger display spaces [7, 21], present world-grounded stereoscopic information that users can interact with directly [6, 21], and even increase privacy by rendering virtual content exclusively through headsets [7]. For instance, Gesslein et al. [21] introduced a system that extends a mobile spreadsheet application with VR, enabling users to view multiple tabs simultaneously and expose dependencies between cells. This paradigm of using VR to augment traditional 2D software has also been applied to data visualization [61, 64] and modeling [38, 55, 65]. The pioneering DART toolkit followed a similar philosophy by extending the 2D timeline-based workflow of Macromedia Director to allow multimedia designers to author AR experiences [36].

inally, this paradigm has recently been applied to presentation systems. Recent studies on XR presentations validate our approach by outlining design considerations for hybrid workflows [23] and demonstrating techniques to enhance spatial understanding [59]. Notably, PoVRPoint [6] utilizes VR's expansive display space to facilitate the search for specific 2D slides and enhance desktop editing tasks. In contrast, SpatialSlides employs the XR environment as a complementary 3D display for spatial content that cannot be effectively conveyed through 2D slides alone. We therefore focus on interactive techniques for capturing 3D perspectives and linking existing presentation slides with specific spatial viewpoints according to the instructor's pedagogical intent.

3 Design Criteria and Concepts

Based on the challenges identified in prior sections and the literature in XR authoring and educational technology, we established core design criteria for our system:

C1 **Accessible Authoring:** Effective tools must enable domain experts to curate spatial narratives using interactions they already know, without requiring programming skills or knowledge of complex 3D software [29, 43, 45].

C2 **Integration with Existing Workflows:** The solution should support the user's primary workflow (e.g., presentation software), leveraging existing skills, content libraries, and file management practices [6, 23, 29, 36].

C3 **Sophisticated Instructional Structures:** The system should support complex instructional flows where multiple, choreographed views of a single 3D model can be linked across several slides to build a concept step-by-step [8, 24].

C4 **Flexible Navigation:** The interaction paradigm should provide a clear, guided path synchronized with the narrative while allowing learners to temporarily break away for self-directed exploration [12, 22, 23, 31, 59].

To ensure seamless integration into the familiar lecture materials, the XR system must prioritize an authoring process that respects the existing 2D slide workflow (leading to C2). Furthermore, because our target users are students and educators rather than developers, the tools must be accessible to those who may lack expertise in 3D modeling software or game engines (leading to C1). Finally, unlike typical AR books where a single model often follows a simple one-to-one mapping with a physical page, lecture slides require more complex delivery. A single concept may need to be deconstructed across multiple specific 3D viewpoints, requiring sophisticated choreography beyond simple model loading (leading to C3). To support active learning, the system must balance this structured guidance with the ability for learners to freely explore spatial content when needed (leading to C4), while providing a mechanism to easily return to the narrative flow.

3.1 Design Concepts

To address the established criteria, our framework implements three key design concepts:

Slide-Driven Spatial Narrative — To support sophisticated instruction (C3), we introduce a mapping between the linear presentation deck and the spatial environment. Fundamentally, this concept establishes the slide deck as the primary state controller: each slide serves as a distinct spatial keyframe, where advancing the slide automatically transports the viewer to the next viewpoint via a smooth transition. This synchronization further extends to internal slide animations, allowing a single slide to trigger a sequence of sub-views for detailed step-by-step explanations. To ensure flexible navigation (C4), we enable learners to temporarily detach for free exploration before seamlessly re-synchronizing with the narrative.

Snapshot Authoring — To support accessible authoring (C1), we map the complex task of 3D camera placement to the familiar mental model of taking a photograph. An author simply positions the model in the immersive view and captures it; this action instantly inserts a static image of that view onto the active slide. Because this proxy image acts as a standard image object, users can leverage existing workflow skills (C2): copying the image inherits the embedded metadata, effectively duplicating the viewpoint, and applying standard presentation animations to these images drives the multi-step spatial choreography described above.

Portable Embedded Metadata — To ensure seamless integration with existing workflows (C2), rather than relying on external save files — which fragment the document and complicate sharing — SpatialSlides embeds spatial metadata directly into native file structures (e.g., image Alt-Text). This ensures the presentation remains a shareable unit that retains backward compatibility with standard viewers while activating immersive features only when the add-on is present.

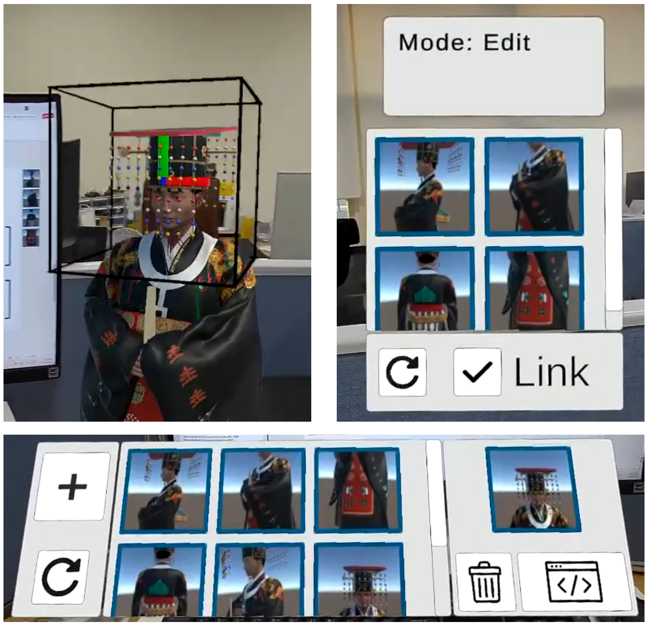

Figure 2: Core UI components implementing the design concepts. (Top-left) The Reference Frame defines the capture boundary for the snapshot authoring metaphor. (Top-right) The Slide Linker visualizes the synchronization link between the active slide and the 3D view. (Bottom) The Data View Manager stores captured snapshots as reusable spatial keyframes.

4 Detailed Design and Implementation

To implement the design concepts introduced in Section 3, SpatialSlides provides a set of immersive user interface (UI) components. The system automatically mirrors the host application's state, switching between Edit Mode (for authoring) and View Mode (for presentation).

4.1 System Overview: Core UI Components

The immersive space is structured around three primary UI components that provide a consistent interface for the user (Fig. 2).

- **Reference Frame:** A designated 3D bounding box that acts as the virtual stage. This frame serves as the coordinate origin for all spatial captures, ensuring that content authored in one physical environment remains consistent when presented in another.
- **Slide Linker:** The primary control panel for managing the synchronization state. It visualizes the connection between the current slide and the AR view, displaying an "empty eye" icon to provide unambiguous feedback when the active slide lacks a linked spatial keyframe.
- **Data View Manager:** A spatial library for all captured 3D views. It displays views as clickable thumbnails, allowing the author to recall, modify, or delete saved perspectives.

4.2 Edit Mode: Linking 3D Views to Slides

The primary goal of Edit Mode is to allow the presenter to define specific 3D perspectives — such as wide establishing shots or detailed close-ups — that link to the content of a specific slide. The presenter executes the snapshot metaphor via a four-step process (Fig. 3) to create synchronized spatial narratives.

4.2.1 Capturing 3D Views. The authoring process begins in the XR space, where the user manipulates the 3D model relative to the Reference Frame to compose the desired shot (Fig. 3a). Using native hand-tracking, the user utilizes pinch-to-grab gestures to physically position, rotate, and scale the virtual artifact. The frame remains visible to provide spatial context. When the capture action is triggered (Fig. 3b), the system saves an image and records the model's transformation matrix (position, rotation, scale) relative to the frame origin, saving it as a reusable spatial keyframe.

4.2.2 Linking and Managing Views on the Desktop. To synchronize this view with the presentation, the user initiates a link action from the AR interface (Fig. 3c). This triggers the system to generate a thumbnail image containing the serialized 3D metadata and insert it onto the active PowerPoint slide (Fig. 3d). Importantly, this thumbnail serves as a self-contained data object. Because the spatial coordinates are embedded within the image's properties (Alt-Text), standard PowerPoint operations become spatial authoring tools: copying the thumbnail to a new slide duplicates the camera angle, while deleting it removes the link.

4.2.3 Sequencing Multiple Views in a Slide. SpatialSlides leverages the native PowerPoint animation timeline to sequence multiple 3D views within a single slide. When a user links additional views, the system automatically inserts each new thumbnail vertically below the previous one. Instructors can sequence these spatial transitions by applying standard "Appear" animations to the thumbnails. In View Mode, the XR environment automatically transitions to the new perspective whenever a thumbnail appears on the screen. Furthermore, authors can seamlessly synchronize these 3D transitions with 2D visual aids, e.g., text blocks or highlighting shapes, by triggering their "Appear" animations simultaneously (Fig. 5). This co-animation ensures that a single click executes both the 2D visual cue and the immersive 3D movement, aligning the learner's textual and spatial attention.

4.3 View Mode: Viewing the Presentation

View Mode is designed to support a fluid transition between a guided, narrative-driven tour and unscripted, free exploration. To reduce visual clutter, authoring components are hidden, and the Reference Frame becomes invisible by default.

4.3.1 Synchronized Navigation. In Guided Mode, the slide deck acts as the driver. As the presenter advances the slides, the 3D model in the AR space automatically interpolates to the perspective linked to the new slide, aligning the learner's spatial attention with the instructor's intent.

4.3.2 Free Exploration and Resume Mechanic. To enable further investigation, the user can break away from the guided path at any time. Grabbing and moving the model automatically triggers Exploration Mode. During this state, the Reference Frame reappears (see Fig. 4), providing a visual cue that the model is no longer synchronized with the narrative. To return to the guided tour, the user can simply drop the model near the Reference Frame, triggering a smooth animation that resets the model to its synchronized state.

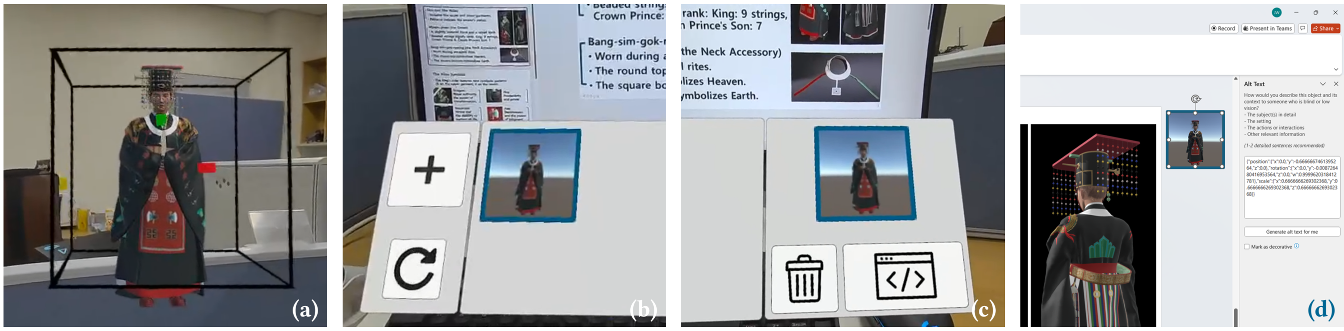

Figure 3: The snapshot authoring process. (a) The author manipulates the model within the Reference Frame to set a desired view. (b) They capture the view's state as a spatial keyframe using the Data View Manager. (c) The user initiates a link action in the Manager, which (d) embeds a thumbnail containing the 3D transform metadata directly onto the desktop slide.

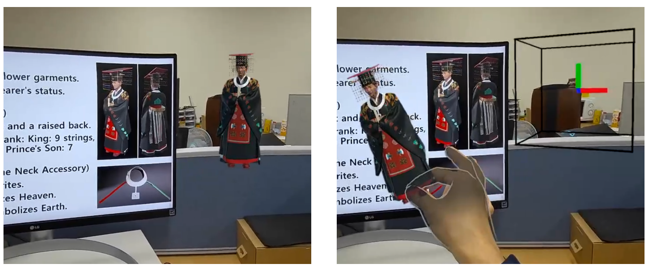

Figure 4: Transitioning between viewing states. (Left) In Guided Mode, the 3D model is synchronized with the slide and the Reference Frame is hidden. (Right) In Exploration Mode, triggered by moving the model, the Reference Frame appears to provide visual feedback.

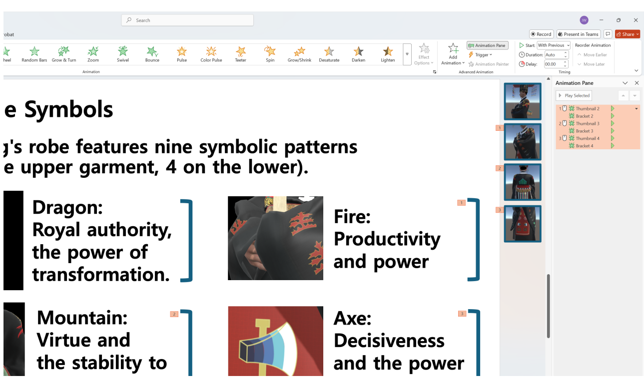

Figure 5: Sequencing multiple views in a slide. The native PowerPoint Animation Pane (right) shows a three-step animation sequence for Slide 3 (Fig. 7). This illustrates how inserted spatial thumbnails are grouped with 2D visual aids (e.g., bracket shapes) under a single "Appear" animation trigger, enabling simultaneous updates across both media.

Alternatively, navigating to a new slide overrides the exploration state, instantly re-synchronizing the view with the new content.

4.3.3 Advanced Presentation Controls. To accommodate different physical environments, the user can reposition the Reference Frame to move the entire virtual stage. Additionally, the synchronization toggle in the Slide Linker allows the presenter to temporarily decouple the model from the slide deck, maintaining a specific 3D view while advancing through multiple text slides.

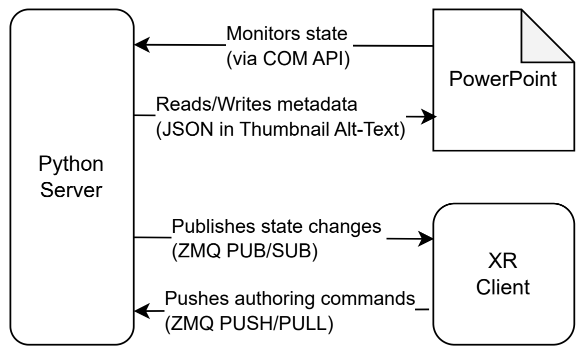

Figure 6: The SpatialSlides architecture.

4.4 System Architecture

SpatialSlides is implemented as a decoupled client-server architecture, comprising two primary components: a Python-based middleware that manages the PowerPoint instance via the Windows Component Object Model (COM) API, and an immersive XR client built in Unity (see Fig. 6). 3D models are managed within the Unity client build. These components communicate via ZeroMQ, establishing a low-latency, bidirectional link. The server uses a PUB/SUB pattern to broadcast state changes (e.g., slide advancement) to the client, and a PUSH/PULL pattern to receive real-time authoring commands (e.g., snapshot capture).

To handle data persistence without breaking the standard `.pptx` file format, we embed the 3D keyframe metadata (position, rotation, scale) as a JSON string directly within the `Alt-Text` property of the slide's proxy images (see Fig. 3d). At runtime, the Python middleware monitors the PowerPoint state for slide changes. When a new slide is detected, the middleware scans the slide's image objects, parses valid JSON strings from their `Alt-Text` fields, and transmits the spatial data to the Unity client. Upon receipt, the client updates the Slide Linker UI to reflect the active connection and immediately interpolates the 3D model to align with the retrieved viewpoint. Detailed class structures and protocols are provided in the Appendix.

5 Usage Scenario: Cultural Heritage

The cultural heritage domain is particularly well-suited for guided presentations. While the general forms of historical artifacts (e.g., traditional garments or mythical sculptures) are often familiar, they contain dense, hidden symbolisms. XR provides a unique opportunity to examine the spatial details of these objects. However, subtle material textures or obscured emblems can still be easily overlooked when learners explore 3D models without proper instructional guidance. To demonstrate how SpatialSlides addresses this pedagogical need, we describe the presentation flow of a module that layers 3D model views to unpack these complex historical details (Fig. 7).

Presentation Flow: The author designs a three-slide presentation focusing on a traditional Joseon King's royal attire. To describe the artifact, the presentation transitions the audience from a macro-level understanding down to micro-level details:

- **Slide 1 (Overview):** Establishes the general historical background using a single, zoomed-out view of the entire garment.
- **Slide 2 (Structural Details):** Focuses on the upper garments (the traditional hat and collar) using a sequence of close-up views.
- **Slide 3 (Symbolism):** Unpacks the meaning of specific emblems distributed across the clothing using localized, micro-level views.

Visual Linking and Pacing: Progressing into dense, micro-level details introduces the challenge of grouping multiple, spatially distinct 3D views within a single slide without visually overwhelming the audience. To manage this density on Slide 3, for instance, the instructor choreographs the scene to reveal one 3D perspective at a time. Driven by the animation synchronization described in Section 4.2.3, the system automatically interpolates the 3D model to focus on each specific emblem sequentially as the presenter advances the slide. Crucially, the presentation establishes a clear visual link between the modalities. This pairing uses the 2D slide elements to answer *what* the audience is learning, while the immersive 3D view immediately shows them *where* it is located on the artifact (Fig. 5).

6 Exploratory Study

We conducted a two-phase mixed-methods study ($N = 16$) to gather initial usability insights, specifically observing how users interact with the integrated viewing experience (Phase 1) and the authoring experience (Phase 2). We selected Korean Cultural Heritage as our evaluation scenario because it balances cultural familiarity with spatial complexity. While the general forms are recognizable to our participants, verifying distributed features (e.g., emblems) necessitates active 3D manipulation (rotation and scaling). Our evaluation was guided by the following research questions:

- **RQ1 (Learner Experience):** To what extent does an integrated, instruction-driven XR system influence learning outcomes, perceived workload, and user preference when compared to unguided 3D exploration?

Table 1: Descriptive statistics and statistical analysis results ($N = 16$). Wilcoxon Signed-Rank tests revealed no significant differences between the Guided and Unguided conditions across all measured metrics (all $p > .05$).

	Unguided	Guided	Wilcoxon		
Variable	M (SD)	M (SD)	W	p	r
Knowledge Recall (%)					
Spatial Recall	85 (20)	95 (9)	13.50	0.130	0.267
Information Recall	93 (14)	86 (17)	29.00	0.417	0.144
System Usability (0-100)					
SUS Score	86.56 (8.98)	84.76 (11.12)	49.50	0.548	0.106
User Engagement (1-5)					
Focused Attention	4.33 (0.72)	4.23 (0.75)	9.00	0.396	0.150
Perceived Usability	4.31 (0.61)	4.17 (0.62)	35.50	0.781	0.049
Aesthetic Appeal	4.40 (0.60)	4.38 (0.56)	21.50	0.905	0.021
Reward Factor	4.59 (0.55)	4.58 (0.39)	9.00	0.750	0.056
Workload (0-100)					
Mental Demand	46.88 (26.76)	36.56 (22.49)	24.00	0.132	0.266
Physical Demand	35.31 (25.46)	28.44 (20.14)	36.50	0.528	0.112
Temporal Demand	21.88 (19.14)	19.06 (17.34)	31.00	0.858	0.032
Performance	27.50 (19.83)	26.25 (21.79)	25.00	0.798	0.045
Effort	37.19 (24.15)	38.75 (26.74)	60.00	1.000	0.000
Frustration	15.62 (13.02)	20.62 (22.65)	14.00	0.308	0.180

- **RQ2 (Authoring Feasibility):** To what extent does the authoring workflow support non-experts in creating intuitive spatial presentations?

Participants: The study was approved by our Institutional Review Board, and all participants provided written informed consent prior to the experiment. We recruited 16 participants (2 female, 14 male, age $M = 24.31, SD = 3.24$) from the campus community. Eligibility criteria required basic computer literacy (mouse and keyboard proficiency) and normal or corrected-to-normal vision suitable for HMD usage. Each participant received KRW 10,000 in cash as compensation. The majority reported low familiarity with XR technology: 5 participants had no prior experience, 9 reported using VR only a few times a year, and 2 reported using VR monthly or more frequently. Prior knowledge regarding the learning topics (Korean sculpture and traditional clothing) was also low, averaging 2.00 ($SD = 0.52$) and 2.62 ($SD = 0.72$) on a 5-point scale.

Apparatus: The experiment was conducted using a Meta Quest 3 headset. We utilized the headset's color passthrough capability to allow the physical monitor and 3D augmented model to be viewed side-by-side. The physical workstation consisted of a 27-inch UHD 4K (3840 × 2160) IPS display paired with a standard keyboard and mouse. To encourage seamless transitions between spatial interactions (manipulating 3D models) and physical interactions (typing or using the mouse), we instructed participants to utilize the system's native hand tracking (Meta Interaction SDK) rather than VR controllers. Participants performed all tasks while seated at the desk, mirroring a typical workspace environment. During the session, the experimenter monitored the participant's view in real-time via the Meta Horizon Casting system.

Procedure: The entire session lasted about 60 minutes. To evaluate both ends of the slide-driven paradigm, participants completed a two-phase sequence and an interview:

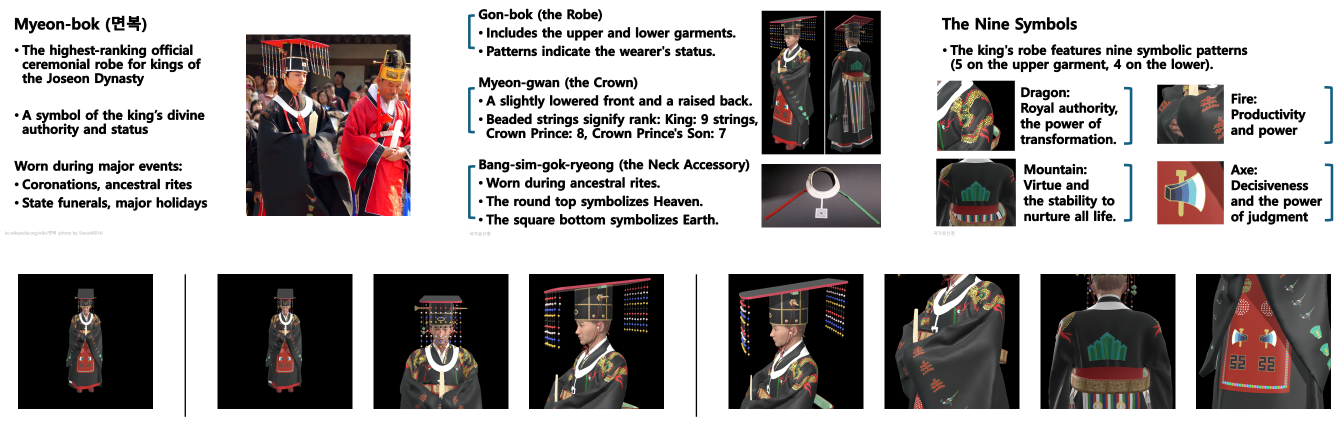

Figure 7: Joseon King's attire presentation module. (Top) The three-slide presentation breaks down the artifact from an overview to structural details. (Bottom) The eight corresponding 3D views are sized equally and arranged sequentially. Vertical dividers map the view clusters to their respective slides: Slide 1 establishes a single macro overview, Slide 2 captures three structural details, and Slide 3 unpacks four distinct emblems. *Image credits: Myeon-bok photograph by Steve46814 (Wikimedia Commons); additional images and the 3D model by the Korea Heritage Service.*

- **Phase 1 (Learner Experience):** Participants assumed the role of a student to evaluate the viewing experience.
- **Phase 2 (Instructor Workflow):** Participants assumed the role of a presenter to evaluate the authoring interface via an exploratory walkthrough.
- **Post-Study Interview:** The session concluded with a semi-structured interview regarding their viewing experience and the system's pedagogical feasibility. The interviews were recorded and transcribed for a six-phase thematic analysis [10]. One researcher performed the coding and the research team collaboratively refined the final overarching themes.

6.1 Phase 1: Learner Experience (RQ1)

6.1.1 Task, Design, and Procedure. To evaluate the integrated viewing experience, participants acted as students tasked with studying provided handouts to prepare for a post-task knowledge test. We developed two learning modules based on historical artifacts: a sculpture of a mythical creature (Appendix) and the royal attire of a Joseon King (Section 5). Each module consisted of a 3-slide handout paired with a 3D model. Participants were encouraged to actively explore the 3D models to verify the spatial forms and structural details described in the slides.

The study employed a within-subjects design with one independent variable: Interaction Condition. To mitigate learning effects, we utilized a balanced Latin Square design to counterbalance the assignment of conditions and the two learning topics. The two interaction conditions were:

- **Integrated-Guided (Our System):** Participants used SpatialSlides with the synchronized, slide-driven navigation features enabled.
- **Standalone-Unguided (Baseline):** Participants used the same viewer but with the synchronization link disabled. This required them to manually manipulate the model to match the text descriptions, simulating a standard standalone model viewer.

For each condition, participants first received a brief tutorial on the specific interaction technique. They then performed the learning task for 10 minutes. Immediately upon completion, they filled out the post-condition questionnaires and the topic-specific knowledge test. A 2-minute break was enforced before repeating the procedure for the second condition to minimize cognitive fatigue. After completing both conditions, participants performed a final comparative evaluation, rating the perceived usefulness of both systems side-by-side and stating their overall preference.

Measures: To evaluate the learner experience, we collected the following dependent variables:

- **Knowledge Recall:** A 10-item multiple-choice knowledge test comprising 5 spatial recall questions (identifying physical locations/structures) and 5 information recall questions (understanding concepts/functions).
- **Subjective Questionnaires:** The System Usability Scale (SUS) [11] to assess usability; the NASA Task Load Index (NASA-TLX) [26] for cognitive and physical workload; and the User Engagement Scale - Short Form (UES-SF) [46] for engagement and focused attention.
- **Preference and Usefulness:** A comparative perceived usefulness rating on a 5-point scale, and a categorical preference selection (Guided, Unguided, or Both) collected at the conclusion of the session.

6.1.2 Quantitative Results. Due to the ordinal nature of our subjective scales and violations of normality indicated by Shapiro-Wilk tests, we utilized non-parametric statistical methods for our primary condition comparisons.

As detailed in Table 1, Wilcoxon Signed-Rank tests revealed no significant differences between the two conditions for Information Recall ($W = 29.00, p = 0.417, r = 0.144$) or Spatial Recall ($W = 13.50, p = 0.130, r = 0.267$). Similarly, we found no significant differences between the two conditions across the System Usability Scale (SUS), User Engagement Scale (UES-SF), and NASA-TLX workload assessments (all $p > .05$, with r ranging from 0.000 to 0.266).

Regarding categorical preference, 10 participants chose the Guided mode, 5 preferred the Unguided mode, and 1 selected both. An exact binomial test excluding the neutral response indicated this split was not statistically significant ($p = 0.302, g = 0.167$). Subjective ratings for perceived usefulness were also not significantly different, with the Guided condition scoring $M = 4.69$ ($SD = 0.48$) and the Unguided condition scoring $M = 4.44$ ($SD = 0.63$) ($W = 46.50, p = 0.247, r = 0.205$).

For spatial knowledge recall, a Wilcoxon Signed-Rank test indicated a significant practice effect; participants scored significantly higher in their second trial ($M = 97.50\%, SD = 6.83\%$) compared to their first ($M = 82.50\%, SD = 19.15\%$) ($W = 21.00, p = 0.013, r = 0.439$). In Trial 1 ($n = 8$ per group), no statistically significant difference was observed between the Guided ($M = 90.00\%, SD = 10.69\%$) and Unguided ($M = 75.00\%, SD = 23.30\%$) conditions ($U = 16.00, p = 0.183, r = 0.333$). No other significant order effects were observed across the remaining metrics.

Finally, a paired t-test confirmed no difference in Spatial Knowledge between the Haechi and Myeonbok modules ($M = 90.00, SD = 16.33$ for both; $t = 0.000, p = 1.000, d = 0.000$). However, Myeonbok yielded slightly higher Conceptual Knowledge scores ($M = 96.25, SD = 8.06$) than Haechi ($M = 82.50, SD = 19.15; t = -2.300, p = 0.036, d = 0.575$).

6.1.3 *Qualitative Insights.*

Post-study interviews revealed three primary themes regarding the learner experience.

Guided vs. Unguided Trade-offs: Interview responses detailed behavioral differences between the modes. Eight participants described the Guided system as a helpful scaffold that clearly communicated the author's educational objectives (P2: "...I could immediately tell, it's talking about this."), describing it as a helpful scaffold. Eleven users valued its convenience, with three highlighting the utility of the automated "view reset" during slide changes. Five participants emphasized that this automated guidance is beneficial for novices unfamiliar with 3D navigation. However, two participants reported that this automated realignment caused visual fatigue when it conflicted with their preferred viewing posture. For the Unguided mode, three users appreciated the freedom to explore on their own, and four participants reported that the lack of guidance promoted active learning (P8: "...I had to touch it, confirm 'it's really here,'... I had to interact more."). Conversely, six participants flagged the manual wayfinding and searching as a cognitive and physical burden.

Exploration Behaviors in Guided Mode: Ten participants reported engaging in free exploration while the guide was active. One user (P3) described the ability to override pre-authored camera views as an "essential" feature. Six participants explained that they broke away from the guide to investigate structural details not visible in the author's frame. Additionally, one participant used this feature to reposition the object and clear their field of view for physical comfort (P9: "Moving it so it wouldn't block my view").

Perceived Engagement and Comprehension: Regarding the overall immersive viewing experience across both conditions, participants highlighted how hands-on physical model manipulation supported their learning process. Eight users noted that this spatial setup promoted active learning, and eight participants linked this physical interaction to improved comprehension. Seven users noted that the immersive nature of the setup made studying captivating (P11: "Even if I didn't feel like studying, it made me study."). Two participants stated that physically interacting with the models helped them better remember the material.

6.2 Phase 2: Instructor Workflow (RQ2)

6.2.1 *Task, Materials, and Procedure.*

Following the learning phase, participants were introduced to the authoring interface. To evaluate the authoring workflow, participants assumed the role of a presenter tasked with creating a new spatial presentation. They were provided a separate 3-slide handout featuring a Buddha sculpture (Appendix). Unlike the learning modules in Phase 1, these slides contained only raw text, with no pre-authored spatial views. Participants were instructed to read the text descriptions and capture and save a 3D view that best visualized the specific details described (e.g., zooming in on a specific hand gesture). Upon completing the authoring task, they filled out the authoring-specific questionnaires (SUS, meCUE [40]). The session then concluded with the semi-structured interview.

6.2.2 *Quantitative Results.*

The mean SUS score was 82.19 ($SD = 7.58, Mdn = 81.25$), which was significantly higher than the acceptability benchmark of 68($W = 136.0, p < .001, r = 1.0$) [5]. On the 7-point Likert scale of the meCUE questionnaire, the mean ratings for usability and perceived usefulness were 5.39 ($SD = 1.23$) and 5.90 ($SD = 0.74$), respectively. Both meCUE dimensions were rated significantly above the neutral midpoint of 4 (Usability: $W = 100.5$, $p = .001$, $r = 0.91$; Usefulness: $W = 136.0$, $p < .001$, $r = 1.0$).

6.2.3 *Qualitative Insights.*

Post-study interviews revealed three primary themes regarding the workflow.

Authoring Concept: Eleven participants described the core view capture concept as clear and intuitive. While the majority rated the authoring process as easy, six participants noted an initial learning curve, describing the workflow as "difficult but learnable" for those unfamiliar with 3D environments. Additionally, three participants required a brief demonstration to fully grasp the linkage between the 3D spatial view and the 2D slide.

Workflow and Interaction Challenges: Participants identified specific bottlenecks regarding interaction fidelity, precision, and modality switching. The most prominent usability issue, cited by nine participants, was the unreliability of native hand-tracking for grasping and manipulating objects (e.g., the pinch gesture failing to release). To mitigate this, users suggested adding axis constraints for rotation (P11) or providing an option to use standard VR controllers for precise authoring (P8). Regarding spatial precision, three participants reported difficulty positioning the 3D model at the exact angle required for their intended view. Furthermore, users suggested features like automatic visual clipping (P2, P6) to handle parts of the model falling outside the intended authoring frame. Finally, one participant (P9) noted that constantly switching between spatial manipulation in XR and 2D slide editing on the desktop (via mouse and keyboard) felt cumbersome, suggesting a need for fully immersive, in-VR authoring controls.

Suggested Applications: When asked about the system's potential, participants frequently mentioned domains involving complex spatial structures. Suggested use cases prominently included product engineering and mechanical demonstrations (7 participants), museum and historical exhibits (6 participants), and scientific education and training simulations. For example, regarding mechanical structures, P11 noted: "...a firearm has internal mechanisms... when I operated it in 3D, I could understand it immediately." Participants also proposed commercial and spatial planning use cases. For instance, P15 suggested applying the system to interactive e-commerce showcases, where 3D models automatically highlight specific features corresponding to text descriptions. Additionally, P12 noted the system's utility in physical environments, stating that viewing augmented models alongside real-world surfaces (e.g., a physical desk) allowed them to verify the scale and placement of virtual objects.

7 Discussion, Limitations, and Future Work

While the quantitative metrics reflect comparable final performance, post-study interviews revealed a distinct trade-off. Participants reported that the Unguided mode promoted active learning by forcing them to manually search and verify spatial features, though several users flagged this constant manual searching as cumbersome or tedious. In contrast, users valued the Guided mode as a helpful scaffold that removed the overhead of visual searching, yet this automation sometimes caused friction when pre-authored views conflicted with a user's preferred physical posture. Rather than passively accepting the automated guide, the majority of participants broke away from the guide to investigate structural details independently, describing this freedom as an "essential" feature. These findings suggest that effective XR educational tools — even those that extend traditional linear study materials like ours — should not rely solely on rigid guided tours. Instead, instructional guidance must be fluid — supporting the learner through automated viewpoint resets while seamlessly permitting temporary, exploratory deviations.

Regarding authoring feasibility, the high system usability (SUS) score and positive meCUE ratings demonstrate that the concept of an integrated XR system was clear, though it may require a brief initial learning curve for novices. Beyond this baseline usability, participants recognized the tool's potential for applications like mechanical demonstrations, historical exhibits, and e-commerce. This suggests that the core interaction loop — pairing explicit text with an automated spatial viewpoint — can be useful in scenarios where one needs to take a step-by-step approach and use multiple modalities (e.g., text, images, 3D views) to explain 3D structures or concepts. Furthermore, viewing models alongside physical desks highlighted AR's unique pedagogical value. By leveraging real-world environmental anchors, learners intuitively grasped the physical scale and context of virtual artifacts. Consequently, these findings indicate that we should actively design for how virtual objects relate to both 2D presentation text and the surrounding physical environment.

Additionally, based on user feedback and observations, building upon the system goals outlined in Section 3, we describe three design takeaways to guide future hybrid presentation tools:

- **Dynamic Ergonomic Adaptation:** Participants frequently fought the automated camera alignment to manage visual occlusion and preserve physical comfort. Future systems must decouple spatial framing from ergonomic scale, incorporating dynamic auto-scaling or collision avoidance that respects the author's intended view while adapting to the learner's physical constraints and workspace setup [18, 35].
- **Enabling Spatial Precision:** Observations of participants struggling to capture precise viewpoints due to natural hand jitter indicate that free-form manipulation is insufficient for precise authoring tasks. Tools must complement macro-level spatial capture with micro-level constraints (e.g., axis snapping, rotational locks) to ensure reliable spatial formatting [2, 28].
- **Bidirectional Interaction Continuity:** Participants experienced difficulty switching between immersive hand tracking and the desktop mouse. To resolve this known Cross-Reality (CR) hurdle [17, 49], future tools should minimize modality switching. Systems should integrate emerging CR techniques, such as SpatialTouch [64], to enable bidirectional control (e.g., mid-air slide navigation) without breaking the user's operational context.

Our quantitative results revealed no significant differences in learning outcomes or workload between the Guided and Unguided conditions. High recall scores suggest a potential ceiling effect (see also Table 1), and the significant improvement observed in the second trial indicates a practice effect, both of which complicate the interpretation of these outcomes. Furthermore, 10 of the 16 users in the Guided condition actively engaged in free exploration, temporarily overriding the pre-authored views. This blending of interaction styles likely blurred the experimental boundaries, which may also explain the lack of statistical difference. Moreover, because Phase 2 lacked a baseline comparison to conventional workflows, our findings reflect perceived usability rather than objective reductions in coordination effort or authoring thresholds. Future studies should incorporate comparative baselines and objective metrics.

Additionally, our participant pool was gender-imbalanced and consisted primarily of university students. Although students possess high familiarity with PowerPoint, evaluating an instructor workflow with learner-profiled participants may not fully capture the practical needs of professional educators. Future work should validate this slide-driven workflow with domain experts in authentic curriculum development. Finally, our learning assessments were developed without formal pilot-testing, and the results revealed that the traditional clothing module (Myeonbok) yielded significantly higher conceptual recall scores than the sculpture module (Haechi). Future studies should employ formally validated instruments to ensure strict equivalence. From a technical perspective, the current prototype relies on the Windows-specific COM API to interface with PowerPoint, and storing metadata in Alt-Text compromises accessibility for screen-reader users. Future iterations will migrate to modern cross-platform APIs to eliminate COM polling latency, extend compatibility, and utilize dedicated, non-visible metadata storage.

8 Conclusion

SpatialSlides bridges traditional lecture materials and immersive 3D content. Rather than treating XR as an isolated activity or a disconnected post-classroom supplement, our integrated approach provides a seamless transition into contextualized 3D exploration. By anchoring virtual content to familiar classroom materials, the system guides users to contextualize and locate spatial information based on prior learning. Furthermore, effective information delivery is inherently multimodal, requiring a synthesis of text, 2D images, and spatial data. We facilitate this integration by leveraging presentation software — the dominant medium for educational text and 2D imagery — as the primary control interface. Because slide decks remain the ubiquitous standard for classroom instruction, handouts, and student review, tools like SpatialSlides ensure that the knowledge captured in these traditional formats can be translated into immersive environments.

User feedback from the exploratory study indicates that this integrated approach clarifies the instructor's intended messages while making the content easier to follow. However, delivering this guided experience requires a nuanced approach. Because users still highly value independent spatial exploration, integrated immersive educational tools should support fluid, seamless transitions between automated scaffolding and self-driven discovery as user needs shift.

Finally, we contribute the open-source framework of SpatialSlides as a reference for the community. By linking immersive clients directly with standard desktop applications, this work provides a practical blueprint for how everyday educational workflows can serve as a bridge to immersive spatial environments.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2026-25490785).

References

[1] Ahmed Alnagrat, Rizalafande Che Ismail, Syed Zulkarnain Syed Idrus, and Rawad Mansour Abdulhafith Alfaqi. 2022. A Review of Extended Reality (XR) Technologies in the Future of Human Education: Current Trend and Future Opportunity. *Journal of Human Centered Technology* 1, 2 (Aug. 2022), 81–96. doi:10.11113/humentech.v1n2.27

[2] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers Graphics* 37, 3 (2013), 121–136. doi:10.1016/j.cag.2012.12.003

[3] Narges Ashtari, Andrea Bunt, Joanna McGrenere, Michael Nebeling, and Parmit K. Chilana. 2020. Creating Augmented and Virtual Reality Applications: Current Practices, Challenges, and Opportunities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, 1–13. doi:10.1145/3313831.3376722

[4] Maribeth Back, Jonathan Cohen, Rich Gold, Steve Harrison, and Scott Minneman. 2001. Listen reader: an electronically augmented paper-based book. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. Association for Computing Machinery, 23–29. doi:10.1145/365024.365031

[5] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction* 24, 6 (July 2008), 574–594. doi:10.1080/10447310802205776

[6] Verena Biener, Travis Gesslein, Daniel Schneider, Felix Kawala, Alexander Otte, Per Ola Kristensson, Michel Pahud, Eyal Ofek, Cuauhtli Campos, Matjaž Kljun, Klen Čopič Pucihar, and Jens Grubert. 2022. PoVRPoint: Authoring Presentations in Mobile Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (May 2022), 2069–2079. doi:10.1109/TVCG.2022.3150474

[7] Verena Biener, Daniel Schneider, Travis Gesslein, Alexander Otte, Bastian Kuth, Per Ola Kristensson, Eyal Ofek, Michel Pahud, and Jens Grubert. 2020. Breaking the Screen: Interaction Across Touchscreen Boundaries in Virtual Reality for Mobile Knowledge Workers. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (Dec. 2020), 3490–3502. doi:10.1109/TVCG.2020.3023567

[8] M. Billinghurst, H. Kato, and I. Poupyrev. 2001. The MagicBook - moving seamlessly between reality and virtuality. *IEEE Computer Graphics and Applications* 21, 3 (May 2001), 6–8. doi:10.1109/38.920621

[9] Doug A. Bowman and Ryan P. McMahan. 2007. Virtual Reality: How Much Immersion Is Enough? *Computer* 40, 7 (July 2007), 36–43. doi:10.1109/MC.2007.257

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. doi:10.1191/1478088706qp063oa

[11] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[12] Ivy Bui, Arunabh Bhattacharya, Si Hui Wong, Harinder R. Singh, and Arpit Agarwal. 2021. Role of Three-Dimensional Visualization Modalities in Medical Education. *Frontiers in Pediatrics* 9 (Dec. 2021). doi:10.3389/fped.2021.760363

[13] Fabio Buttussi and Luca Chittaro. 2018. Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. *IEEE Transactions on Visualization and Computer Graphics* 24, 2 (Feb. 2018), 1063–1076. doi:10.1109/TVCG.2017.2653117

[14] Mengyu Chen, Marko Peljhan, and Misha Sra. 2021. EntangleVR: A Visual Programming Interface for Virtual Reality Interactive Scene Generation. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology (VRST '21)*. Association for Computing Machinery, 1–6. doi:10.1145/3489849.3489872

[15] Stephanie Claudino Daffara, Anna Brewer, Balasaravanan Thoravi Kumaravel, and Bjoern Hartmann. 2020. Living Paper: Authoring AR Narratives Across Digital and Tangible Media. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, 1–10. doi:10.1145/3334480.3383091

[16] Hugo Coelho, Pedro Monteiro, Guilherme Gonçalves, Miguel Melo, and Maximino Bessa. 2022. Authoring tools for virtual reality experiences: a systematic review. *Multimedia Tools and Applications* 81, 19 (Aug. 2022), 28037–28060. doi:10.1007/s11042-022-12829-9

[17] Robbe Cools, Inne Maerevoet, Matt Gottsacker, and Adalberto L. Simeone. 2025. Comparison of Cross-Reality Transition Techniques Between 3D and 2D Display Spaces in Desktop-AR Systems. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (may 2025), 2798–2805. doi:10.1109/TVCG.2025.3549907

[18] Shakiba Davari, Feiyu Lu, and Doug A. Bowman. 2020. Occlusion Management Techniques for Everyday Glanceable AR Interfaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 324–330. doi:10.1109/VRW50115.2020.00072

[19] Fabian Froehlich, Chris Hovey, Sameen Reza, and Jan L. Plass. 2024. Beyond Slideshows – Investigating the Impact of Immersive Virtual Reality on Science Learning. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 945–950. doi:10.1109/VR58804.2024.00113

[20] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. 2017. Working with Augmented Reality?: A Long-Term Analysis of In-Situ Instructions at the Assembly Workplace. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '17)*. Association for Computing Machinery, 222–229. doi:10.1145/3056540.3056548

[21] Travis Gesslein, Verena Biener, Philipp Gagel, Daniel Schneider, Per Ola Kristensson, Eyal Ofek, Michel Pahud, and Jens Grubert. 2020. Pen-based Interaction with Spreadsheets in Mobile Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 361–373. doi:10.1109/ISMAR50242.2020.00063

[22] Vicki S. Gier and David S. Kreiner. 2009. Incorporating Active Learning with PowerPoint-Based Lectures Using Content-Based Questions. *Teaching of Psychology* 36, 2 (April 2009), 134–139. doi:10.1080/00986280902739792

[23] Matt Gottsacker, Mengyu Chen, David Saffo, Feiyu Lu, Benjamin Lee, and Blair MacIntyre. 2025. Examining the Effects of Immersive and Non-Immersive Presenter Modalities on Engagement and Social Interaction in Co-located Augmented Presentations. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3706598.3713346

[24] Raphaël Grasset, Andreas Dünser, and Mark Billinghurst. 2008. Edutainment with a mixed reality book: a visually augmented illustrative childrens' book. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology (ACE '08)*. Association for Computing Machinery, 292–295. doi:10.1145/1501750.1501819

[25] Shilpi Gupta and Christopher Jaynes. 2006. The universal media book: tracking and augmenting moving surfaces with projected information. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*. 177–180. doi:10.1109/ISMAR.2006.297811

[26] 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.

[27] Christian Hartmann, Younes Orli-Idrissi, Laura Claudia Johanna Pflieger, and Maria Bannert. 2023. Imagine & immerse yourself: Does visuospatial imagery moderate learning in virtual reality? *Computers & Education* 207 (Dec. 2023), 104909. doi:10.1016/j.compedu.2023.104909

[28] Joseph J. LaViola Jr, Ernst Kruijff, Ryan P. McMahan, Doug Bowman, and Ivan P. Poupyrev. 2017. *3D User Interfaces: Theory and Practice.* Addison-Wesley Professional. Google-Books-ID: fxWSDgAAQBAJ.

[29] Veronika Krauß, Alexander Boden, Leif Oppermann, and René Reiners. 2021. Current Practices, Challenges, and Design Implications for Collaborative AR/VR Application Development. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, 1–15. doi:10.1145/3411764.3445335

[30] Jule M. Krüger and Daniel Bodemer. 2022. Application and Investigation of Multimedia Design Principles in Augmented Reality Learning Environments. *Information* 13, 2 (Feb. 2022), 74. doi:10.3390/info13020074

[31] Dina Kurzweil, Karen Marcellas, Brandon Henry, and Eric Meyer. 2020. Evidence-Based Guidelines for Recording Slide-Based Lectures. *Medical Science Educator* 30, 4 (Dec. 2020), 1611–1616. doi:10.1007/s40670-020-01032-w

[32] Bireswar Laha, Kriti Sensharma, James D. Schiffbauer, and Doug A. Bowman. 2012. Effects of Immersion on Visual Analysis of Volume Data. *IEEE Transactions on Visualization and Computer Graphics* 18, 4 (April 2012), 597–606. doi:10.1109/TVCG.2012.42

[33] Germán Leiva, Vittoria Frau, Unnikrishnan Radhakrishnan, and Mille Skovhus Lunding. 2025. Mucho: A Timeline-Based Immersive Tool for Prototyping Interactive Multimodal XR Experiences. *Proceedings of the ACM on Human-Computer Interaction* 9, 4 (June 2025), 1–20. doi:10.1145/3734185

[34] Zhen Li, Michelle Annett, Ken Hinckley, Karan Singh, and Daniel Wigdor. 2019. HoloDoc: Enabling Mixed Reality Workspaces that Harness Physical and Digital Content. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, 1–14. doi:10.1145/3290605.3300917

[35] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 147–160. doi:10.1145/3332165.3347945

[36] Blair MacIntyre, Maribeth Gandy, Steven Dow, and Jay David Bolter. 2005. DART: a toolkit for rapid design exploration of augmented reality experiences. In *ACM SIGGRAPH 2005 Papers (SIGGRAPH '05)*. Association for Computing Machinery, 932. doi:10.1145/1186822.1073288

[37] C. Marshall. 2005. Reading and Interactivity in the Digital Library: Creating an experience that transcends paper. In *Proceedings of the CLIR/Kanazawa Institute of Technology Roundtable*, Vol. 5. 1–20. https://api.semanticscholar.org/CorpusID:473212

[38] Fabrice Matulic and Daniel Vogel. 2023. Pen+Touch+Midair: Cross-Space Hybrid Bimanual Interaction on Horizontal Surfaces in Virtual Reality. In *Graphics Interface 2023*. https://openreview.net/forum?id=TPhs6nnLZUm

[39] Richard E. Mayer. 2020. *Multimedia Learning* (3 ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355

[40] Michael Minge, Manfred Thüring, Ingmar Wagner, and Carina V Kuhr. 2016. The meCUE questionnaire: a modular tool for measuring user experience. In *Advances in Ergonomics Modeling, Usability & Special Populations: Proceedings of the AHFE 2016 International Conference on Ergonomics Modeling, Usability & Special Populations, July 27-31, 2016, Walt Disney World®, Florida, USA*. Springer, 115–128.

[41] Kelly Minotti, Daniel Xuan Hien Mai, Guillaume Loup, Amine Chellali, Marie-Hélène Ferrer, and Samir Otmane. 2025. The Immersive Debriefing: Comparative Evaluation of Full and Segmented Redo Methods in Virtual Reality. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 109–119. doi:10.1109/ISMAR67309.2025.00024 ISSN: 2473-0726.

[42] Kyzyl Monteiro, Ritik Vatsal, Neil Chulpongsatorn, Aman Parnami, and Ryo Suzuki. 2023. Teachable Reality: Prototyping Tangible Augmented Reality with Everyday Objects by Leveraging Interactive Machine Teaching. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, 1–15. doi:10.1145/3544548.3581449

[43] Brad Myers, Scott E. Hudson, and Randy Pausch. 2000. Past, present, and future of user interface software tools. *ACM Transactions on Computer-Human Interaction* 7, 1 (March 2000), 3–28. doi:10.1145/344949.344959

[44] Brad A. Myers, Richard McDaniel, and David Wolber. 2000. Programming by example: intelligence in demonstrational interfaces. *Commun. ACM* 43, 3 (March 2000), 82–89. doi:10.1145/330534.330545

[45] Michael Nebeling and Maximilian Speicher. 2018. The Trouble with Augmented Reality/Virtual Reality Authoring Tools. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 333–337. doi:10.1109/ISMAR-Adjunct.2018.00098

[46] Heather L O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39.

[47] Jocelyn Parong and Richard E. Mayer. 2018. Learning science in immersive virtual reality. *Journal of Educational Psychology* 110, 6 (2018), 785–797. doi:10.1037/edu0000241

[48] Shwetha Rajaram and Michael Nebeling. 2022. Paper Trail: An Immersive Authoring System for Augmented Reality Instructional Experiences. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, 1–16. doi:10.1145/3491102.3517486

[49] Tobias Rau, Tobias Isenberg, Andreas Köhn, Michael Sedlmair, and Benjamin Lee. 2025. Traversing Dual Realities: Investigating Techniques for Transitioning 3D Objects between Desktop and Augmented Reality Environments. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3706598.3713949

[50] Tomoki "Issac" Saso, Kenji Iguchi, and Masa Inakage. 2003. Little red: storytelling in mixed reality. In *ACM SIGGRAPH 2003 Sketches & Applications (SIGGRAPH '03)*. Association for Computing Machinery, 1. doi:10.1145/965400.965573

[51] Philip Schuchardt and Doug A. Bowman. 2007. The benefits of immersion for spatial understanding of complex underground cave systems. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology (VRST '07)*. Association for Computing Machinery, 121–124. doi:10.1145/1315184.1315205

[52] Songjia Shen, Hsiang-Ting Chen, William Raffe, and Tuck Wah Leong. 2021. Effects of Level of Immersion on Virtual Training Transfer of Bimanual Assembly Tasks. *Frontiers in Virtual Reality* 2 (May 2021). doi:10.3389/frvir.2021.597487

[53] Songjia Shen, Chek Tien Tan, Hsiang-Ting Chen, William L Raffe, and Tuck Wah Leong. 2025. Educator Perceptions of XRAuthor: An Accessible Tool for Authoring Learning Content with Different Immersion Levels. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, 1–11. doi:10.1145/3706598.3713982

[54] Ajith Sowndararajan, Rongrong Wang, and Doug A. Bowman. 2008. Quantifying the benefits of immersion for procedural training. In *Proceedings of the 2008 workshop on Immersive projection technologies/Emerging display technologiges (IPT/EDT '08)*. Association for Computing Machinery, 1–4. doi:10.1145/1394669.1394672

[55] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. 2019. TabletInVR: Exploring the Design Space for Using a Multi-Touch Tablet in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, 1–13. doi:10.1145/3290605.3300243

[56] Nobuko Taketa, Kenichi Hayashi, Hirokazu Kato, and Shogo Noshida. 2007. Virtual Pop-Up Book Based on Augmented Reality. In *Human Interface and the Management of Information. Interacting in Information Environments*. Springer, 475–484. doi:10.1007/978-3-540-73354-6_52

[57] Robert B. Trelease. 2016. From chalkboard, slides, and paper to e-learning: How computing technologies have transformed anatomical sciences education. *Anatomical Sciences Education* 9, 6 (2016), 583–602. doi:10.1002/ase.1620

[58] Silviu Vert and Diana Andone. 2019. Virtual Reality Authoring Tools for Educators. *ITM Web of Conferences* 29 (2019). doi:10.1051/itmconf/20192903008

[59] Nam-Dang Vo, Van-Vinh Thai, Nam Hoai Do, Viet-Tham Huynh, Anthony Tang, and Khanh-Duy Le. 2025. Enhancing Spatial Understanding in Mixed-Reality Presentations. In *Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology (VRST '25)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3756884.3766024

[60] Tianyi Wang, Xun Qian, Fengming He, Xiyun Hu, Yuanzhi Cao, and Karthik Ramani. 2021. GesturAR: An Authoring System for Creating Freehand Interactive Augmented Reality Applications. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, 552–567. doi:10.1145/3472749.3474769

[61] Xiyao Wang, David Rousseau, Lonni Besançon, Mickael Sereno, Mehdi Ammi, and Tobias Isenberg. 2020. Towards an Understanding of Augmented Reality Extensions for Existing 3D Data Analysis Tools. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, 1–13. doi:10.1145/3313831.3376657

[62] Lei Zhang and Steve Oney. 2019. Studying the Benefits and Challenges of Immersive Dataflow Programming. In *2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. 223–227. doi:10.1109/VLHCC.2019.8818856

[63] Lei Zhang and Steve Oney. 2020. FlowMatic: An Immersive Authoring Tool for Creating Interactive Scenes in Virtual Reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. Association for Computing Machinery, 342–353. doi:10.1145/3379337.3415824

[64] Lixiang Zhao, Tobias Isenberg, Fuqi Xie, Hai-Ning Liang, and Lingyun Yu. 2025. SpatialTouch: Exploring Spatial Data Visualizations in Cross-Reality. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (jan 2025), 897–907. doi:10.1109/TVCG.2024.3456368

[65] Qian Zhou, George Fitzmaurice, and Fraser Anderson. 2022. In-Depth Mouse: Integrating Desktop Mouse into Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, 1–17. doi:10.1145/3491102.3501884