

Article

A More Immersive Experience, Deeper Learning? - A Comparative Study of Multimodal Interaction Combinations and Cognitive Mechanisms in UE5 Digital Exhibition Halls

Xuzhe Sun^{1,*}

¹ Shenyang Aerospace University, Shenyang, China

* Correspondence: Xuzhe Sun, Shenyang Aerospace University, Shenyang, China

Abstract: Unreal Engine 5 (UE5) facilitates the creation of highly immersive digital exhibition halls. However, an overabundance of visual, auditory, and haptic cues might elevate cognitive load, thereby diminishing learning outcomes. By drawing upon Multiple Resource Theory and Cognitive Load Theory, this study posits that a greater number of modalities does not inherently equate to enhanced efficacy. Furthermore, it suggests that the combination of visual and haptic interactions, denoted as V+H, could potentially offer a more favorable equilibrium between immersion and processing demand. A two-stage experimental approach is employed to compare the conditions of V+A, V+H, and V+A+H. Subsequently, cognitive load, attention allocation, and task self-efficacy are examined as mediating mechanisms. This study provides design guidance for achieving efficient multimodal interaction within UE5 exhibitions.

Keywords: Unreal Engine 5; Digital Exhibition Hall; Multimodal Interaction; Cognitive Load; Multiple Resource Theory

1. Introduction

Unreal Engine 5, abbreviated as UE5, can achieve highly detailed reconstruction of digital exhibition halls by utilizing Nanite geometry and Lumen global illumination. However, when high visual fidelity is combined with spatial audio and interactive feedback, it gives rise to a significant issue: does stronger immersion enhance learning?, or could it overload limited cognitive resources?

Evidence from mixed-reality exhibitions shows that multisensory interaction can enhance experience but may also cause cognitive overload when multiple channels are presented simultaneously [1]. Research on the Palace Museum Digital Cultural Relics Database likewise reports limited immersive interaction and low information absorption despite high-quality digital content [2]. Thus, immersion and learning are not necessarily positively correlated.

The design goal is therefore not maximum sensory stimulation, but coordinated information delivery. Studies in aviation simulation and autonomous-vehicle interfaces similarly suggest that performance depends more on effective cross-modal coordination than on simply increasing the number of modalities [3,4].

Therefore, this study explores whether the cue combinations of V+A, V+H, and V+A+H result in varying degrees of visitor task proficiency. It also investigates the cognitive and psychological mechanisms through which these effects manifest.

2. Literature Review

2.1. Visitor Proficiency in Digital Exhibition Halls

Visitor proficiency in digital exhibition halls refers to the ability to complete exhibit exploration, obtain information accurately, and retain knowledge within a limited time. It is commonly assessed through task completion rate, information accuracy, and reaction

Received: 04 August 2026

Revised: 15 September 2026

Accepted: 27 September 2026

Published: 30 September 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

time. UE5-based virtual exhibitions can support user experience and learning, yet current digital heritage systems often remain interaction-poor and show differences across user groups [2]. This suggests a need to optimize multimodal interaction for both efficiency and comprehension.

2.2. Multimodal Interaction and Cognitive Load Theory

Cognitive Load Theory suggests that working memory has its limitations. Once the demands for information processing surpass its capacity, learning will decline [5]. Within UE5 exhibitions, visual content, voice guidance, and haptic feedback might vie for cognitive resources. Should they be presented simultaneously at a high intensity, they could potentially give rise to extraneous load which is not related to the learning task.

Multiple Resource Theory posits that attention is spread out across perceptual channels [6]. Cues that are well-coordinated can alleviate the pressure on a single channel. However, redundant or competing signals have the potential to impair performance. Research on MR exhibitions suggests that appropriately combined multisensory cues can enhance information acquisition while avoiding unnecessary redundancy [1], which supports the proposition that coordinated combinations might surpass simple modality stacking.

2.3. Attention Allocation and Task Self-Efficacy

Attention allocation refers to the ability to maintain and switch focus among multiple information sources. Task self-efficacy is about confidence in completing a task [7]. In digital exhibitions, clear and coordinated feedback can support attention, enhance confidence, and encourage active exploration. Related AI navigation research indicates that decreasing cognitive and psychological burdens, while also increasing trust, can enhance task performance [8], thus offering a foundation for testing these variables as mediators.

2.4. Research Hypotheses and Model

Based on the above theories, this study proposes the following hypotheses:

H1: Within UE5 digital exhibition halls, when compared to the visual plus auditory (V+A) combination and the trimodal (V+A+H) combination, the visual plus haptic (V+H) combination can notably improve visitors' task proficiency.

H2: A visitor whose cognitive load is relatively low tends to exhibit higher task proficiency.

H3: Visitors with stronger attention allocation abilities tend to have higher task proficiency.

H4: A visitor's task self-efficacy is stronger, and their task proficiency is higher.

The research model is presented in Figure 1: Multimodal prompt combinations, which include V+A, V+H, and V+A+H, exert influence on visitors' task proficiency via three distinct pathways. These pathways are cognitive load that has a negative impact, attention allocation that brings about a positive effect, as well as task self-efficacy that also results in a positive outcome.

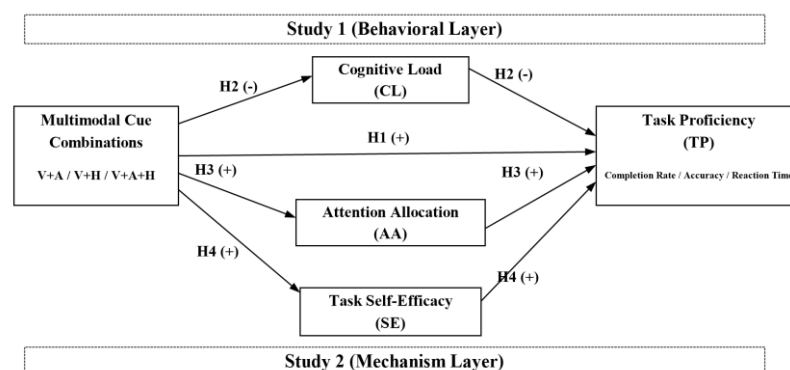


Figure 1. Research Model

3. Research Design

3.1. Study 1: Behavioral Performance Comparison

Study 1 adopts a three-group between-subjects design to conduct comparisons among V+A, V+H, and V+A+H. Through G*Power analysis, which involves 3 groups multiplied by 3 dependent variables, with f^2 set at 0.15, α at 0.05, and power at 0.80, it was determined that the minimum required sample size is 51. A total of seventy-five experienced volunteers were recruited for this study. Ultimately, there were 60 valid participants remaining after screening, with each group consisting of 20 individuals. They each finished an exploration task within a standardized UE5 exhibition that comprises eight exhibit nodes covering history, art, and technology.

The independent variable is cue combination. Visual cues use UI highlighting and dynamic lighting; auditory cues use spatialized voice guidance; haptic cues use controller vibration. Outcomes are task completion rate, information accuracy, and reaction time from cue onset to correct interaction (As shown in Table 1).

Table 1. Operationalized Measurement Indicators for Task Proficiency

Measurement Indicator	Operational Definition	Calculation/Recording Method	Indicator Direction
Task Completion Rate	visitors' completion degree of exploration tasks at 8 key exhibit nodes	Number of completed nodes $\div 8$	Higher is better
Information Accuracy	The extent to which visitors can accurately obtain and retain information about exhibits.	Correctness rate during subsequent knowledge tests	Higher is better
Reaction Time	visitors' time from prompt appearance to triggering correct interaction	System-recorded response time	Lower is better

3.2. Study 2: Psychological Mechanism Validation

Study 2 focuses on the V+H condition with 60 valid participants. Structural Equation Modeling (SEM) tests the mediating roles of cognitive load, attention allocation, and task self-efficacy. Cognitive load is measured with a simplified six-dimension NASA-TLX scale; attention allocation and task self-efficacy use three-item scales. All subjective measures use a 5-point Likert format (As shown in Table 2).

Table 2. Variables and Measurement Methods in Study 2

Variable	Type	Measure	Items	Data
Cognitive Load	Mediator	Short NASA-TLX	6 dimensions	5-point Likert
Attention Allocation	Mediator	Attention scale	3 items	5-point Likert
Task Self-Efficacy	Mediator	Self-efficacy scale	3 items	5-point Likert
Task Proficiency	Outcome	Behavioral indicators	Completion accuracy, reaction time.	Objective data

4. Expected Results and Discussion

4.1. Expected Findings

The V+H condition is anticipated to surpass V+A and V+A+H in terms of completion rate, accuracy, reaction time [6]. Haptic feedback can transmit information via a relatively

independent channel, thus diminishing competition for visual resources. Conversely, simultaneous three-channel input might elevate distraction and cognitive load, which aligns with the results obtained from MR exhibitions [1].

SEM is expected to show a negative effect of cognitive load on task proficiency ($\beta \approx -0.70$) and positive effects of attention allocation and task self-efficacy ($\beta \approx 0.50$) [3,8]. These variables are expected to mediate the relationship between cue combination and performance, indicating that effective multimodal design depends on matching information delivery to users' cognitive capacity (As shown in Table 3).

Table 3. The Correspondence That Exists between the Expected Outcomes and the Research Hypotheses.

Hypothesis/Path	Expected Result	Expected Explanation
H1: V+H vs. V+A / V+A+H	V+H demonstrates superior overall performance in completion rate, accuracy, and reaction time	Tactile feedback, functioning as a channel that is relatively independent, has the potential to decrease the resource competition occurring within a single channel.
H2: Cognitive Load $\rightarrow$ Task Proficiency.	Anticipated notable adverse influence	Higher cognitive load can lead to a greater likelihood of decline in task performance.
H3: Attention Allocation $\rightarrow$ Task Proficiency.	Expected significant positive impact.	Efficient allocation of attention is conducive to obtaining exhibition information and engaging in interactive responses.
H4: Task Self-Efficacy influences Task Proficiency.	Expected a significant positive impact.	Enhanced self-efficacy can foster active exploration, thereby improving task performance.

4.2. Theoretical Contributions

This study links Cognitive Load Theory and Multiple Resource Theory to UE5 exhibition design, emphasizing that modal synergy is more important than modal quantity [5,6]. It also extends the cognitive load–self-efficacy–task performance pathway to digital exhibitions and aligns with AR research showing that coordinated interaction and 3D presentation can improve engagement and learning [9].

4.3. Practical Implications

First, a tactile-first strategy can strengthen controller vibration at key interaction nodes and partially replace audio prompts, reducing competition between auditory and visual information [2].

Second, cognitive-load-sensitive design can use UE5 real-time rendering to adjust information density when browsing speed or interaction errors rise, helping maintain a balance between immersion and processing demand [1].

Third, clear and continuous haptic feedback can strengthen task self-efficacy, thereby maintaining visitors' willingness to explore [7] (As shown in Table 4).

Table 4. UE5 Digital Exhibition Hall Multimodal Interaction Design Strategies

Design Strategy	Issue	UE5 Implementation	Expected Function
Tactile-First	Audio-visual overload	Make use of vibrations at important nodes, decrease audio prompts.	Lower channel competition
Load-Sensitive Design	Excessive information density	Modify the density according to browsing speed, error rate.	Balance immersion with learning.

Efficacy Incentive	Low sense of control	Employ distinct continuous haptic or reward feedback.	Raise self-efficacy, engage in exploration.
-----------------------	-------------------------	--	---

4.4. Limitations and Future Outlook

The study is limited by its predominantly young sample, standardized exhibition setting, and reliance on self-report measures. Future work should include older and younger users, test more complex contexts, and combine behavioral data with eye tracking, EEG, EDA, sEMG, or HRV to support real-time cognitive-load monitoring and adaptive interaction [1,2].

5. Conclusion

UE5 expands the immersive potential of digital exhibitions, but more immersion does not necessarily produce deeper learning. Across simulation, vehicle HMI, MR exhibitions, and digital heritage research, coordinated multimodal design is more important than sensory accumulation. UE5 exhibition interfaces should therefore prioritize balanced cue combinations, manageable cognitive load, and feedback that supports attention and self-efficacy, so that immersion serves rather than competes with learning.

References

1. "Effects of immersion level and sound type on user experience in mixed reality interactive exhibitions: Evidence from psychophysiology," *Displays*, Elsevier, 2026.
2. "Multimodal interaction enhancement of digital cultural heritage system: User behavior analysis and interface reconstruction of the heritage scanning library of the palace museum," *Scientific Reports*, Springer Nature, vol. 16, p. 10654, 2026.
3. L. Bu, W. Zhang, Y. Liu, S. Qiao, and H. Li, "Better cues for better pilots: Comparing multimodal combinations and exploring cognitive mechanisms in general aviation simulation," *International Journal of Human-Computer Interaction*, p. 124, 2026.
4. L. Bu, B. Cui, W. Pan, H. Chen, S. Xia, and H. Li, "User-centered multimodal interaction design for autonomous vehicles: A focus on cognitive load and accessibility for users with severe visual impairment," *Universal Access in the Information Society*, vol. 24, no. 4, pp. 3431–3448, 2025.
5. J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988.
6. C. D. Wickens, "Multiple resources and mental workload," *Human Factors*, vol. 50, no. 3, pp. 449–455, 2008.
7. A. Bandura, *Self-efficacy: The exercise of control*. Macmillan, 1997.
8. H. Li, Y. Hou, W. Pan, H. Chen, J. Gai, and L. Bu, "Exploring the impact of anthropomorphism in AI navigation systems on elderly users' efficiency, trust, satisfaction, and cognitive load," *International Journal of Human-Computer Interaction*, vol. 42, no. 3, pp. 1503–1522, 2026.
9. M.-F. Lee, G.-S. Chen, H.-C. Chen, and J.-Z. Chen, "The intersection of digital wellbeing and collection exhibition: A study on the impact of AR interactive display models on visitor experience," *Computer Science and Information Systems*, vol. 22, no. 4, p. 1637, 2025.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.