Paper vs. Mixed Reality: a comparative study on Knowledge Acquisition, engagement and user comfort in non-STEM contexts

Edoardo Cecchini, Giuseppe Di Stefano

Abstract—This study investigates whether Mixed Reality (MR) can support deep knowledge acquisition in non-STEM academic contexts when integrated with traditional paper-based reading. While XR research has predominantly focused on STEM disciplines, we examine its potential for subjects requiring complex spatial imagination rather than symbolic manipulation.
Using the Elizabethan Globe Theatre as a case study, we compared traditional paper reading with an MR condition in which a 3D architectural model was spatially anchored near the physical text using a Meta Quest 3 headset. The experiment evaluated three constructs: Knowledge Acquisition (KA), Perceived Self-Efficacy (PSE), and Perceived Cognitive Load (PCL), with KA assessed across declarative, procedural/strategic, and a deliberately introduced declarative–spatial knowledge level.
Results indicate no advantage of MR for declarative or procedural–strategic knowledge, but a promising trend for declarative spatial knowledge and a clear increase in voluntary engagement, behaviorally reflected in longer reading durations. However, MR participants reported higher physical demand and lower perceived self-efficacy, likely due to medium novelty and current pass-through display limitations. Overall, the findings suggest that MR can enhance engagement and spatial understanding in humanities learning, while current hardware constraints limit its sustainability for reading-intensive tasks.

Keywords—*Augmented Reality, Reading Comprehension, Humanities Education, Knowledge Acquisition, NASA-TLX, Meta Quest 3*

1. Introduction

Traditional paper-based reading offers well documented cognitive benefits: it provides a distraction-free environment, stable visual presentation, and facilitates deep comprehension [2]. However, printed materials are inherently limited in their ability to represent complex spatial relationships and abstract concepts that remain "invisible" to readers lacking strong visualization skills. Digital reading can partially address these limitations through multimedia integration, yet research shows that digital interfaces promote shallow processing and skimming behaviors, ultimately hindering students' ability to grasp complex conceptual relationships [3]. Our study seeks to synthesize the cognitive depth afforded by physical reading with the immersive potential of Extended Reality (XR). While the majority of XR research concentrates on STEM disciplines, we explore its potential in non-STEM contexts that demand complex spatial imagination. Following established frameworks in MR education research, Mixed Reality (MR) technologies yield educational benefits primarily when applied to learning abstract or spatially complex concepts requiring three-dimensional visualization [5] where traditional 2D representations prove inadequate. We selected the Elizabethan Globe Theatre as our subject matter, which integrate a sophisticated vertical cosmology stratified from "Hell" beneath the stage to the "Heavens" above, with thrust stage geometry requiring considerable mental reconstruction skills that XR can effectively support. Recognizing that successful XR implementation depends critically on interaction design, we deliberately minimized interface complexity to preserve deep-focus benefits. Our design strategy addresses the distinction between L1 (Interface) cognitive load, the extraneous mental effort required to master the technology itself, and L2 (Domain) cognitive load, the desirable cognitive effort dedicated to processing the subject [7]. Participants in the MR condition received printed text and a 3D model of the theatre with labeled sections, anchored above a QR code marker. Interaction was limited to physical manipulation of the marker and natural head movements, intentionally avoiding graphical user interfaces or complex gestural controls to minimize L1 cognitive load and maximize L2 learning engagement. This research extends beyond measuring declarative knowledge to comprehensively assess multiple levels of knowledge acquisition following Shavelson et al.'s conceptual framework[1]. We examine declarative spatial knowledge (architectural terminology and spatial relationships) as well as procedural and strategic knowledge, the ability to apply spatial understanding to solve complex problems integrating multiple architectural elements and environmental constraints, such as optimizing actor positioning for audience sightlines. In addition, we consider a spatially grounded form of declarative knowledge, concerning the recall of the three-dimensional placement and relative organization of architectural components. Finally, we measure Perceived Self-Efficacy (PSE) to assess whether MR model support enhanced learners' confidence in their knowledge mastery, and Perceived Cognitive Load (PCL), a fundamental consideration when employing head-mounted displays in educational contexts.

2. Scientific question

The research question that this study tries to answer is: How do traditional paper-based learning environments and mixed-reality environments applied to paper reading differ in their impact on university students' Perceived Cognitive Load (PCL), Perceived Self-Efficacy (PSE), and Knowledge Acquisition (KA) in a non-STEM educational context?

Independent Variable: Learning environment (MR-augmented reading vs. traditional paper reading)

Dependent Variables:

- Knowledge Acquisition (KA): assessed across declarative, declarative-spatial, and procedural-strategic knowledge levels
- Perceived Self-Efficacy (PSE): learner confidence in knowledge mastery
- Perceived Cognitive Load (PCL): overall workload felt during the task

3. Related Works

One of the primary challenges in XR-based learning is managing the learner's limited working memory. The L1–L2 cognitive load model [7] suggests that effective educational design must prioritize L2 domain-specific learning by minimizing the cognitive effort required for L1 interface mastery. Introducing unnecessary technological complexity next to inherently demanding material can create a dysfunctional experience that decrease student motivation and undermines knowledge transfer. This theoretical friction directly justifies our minimal interaction design strategy and is the reason behind our investigation of Perceived Cognitive Load (PCL) through questionnaires. XR educational research has concentrated predominantly on STEM disciplines because empirical evidence demonstrates these technologies prove effective primarily for "invisible" and spatially complex concepts [5] where three-dimensional visualization or interaction is crucial for comprehension. Research indicates that when immersive tools are applied to subjects not requiring such visualization, they may provide negligible benefits or even prove counterproductive as a distraction. This principle directly guided our selection of the Elizabethan Globe Theatre; it requires sophisticated spatial reasoning and visualization skill that traditional 2D media is often inadequate to support. Our study thus addresses a notable gap where architectural and humanities applications remain significantly underexplored compared to science and engineering. Grounding the evaluation in Shavelson et al.'s (2008) framework [1], this study initially identifies four levels of Knowledge Acquisition: declarative (factual recall), procedural (steps for tasks), schematic (connecting concepts), and strategic (application in new contexts). However, we adapted this framework for our specific case study by excluding the schematic level, as it was considered less applicable and difficult to test effectively within the scope of our reading-based experiment. We also chose to consider procedural and strategic knowledge as a single, integrated category to better evaluate the application of architectural and environmental constraints. To further enrich the original framework, we introduced a novel "declarative-spatial" level, which aims to isolate the recall of three-dimensional placement and the relative organization of components. This extension was done because literature reviews indicate that while XR technologies are increasingly studied, approximately 50% of research prioritizes usability over learning outcomes, with those that do measure Knowledge Acquisition often focusing only on simple declarative knowledge [6]. Since research suggests that basic factual recall is the easiest level to achieve and often neglects the deeper conceptual integration found in procedural and strategic knowledge, our assessment was deliberately designed to include these higher-order questions alongside our proposed spatial metrics to capture a more comprehensive picture of learning in non-STEM contexts. The Cognitive-Affective Model of Immersive Learning (CAMIL) [4] identifies presence and agency as the primary affordances driving learning outcomes.

Our design's core strength lies in maximizing presence by preserving the natural reading context; utilizing physical paper provides stable haptic landmarks that minimize cognitive strain and maintain the deep analytical focus traditionally associated with print. Regarding agency, resource constraints necessitated a simplified implementation via a physical QR-coded cube; however, this intentional control is hypothesized to act as an 'affective buffer,' particularly as the unique ability to simultaneously 'read and see' situated 3D visualizations is expected to boost Perceived Self-Efficacy (PSE) by demystifying the Globe Theatre's complex architectural structure.

4. Methodology & Experimental Design

4.1. Overview

This section outlines the methodological structure of the experiments conducted for this research. To evaluate the impact of immersive technology on learning, a comparative study was designed, consisting of two distinct experimental conditions. The Mixed Reality (MR) Group, involves the execution of the reading task through a Head-Mounted Display (HMD), where digital 3D architectural models are integrated into the physical workspace. The second condition, the Control Group, requires participants to perform the same reading task using traditional media, specifically, printed text and static 2D images. By mirroring the informational content across both conditions, the study aims to isolate the effects of spatial visualization provided by the MR environment on Knowledge Acquisition (KA), Perceived Self-Efficacy (PSE), and Perceived Cognitive Load (PCL).

4.2. Participants

The study sample consists of 14 university students aged 18 to 28, representing a cross-section of STEM, psychology, and humanities disciplines. This diverse selection is strategic, with the aim of addressing the research gap regarding XR applications in non-STEM academic contexts. To maintain consistency and ensure the validity of the results, all participants were non-native English speakers with an intermediate proficiency level, ranging from B1 to B2. Furthermore, the recruitment process prioritized students who habitually use physical textbooks for their academic preparation, ensuring that the control group's experience accurately reflects a standard learning environment.

4.3. Materials and Apparatus

The experiment utilized a combination of standardized learning materials and condition-specific apparatus.

Shared Materials (Both Conditions):

- **Printed text:** *"The Architecture of Spectacle: Spatial Dynamics of the Elizabethan Globe"* (approx. 1,200 words, B1 English level)

Figure 1. MR condition reading setup

- **Knowledge assessment:** 10-question multiple-choice test (4 options per question), targeting declarative and procedural/strategic knowledge levels
- **PSE questionnaire:** Single question with 5-point Likert scale item
- **NASA-TLX questionnaire:** 6 subscales with open-ended section
- Stopwatch for time-on-task measurement

MR condition specific apparatus:

- Meta Quest 3 head-mounted display with passthrough mixed reality capability
- Physical QR-coded cube for spatial anchoring
- Custom MR application displaying explorable 3D model of Globe Theatre with labeled architectural elements (Pit, Galleries, Hell, Heavens, Upper Stage, etc.)

Control condition specific materials:

- 2D static image of Globe Theatre structure (printed, same architectural elements visible as 3D model)

4.4. Experimental Procedure

Both conditions followed an identical protocol with only the visualization medium differing:

Phase 1 - Setup and familiarization:

- **MR condition:** Participants wore the Meta Quest 3 and completed a brief adaptation phase, anchoring the 3D Globe Theatre model to the QR cube positioned ergonomically on their desk. This was made to mitigate the "novelty effect", where early excitement or technical confusion can inflate cognitive load.
- **Control condition:** Participants were seated with printed materials (text + 2D image) arranged on the desk.

Phase 2 - Instructional briefing: Both groups received identical instructions via an initial survey, informing them that their reading would be followed by a knowledge assessment.

Phase 3 - Reading task: Participants read the article at their own pace. Time-on-task was discretely recorded using a stopwatch to measure engagement (ecological validity preserved by making the participants unaware of timing).

- **MR group:** Could physically manipulate the QR cube and use natural head movements to examine the 3D model from multiple angles while reading.
- **Control group:** Referenced the static 2D image alongside the text.

Phase 4 - Assessment sequence:
Immediately following the reading task, participants completed:

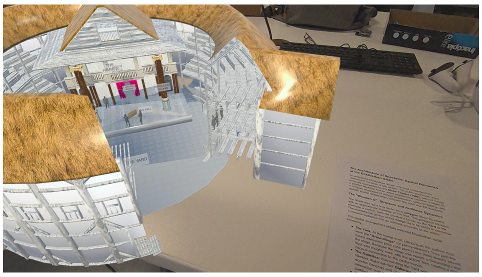

Figure 2. MR application view

1. PSE self-rating (1-5 Likert scale): "How confident are you in your understanding of the text you read?"
2. Knowledge acquisition test (10 questions)
3. NASA-TLX questionnaire supplemented with open-ended prompts to capture the rationale behind the workload ratings
4. Demographic survey to assess prior experience with augmented reality and provide a foundational knowledge assessment, including participants' English proficiency levels and their existing familiarity with the subject studied

4.5. Measurement Instruments

To capture complementary dimensions of the learning experience, the study employs a set of measurement instruments aligned with the three core dependent variables: Knowledge Acquisition (KA), Perceived Self-Efficacy (PSE), and Perceived Cognitive Load (PCL). This combination was deliberately selected to address a limitation observed in prior XR learning research, which often prioritizes usability metrics or simple factual recall while underrepresenting deeper forms of learning and cognitive sustainability. In addition, a behavioral measure (Time on Task) was included as a secondary indicator of engagement. Together, these instruments provide a structured and multi-layered evaluation of learning outcomes and cognitive experience.

4.5.1. Knowledge Acquisition (KA)

Knowledge Acquisition was assessed through a 10-item multiple-choice test, with four response options per item. The test was designed, as discussed before, with the explicit goal of evaluating not only performance differences between conditions, but also qualitative differences in the type of knowledge acquired. To this end, questions were distributed across different knowledge categories:

- **Declarative Knowledge (4 questions):** These questions target factual recall and terminology related to the architectural structure of the Globe Theatre. They primarily assess basic memorization and do not require complex spatial reasoning or three-dimensional visualization.
- **Procedural and Strategic Knowledge (5 questions):** This category includes questions that require the integration of multiple spatial elements of the theater and the application of architectural and environmental constraints. Rather than testing isolated factual knowledge, these questions assess whether participants can reason about how space, structure, and audience placement interact during a performance. For example, participants are asked to determine optimal staging choices, such as selecting the best position for an actor to maximize visibility across different audience sections or identifying the most suitable area to convey specific narrative conditions.
- **Declarative Spatial Knowledge (1 question):** This category focuses on declarative knowledge that is explicitly spatial in nature, requiring participants to recall where specific architectural elements are located within the theater and how they are positioned relative to other components. While this category is not explicitly part of the original framework, it was intentionally included for motivations discussed in Section 3.

4.5.2. Perceived Self-Efficacy (PSE)

Single-item 5-point Likert scale administered immediately post-reading, pre-test: "How much do you feel prepared and confident in your ability to accurately answer questions about the structure and function of the Shakespearean theatre?"

4.5.3. Perceived Cognitive Load (PCL)

NASA-TLX questionnaire consisting of six subscales (mental demand, physical demand, temporal demand, perceived performance, effort, and frustration), with the option for participants to briefly explain their ratings.

4.5.4. Time on task

Reading duration (in minutes) recorded as a behavioral indicator of engagement, depth of involvement with the learning material, and willingness to take time for exploration. This measure is also analyzed in relation to the temporal demand subscale of the NASA-TLX to assess whether participants felt hurried or time-pressured while completing the task.

5. Results

This section presents the findings from the comparative study between Mixed Reality (MR) and traditional paper-based learning environments. Given the limited sample size ($N = 14$), we employed a non-parametric statistical approach. Statistical significance was determined using $p < 0.05$, while effect sizes (Cohen's d) were calculated to identify meaningful trends. Following established conventions, effect sizes of $|d| > 0.5$ indicate substantial practical differences, even when statistical significance is not achieved due to limited statistical power.

5.1. Hypotheses

Based on the theoretical framework and related literature, we formulated the following hypotheses, organized by their centrality to the research question:

- **H1 - Knowledge Acquisition:**
  - **H1.1 - Declarative Knowledge Acquisition:** Comparable performance is expected between the MR and control groups, as minimal advantage from spatial visualization is anticipated at this level.
  - **H1.2 - Procedural and Strategic Knowledge:** An advantage for the MR condition is expected for this category of questions. The availability of an interactive 3D visualization directly supports spatial reasoning by allowing users to perceive depth, relative positioning, and visibility constraints more intuitively. In contrast, these relationships are significantly more difficult to infer from static, non-interactive 2D diagrams.
  - **H1.3 - Declarative Spatial Knowledge:** An advantage for the MR condition is expected for this category. The ability to directly observe the three-dimensional placement of architectural components while accessing descriptive information is hypothesized to support stronger spatial encoding within this introduced declarative–spatial distinction.
- **H2 - Perceived Self-Efficacy:** MR will enhance PSE, as the ability to simultaneously read and visualize 3D models demystifies complex architectural relationships, boosting learner confidence.
- **H3 - Perceived Cognitive Load:** Despite minimal interaction design, MR will show higher overall PCL, with physical demand being the primary contributor due to current HMD limitations.
- **H4 - Time on Task:** MR participants will spend more time engaged with the material, indicating enhanced intrinsic motivation and voluntary exploration.

5.2. Knowledge Acquisition (KA)

5.2.1. H1.1 - Declarative Knowledge

Both groups achieved identical performance on declarative knowledge questions, with a mean accuracy rate of 89% ($d = 0.00$). This fully supports H1.1, confirming that MR provides no measurable advantage for factual recall and terminology that do not require spatial reasoning. This aligns with existing XR literature showing that immersive technologies offer limited value when learning objectives lack spatial complexity.

5.2.2. H1.2 - Procedural and Strategic Knowledge

Procedural and strategic questions, requiring integration of multiple spatial elements and application of architectural constraints, proved most challenging across both conditions. The MR group achieved 54% accuracy compared to 51% for the paper-based group. This minimal difference was not statistically significant ($p > 0.05$) with a negligible effect size ($d = 0.16$), failing to support H1.2.

Although we hypothesized that interactive 3D visualization would support spatial reasoning and facilitate the transfer of learned knowledge to novel contexts, the results provide no evidence for this advantage. Several explanations are plausible. Prior literature consistently highlights that learning outcomes in XR environments are highly sensitive to interaction design, suggesting that the limited interactivity of our prototype may have constrained its educational effectiveness. Additionally, existing research indicates that such benefits are often more pronounced in STEM domains, where spatial manipulation is more tightly coupled with problem-solving demands. The very small observed effect size further suggests that this outcome cannot be attributed solely to limited sample size.

5.2.3. H1.3 - Declarative Spatial Knowledge

The most promising result emerged for declarative spatial knowledge, assessing recall of 3D placement and relative positioning of architectural elements. The MR group achieved 86% accuracy compared to 57% for the paper-based group, a 29 percentage point advantage.

While not statistically significant ($p = 0.290$), likely due to small sample size and single-question assessment, the medium-to-large effect size ($d = 0.62$) suggests a meaningful trend supporting H1.3. The ability to directly observe three-dimensional relationships while reading appears to facilitate stronger spatial encoding compared to static 2D representations. This finding indicates MR's potential for supporting spatial understanding in non-STEM contexts, though larger studies with multiple spatial items are needed for confirmation.

Figure 3 illustrates the comparative performance across all knowledge categories.

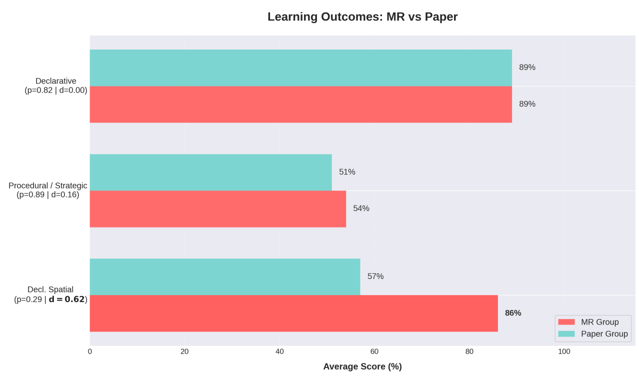

Figure 3. Knowledge Acquisition comparison between MR and paper-based conditions in the different categories

5.3. Perceived Self-Efficacy (PSE)

Statistical analysis revealed a significant difference ($p = 0.045$) contradicting H2. The paper-based group reported higher perceived self-efficacy ($M = 3.86$) compared to the MR group ($M = 3.29$), with a very large effect size ($d = 1.31$).

This unexpected finding contradicts our hypothesis grounded in the CAMIL framework. However, post-experiment demographic data revealed that all participants reported zero or minimal prior experience with head-mounted displays, suggesting that differences in confidence are likely attributable to familiarity with the learning medium itself. Participants engaged with paper-based reading through a methodology they have repeatedly relied upon throughout their academic careers, fostering comfort, predictability, and consequently higher self-confidence. In contrast, the MR condition introduced a novel learning environment that appears to have induced metacognitive uncertainty, reducing participants' confidence in their preparation despite the possibility of improved learning outcomes.

This difference is visualized in Figure 4.

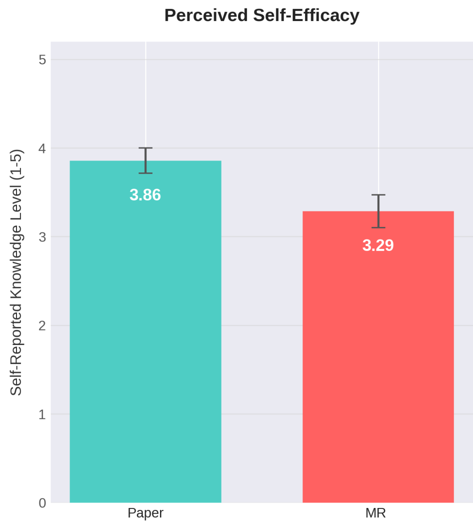

Figure 4. PSE ratings comparison between MR and paper-based learning conditions (5-point Likert scale)

5.4. Perceived Cognitive Load (PCL)

PCL was assessed using the NASA-TLX questionnaire, evaluating six workload dimensions. Our primary focus was physical demand, as hypothesized in H3.

5.4.1. Physical Demand

A statistically significant difference emerged in physical demand ($p = 0.019$) with a very large effect size ($d = 1.65$), strongly supporting H3. MR participants reported markedly higher physical discomfort ($M = 3.29$) compared to paper-based participants ($M = 1.43$) on the 5-point scale.

Qualitative feedback from the open-ended NASA-TLX section provided crucial insights. A substantial proportion of MR participants reported eye strain and difficulty achieving comfortable focus:

- *"It was a little out of focus and hard to read"*
- *"It was demanding principally because of the fact that it was kind of blurred"*

This reflects a fundamental limitation of current MR technology for reading tasks: text must be viewed through the headset's passthrough cameras and optical system, which provide lower visual quality and require unnatural focal accommodation compared to reading directly from paper. This represents a major barrier to comfortable extended use for reading-intensive educational tasks.

5.4.2. Frustration

Frustration ratings showed slight elevation in the MR condition but were not statistically significant with a small effect size ($d < 0.5$). Despite physical discomfort, participants did not experience substantial frustration, suggesting they recognized the potential value of spatial visualization despite hardware limitations.

The distribution of cognitive load across all NASA-TLX dimensions is shown in Figure 5.

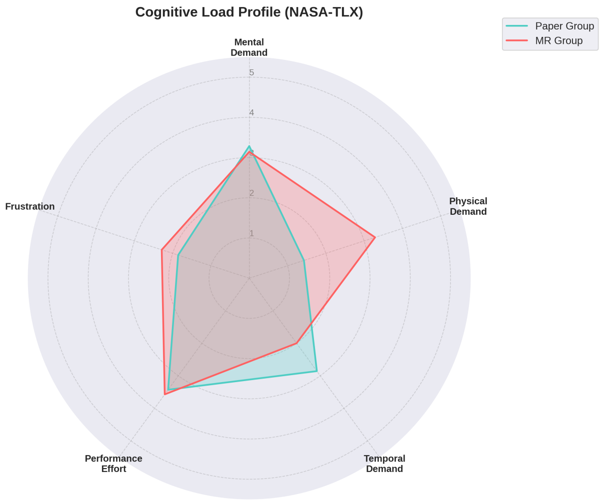

Figure 5. NASA-TLX subscale ratings across six dimensions for MR and paper-based conditions

5.5. Time on Task

Time on task was recorded discreetly to measure behavioral engagement. Results strongly support H4, revealing a distinctive pattern indicative of voluntary exploration and intrinsic motivation.

5.5.1. Reading Duration

MR participants spent an average of 2.5 minutes longer on the reading task ($M = 10.92$ min) compared to paper-based participants ($M = 8.63$ min). While not statistically significant ($p = 0.197$), the large effect size ($d = 0.81$) indicates a substantial trend toward increased time investment with augmented materials.

5.5.2. Temporal Demand

Interestingly, the NASA-TLX Temporal Demand subscale (Fig. 5) revealed that MR participants perceived lower time pressure ($M = 2.00$) compared to paper-based participants ($M = 2.86$). This difference approached significance ($p = 0.084$) with a very large effect size ($d = 1.13$).

5.5.3. Interpretation

The convergence of these metrics provides compelling evidence for H4. Participants spent more time while feeling less rushed.

If additional time resulted from difficulty or interface frustration, temporal demand would have been elevated, reflecting urgency to finish. Instead, the opposite pattern emerged: participants chose to engage more deeply with the material, freely exploring the 3D model and its relationship to the text without feeling pressured. This behavioral signature, extended voluntary engagement combined with reduced time pressure, indicates intrinsic motivation and genuine interest in the learning material, consistent with theories of effective learning environments.

5.6. Summary of Findings

Table 1 provides an overview of hypothesis testing results.

Key Findings:

- **Knowledge Acquisition:** MR showed no advantage for declarative knowledge (H1.1 supported), no meaningful effect for procedural/strategic knowledge (H1.2 not supported), but a promising trend for spatial knowledge (H1.3, $d = 0.62$).

Table 1. Summary of Hypothesis Testing Results

Hypothesis	Variable	Support
H1.1	Declarative KA	Supported
H1.2	Procedural/Strategic KA	Not Supported
H1.3	Spatial KA	Promising Trend
H2	Self-Efficacy	*Contradicted
H3	Physical Demand	*Supported
H4	Engagement	Supported (Trends)

* Statistically significant ($p < 0.05$); Trends indicated by $|d| > 0.5$

- **Self-Efficacy:** Paper-based reading produced significantly higher confidence (H2 contradicted, $d = 1.31$), likely due to medium familiarity rather than learning effectiveness. Universal lack of prior HMD experience suggests technological novelty suppressed metacognitive confidence.
- **Cognitive Load:** MR introduced significantly higher physical demand (H3 supported, $d = 1.65$), primarily from visual discomfort and focal issues. Nevertheless, the slightly higher performance scores achieved by the MR group suggest that the minimal interaction design successfully avoided interface-related burden (L1 load).
- **Engagement:** MR prompted longer voluntary engagement ($d = 0.81$) with reduced time pressure ($d = 1.13$), creating a "quality time signature" indicative of intrinsic motivation (H4 supported).

6. Conclusion

This study examined the impact of Mixed Reality–augmented reading on knowledge acquisition, perceived self-efficacy, cognitive load, and engagement in a non-STEM educational context. The results indicate that MR does not provide measurable advantages for declarative or procedural–strategic knowledge acquisition under the present design, while showing a promising trend for declarative spatial knowledge and a clear increase in voluntary engagement. These findings highlight the importance of distinguishing between different types of knowledge when evaluating XR-based learning interventions, particularly beyond STEM domains.

In essence, the results suggest, consistent with prior literature on XR learning, that the effectiveness of MR is strongly mediated by interaction design, while also being influenced by learners' familiarity with the medium. While the minimal interaction strategy successfully reduced L1 (interface-related) cognitive load, enabling participants to focus on the learning task itself, current pass-through MR technology introduced substantial physical discomfort, primarily due to visual strain and focal accommodation. This limitation is particularly critical from a sustainability perspective, as real-world study sessions typically extend far beyond the short duration of an experimental task. Notably, and in contrast with common expectations in the literature, MR appears to have had a slight negative effect on perceptions of self-efficacy, despite increased engagement and exploratory behavior.

Future research should investigate whether these constraints can be mitigated through alternative hardware configurations, such as optical see-through MR displays that allow direct viewing of physical text, or through interaction designs that increase meaningful agency without reintroducing excessive cognitive load. Larger-scale studies are also needed to confirm the observed trends in spatial knowledge acquisition and engagement, and to better understand how XR technologies can support higher-order learning in humanities and other non-STEM contexts.

References

[1] A. Mangen, B. R. Walgermo, and K. Brønnick (2013) *Reading linear texts on paper versus computer screen: Effects on reading comprehension*, International Journal of Educational Research, vol. 58, pp. 61–68.

[2] L. M. Singer and P. A. Alexander (2017) *Reading across mediums: Effects of reading digital and print texts on comprehension and calibration*, The Journal of Experimental Education, vol. 85, no. 1, pp. 155–172.

[3] G. Makransky and G. B. Petersen (2021) *The Cognitive Affective Model of Immersive Learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality*, Educational Psychology Review, vol. 33, no. 3, pp. 937–958.

[4] R. J. Shavelson, D. B. Young, C. C. Ayala, et al. (2008) *On the Impact of Curriculum-Embedded Formative Assessment on Learning: A Collaboration Between Curriculum and Assessment Developers*, Applied Measurement in Education, vol. 21, no. 4, pp. 295–314.

[5] M. Chandramouli, A. Zafar, and A. Williams (2025) *A Detailed Review of the Design and Evaluation of XR Applications in STEM Education and Training*, Electronics, vol. 14, no. 19, 3818.

[6] P. Acevedo, A. J. Magana, B. Benes, and C. Mousas (2024) *A Systematic Review of Immersive Virtual Reality in STEM Education: Advantages and Disadvantages on Learning and User Experience*, IEEE Access, vol. 12, pp. 189359–189386.

[7] X. Li, Y. Han, Y. Wu, K. Yue, and Y. Liu (2024) *User study of an AR reading aid system to promote deep reading*, Computers & Education: X Reality, vol. 5, 100086.

[8] J. W. Lai and K. H. Cheong (2022) *Educational Opportunities and Challenges in Augmented Reality: Featuring Implementations in Physics Education*, IEEE Access, vol. 10, pp. 43143–43158.