Designing Gaze-Adaptive Immersive Learning Support

Zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Shi Liu, M. Sc.

Tag der mündlichen Prüfung	23.04.2026
Referent:	Prof. Dr. Alexander Mädche
Korreferent:	PD. Dr. Simone-Nadine Löffler

Karlsruhe, April 2026

Abstract

The field of immersive learning has emerged through the application of technologies such as Virtual Reality (VR) and Mixed Reality (MR) to create interactive educational experiences. In contrast to passive content consumption, typical for learning on 2D screens or with printed materials, immersive learning enables presence and active exploration of abstract concepts within situated environments. However, challenges remain in these systems: the interactive elements and high immersion that make immersive learning effective can also overwhelm learners, leading to distraction and cognitive overload. To address this, gaze-adaptive support has been recognized as an effective design intervention to facilitate cognitive processing. Yet, while modern head-mounted displays (HMDs) used for immersive learning are increasingly equipped with eye-trackers, there is a lack of knowledge on how to transfer technical capabilities and eye-tracking data into learner-centered designs that effectively support immersive learning. Therefore, this dissertation investigates three research gaps: i) a conceptualization of the state-of-the-art of mixed reality in higher education, ii) the design of gaze-adaptive support, and iii) the extension of gaze-adaptive support to ex-situ learning with cross-device interaction. The research is conducted through four studies. The first study establishes a conceptual framework through a systematic review of head-mounted MR systems in higher education. The second study implements a gaze-adaptive support design in immersive learning systems, introducing attention feedback to promote self-reflection. Recognizing that learning extends beyond the immersive session itself, the third study introduces a cross-device interaction technique that facilitates seamless note-taking between a headset and a tablet. Finally, the fourth study presents the learning ecosystem AttentiveLearn, comprising an immersive virtual classroom, an attention-aware personalization pipeline, and an integrated mobile assistant. This ecosystem utilizes eye-tracking data captured in VR to personalize post-lecture quizzes and was evaluated in a four-week field study. This dissertation contributes to the field of Human-Computer Interaction (HCI). First, the dissertation provides a conceptual framework for MR system design in higher education. Second, it offers empirical insights into improving learners' attention management, user experience, and learning outcomes through gaze-adaptive support. Lastly, it contributes a set of artifacts and technical pipelines for cross-device and gaze-adaptive support. Overall, these findings offer researchers and practitioners insights to create learner-centered immersive learning experiences.

Contents

List of Figures

List of Tables

List of Abbreviations

ACM	Association for Computing Machinery
ADI	Attention Distribution Index
AI	Artificial Intelligence
ANOVA	Analysis of variance
AOI	Area of Interest
API	Application Programming Interface
AR	Augmented Reality
BF	Bayes Factor
BJET	British Journal of Educational Technology
CAD	Computer-Aided Design
CAMIL	Cognitive Affective Model of Immersive Learning
CEFR	Common European Framework of Reference for Languages
CHI	ACM Conference on Human Factors in Computing Systems
CI	Confidence Interval
CSV	Comma-Separated Values
CV	Coefficient of Variation
ETHE	International Journal of Educational Technology in Higher Education
ETR&D	Educational Technology Research and Development
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HIP	Human Information Processing
HMD	Head-mounted Displays
HTTP	Hypertext Transfer Protocol
IEEE	Institute of Electrical and Electronics Engineers
IPS	Individual Procedure Score
IRB	Institutional Review Board
IS	Information Systems
JILR	Journal of Interactive Learning Environments

JSON	JavaScript Object Notation
JCAL	Journal of Computer Assisted Language Learning
K-12	Kindergarten to twelfth grade
KD2Lab	Karlsruhe Design and Decision Lab
LLM	Large Language Model
M	Mean
MD	Mean Difference
MOOC	Massive Open Online Course
MRTK	Mixed Reality Toolkit
MSLQ	Motivated Strategies for Learning Questionnaire
MR	Mixed Reality
NFI	New Feedback Items
NASA-TLX	NASA Task Load Index
OST	Optical See-Through
PC	Participants in the control group
PT	Participants in the treatment group
RM ANOVA	Repeated Measures Analysis of variance
RQ	Reserach Question
SD	Standard Deviation
STEM	Science, Technology, Engineering and Mathematics
SUS	System Usability Scale
TA	Thematic Analysis
TNFI	Time New Feedback Items
UEQ	User Experience Questionnaire
UEQ-S	User Experience Questionnaire (Short)
UES	The User Engagement Scale
UES-SF	The User Engagement Scale (Short Form)
VST	Video See-Through
VR	Virtual Reality
XR	Extended Reality

1 Introduction[1]

1.1 Motivation

During the past decades, the landscape of education has witnessed a transformation, driven by the ongoing advancement of digital technologies (Alenezi, 2023; Peters, 2000). As digital learning becomes adopted and integrated with traditional methods, research in Human-Computer Interaction (HCI) and educational technologies has increasingly focused on novel pedagogical models, ranging from purely digital learning activities to complex hybrid setups (Mayer, 2014b; Raes et al., 2020). Among these emerging paradigms, *immersive learning* has established itself as a new research frontier. Immersive learning conceptualizes education not merely as the passive consumption of learning content through a screen, as commonly seen in other digital learning formats such as massive open online courses (MOOCs), but as a set of active phenomenological experiences grounded in the affordance of *presence* (Mystakidis and Lympouridis, 2023). This paradigm is realized through a spectrum of immersive technologies, including Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), which collectively allow learners to explore and consume learning material by immersing themselves in the learning context rather than passive observation.

The premise of immersive learning lies in its unique ability to foster highly interactive environments that generate a sensation of "being there": a higher level of presence compared to traditional 2D interfaces. This enables learning to occur within "authentic situations", thereby improving learners' situated cognition during the learning process (Brown et al., 1989, p. 33). Evidence from industry reports suggests that when immersive learning experiences are successfully delivered, they offer a pathway to more cost-effective, scalable, and results-oriented education (Dalton and Ruhl, 2021). Existing studies indicate that learners adopting immersive learning systems acquire skills faster (Laumann et al., 2024), demonstrate greater confidence in applying learned skills in real-world scenarios (Dafoulas et al., 2019), and exhibit higher motivation during learning sessions (Deng et al., 2025; Mohammadhossein et al., 2022).

These benefits have been empirically observed across various educational domains. For instance, in fields where the visualization of abstract concepts is critical, such as physics, im-

[1]This chapter is based on the following studies that have been published, accepted, or under review: Liu et al. (2025b), Liu et al. (2025c), Liu et al. (n.d.), Liu et al. (2026)

mersive learning has proven beneficial for knowledge retention (Tauseef et al., 2024). By allowing students to interact with intangible forces or manipulate variables in a virtual laboratory, immersive systems render complex theoretical concepts tangible, leading to more effective conceptual understanding compared to static textbooks. Similarly, in high-stakes fields including medical training, where real-world practice is often constrained by safety risks, ethical concerns, or costs, simulation-based immersive learning has become a well-adopted strategy (Chernikova et al., 2020). Medical students can practice surgical procedures repeatedly in a risk-free environment, learning from mistakes that would be unacceptable in a clinical setting (Gieser et al., 2024). Furthermore, immersive learning provides a pragmatic solution to the constraints of physical location and time. As demonstrated during the COVID-19 pandemic, VR enables synchronous and asynchronous collaborative learning experiences that bridge geographical distances (Gao et al., 2021). Compared to standard video conferencing, VR systems allow for embodied interaction and a shared sense of space, which are crucial for collaborative problem-solving (Zhang et al., 2022). Beyond these examples, applications extend to music education (Banquiero et al., 2024), programming education (Radu et al., 2021a), and motor training (Lin et al., 2021).

Across these diverse domains, a common finding holds true: immersive learning reliably enhances learners' perceived presence and situational interest. However, despite these advancements, the pedagogical efficacy of these systems presents a complex nuance. While motivational benefits are well-documented, the impact on actual learning outcomes and long-term knowledge retention remains inconsistent (Makransky et al., 2019). A closer examination of existing immersive learning systems reveals design patterns that prioritize short-term engagement over sustained learning (Radianti et al., 2020). Many systems are effective as "one-off" demonstrations but lack the pedagogical scaffolding required for curriculum-integrated learning journeys. Furthermore, the different interactivity levels, ranging from passive observation to complex object manipulation in immersive learning systems, pose a dilemma: while high immersion can promote engagement, it does not automatically equate to learning success (Liu et al., 2025c). Without proper guidance, the richness of the virtual environment can lead to distractions and cognitive overload rather than focused study.

This highlights the critical need for adaptive learning support. In digital learning, adaptive support based on individual characteristics and learning progress is recognized as a key driver of success, predicated on the understanding that learners' cognitive and metacognitive processes are unique (Mayer, 2014a). In the context of immersive learning, the need for such support is even more pronounced: the affordances that make immersive learning effective simultaneously impose cognitive demands on learners (Kockord and Bodensiek, 2021). Therefore, a recent research trend advocates shifting the focus from merely designing immersive *learning content* to designing adaptive *learning support* (Makransky and Petersen, 2021). The goal is to optimize learners' cognitive and metacognitive processes during the immersive experience, ensuring that the technology supports, rather than hinders, the learning process.

A conceptual foundation for understanding these dynamics is provided by the Cognitive Affective Model for Immersive Learning (CAMIL), proposed by Makransky and Petersen (2021). The CAMIL model states that learning outcomes are determined by the interplay of various affective and cognitive factors, including situational interest, motivation, cognitive load, self-regulation, and embodiment. Crucially, *attention* serves as a central moderator among these factors: focused attention is the prerequisite for situational interest (Hidi and Renninger, 2006). Furthermore, instructional design principles, such as the signaling principle, aim specifically to guide learner attention to manage cognitive load (Mayer, 2014b). Finally, maintaining focused attention is essential for self-regulation, allowing learners to resist distractions and focus on learning tasks (Johnson and Davies, 2014). Therefore, to design adaptive learner-centered support, we must first establish a method to interpret and support the learner's attention.

To bridge the gap between theory and adaptive system design, eye-tracking technology offers a window into the learner's attention state. *Gaze-adaptive support* utilizes real-time eye-tracking data not just as an interaction method, but as a continuous source of information for user-centered support (Kurzhals et al., 2020). By analyzing metrics such as fixation duration, saccades, and pupil dilation, systems can infer users' visual attention and underlying cognitive load (Duchowski, 2007; Rayner, 1998). While eye-tracking has been explored for cognitive support in desktop settings - ranging from attentive user interfaces (Roda and Thomas, 2006) to gaze-adaptive reading assistance (Langner et al., 2023) - its application in immersive learning remains limited. With the integration of eye-trackers into modern VR and MR head-mounted displays (HMDs), the technical barrier has been significantly lowered (Abeysinghe et al., 2025), yet design knowledge for effective learner-centered interventions has not kept pace. Existing gaze-adaptive immersive systems primarily focus on non-learning contexts (Lindlbauer et al., 2019; Seele et al., 2017) or isolated interaction techniques (Han et al., 2022), often lacking empirical validation of learning outcomes or integration into a broader pedagogical framework

(Abeysinghe et al., 2025).

Based on the current state of research, I recognize three challenges (C1–C3) that currently hinder the effective integration of adaptive learning support in immersive learning:

- **C1: The Challenge of Attention Management.** The affordances that make immersive learning effective - visuals, spatial audio, and embodiment - simultaneously impose cognitive demands. Learners are susceptible to cognitive overload, because immersive visualization and interactive elements can divert attention from core learning objectives. The challenge lies in managing this rich input to support, rather than overwhelm learners' limited cognitive resources.
- **C2: The Lack of Gaze-Adaptive Support.** Most current immersive systems adopt a static, one-size-fits-all approach of learning support. Despite the availability of embedded eye-tracking hardware, there is a lack of intelligent pipelines that adapt to the individual learner's attentional state. The challenge is to translate raw gaze data into meaningful pedagogical interventions that scaffold the learning process dynamically.
- **C3: The Isolated Learning Experience.** Immersive learning often suffers from a silo effect, where the experience ends when the headset is removed. Current designs rarely account for the full learning journey, which includes knowledge consolidation, reflection, and review. The challenge is to break this isolation by designing cross-device ecosystems that bridge the in-situ immersive session with ex-situ activities on ubiquitous devices, ensuring knowledge retention and transfer.

Therefore, this dissertation aims to address these identified challenges by adopting a holistic perspective. By investigating gaze-adaptive interventions across different degrees of immersion and learning setups, ranging from single-device interactions to cross-device ecosystems, this work seeks to provide a conceptual framework, novel artifacts with empirical evidence, and actionable design knowledge. The research objective is to move beyond the novelty of immersive learning and create learner-centered experiences that effectively understand, adapt to, and support the learner's attention management throughout the learning journey.

1.2 Research Gaps and Questions

The challenges outlined in the motivation highlight the complex landscape of immersive learning and highlight how gaze-adaptive support can promote better user experiences and learning outcomes. To summarize the problem space, the novelty of immersive learning systems and their design affordances can cause cognitive overload and attention management problems (C1). While educational research acknowledges that cognitive processes in learning are complex and no single strategy addresses the overall cognitive challenge (Schunk, 2012), attention management remains a critical prerequisite for success in novel learning environments such as immersive learning. Furthermore, while the value of biosignal-adaptive systems, including gaze-adaptive support using eye-tracking, has been well-researched in desktop settings (Duchowski, 2007; Schultz and Maedche, 2023), there is a lack of research investigating how to transfer these techniques to optimize immersive learning experiences (C2). Finally, learner support must not be treated as an isolated design intervention confined to the immersive environment. As learning is an inherently continuous and ubiquitous process that occurs both during and after a session (Milrad et al., 2013), gaze-adaptive support must extend beyond the VR or MR headset. Bridging the immersive experience with continuous, cross-device support is therefore essential for learning success (C3). Based on these three challenges, this dissertation proposes the following guiding research question (RQ):

> **Main RQ:** How to design gaze-adaptive support to effectively assist learners in immersive learning?

To address the overarching main research question, this dissertation comprises four distinct studies, each guided by specific sub-questions. Collectively, these studies offer a systematic investigation of user challenges in immersive learning, addressing them across diverse technological and contextual configurations. First, regarding the technological scope, the research covers the full breadth of the *Reality-Virtuality Continuum* (Milgram and Fumio, 1994), encompassing AR, MR, and VR. Second, recognizing that learning is a continuous process that transcends any single device (Milrad et al., 2013), this work extends its scope beyond the immersive headset to broader device types. Drawing on the taxonomy of cross-device interaction by Brudy et al. (2019), the studies investigate support mechanisms ranging from *single-device* *interactions* (isolated HMD use, e.g., Study II) to complex *cross-device interactions*. These include both “synchronous” (parallel use of devices, e.g., Study III) and “sequential” (shifting tasks between devices over time, e.g., Study IV) scenarios (Brudy et al., 2019, p. 4). This holistic approach aligns with the pedagogical concept of seamless learning, ensuring that learners are supported throughout different stages of their learning journey - a research area that, to the best of my knowledge, has yet to be systematically explored in the field of gaze-adaptive immersive support.

To establish a solid conceptual foundation, my research begins with a comprehensive overview of the immersive learning landscape. Within the *Reality-Virtuality Continuum*, existing research have been conducted to examine the state of the art of immersive VR learning systems (Radianti et al., 2020) and AR systems implemented on handheld devices such as mobile phones or tablets (Akçayır and Akçayır, 2017a). However, MR learning systems implemented with HMDs, which blend physical surroundings and virtual elements while retaining hands-free interaction (Speicher et al., 2019), have not been adequately investigated regarding their design dimensions, application domains, and specific user challenges. Given that MR represents a distinct and integral segment of the immersive learning spectrum, this lack of review represents the first research gap: **Gap 1: Lack of a Comprehensive Understanding of Mixed Reality Learning Systems.** To address this, Study I conducts a systematic literature review specifically targeting head-mounted MR systems in higher education, analyzing current design patterns, user challenges, and evaluation methodologies. Consequently, I formulate the following research question for the first study:

RQ1: What is the state of the art of mixed reality learning systems in higher education?

- **RQ1a:** Which types of head-mounted MR technologies have been applied in higher education across different fields?
- **RQ1b:** Which learning paradigms and theories have guided the design of MR learning systems in higher education?
- **RQ1c:** In which fields of higher education have MR learning systems been implemented and studied?
- **RQ1d:** What design features are commonly integrated into MR learning systems for higher education?
- **RQ1e:** What research methods and study designs have been used to evaluate the learning outcomes of MR learning systems?

Following the insights from the first study, user challenges regarding attention management and the potential for eye-tracking interventions became evident. To go beyond conceptual understanding to design instantiations, the subsequent studies focus on designing, implementing and empirically evaluating gaze-adaptive support systems for immersive learning, following a human-centered design approach. This leads to the identification of the second research gap: **Gap 2: Insufficient Design Knowledge for Gaze-Adaptive Support in Immersive Learning.** Although gaze-based interaction and adaptive support are established paradigms in desktop settings, it remains unclear how to effectively transfer and adapt these techniques to immersive environments. As a result, current research lacks tangible artifacts designed to support learners' attention management in immersive learning environments. Moreover, the few existing studies that have proposed gaze-adaptive support for immersive learning, including the study by Han et al. (2022), lack the empirical evidence necessary to derive actionable design knowledge for future research. To address this, I first investigate *attention feedback*. Attention feedback based on eye-tracking is a well-established gaze-adaptive support technique in desktop settings, and it can effectively promote self-reflection in the revisit phase of information processing, eventually leading to better attention management (Toreini et al., 2022). However, its application as a support mechanism in immersive learning remains unexplored. Therefore, I formulate the second research question:

RQ2: How to design attention feedback for immersive learning to improve students' self-reflection?

While Study II addressed in-situ support during immersive learning with attention feedback, its findings reveal a further research direction: adaptive support should not end when the user removes the HMD and exits the immersive learning environment. This perspective aligns with the conceptual framework of *seamless learning* (Albeedan et al., 2024; Sharples, 2015), which views learning as a continuous process independent of a single location or device. Currently, this continuity is often broken in immersive learning designs. For example, in Study I, users could review their learning progress with attention feedback as a gaze-adaptive support technique. However, both the support and the information on individual learning progress became inaccessible after users removed the headset. Thus, I identify the third research gap: **Gap 3: The Disconnect Between Immersive Experiences and the Ex-Situ Learning Journey.** To bridge the gap between immersive experiences and external learning contexts, I reviewed strategies capable of connecting learning activities across different times and spaces and identified *note-taking* as an established support technique. Therefore, Study III investigates note-taking as another gaze-adaptive design. Note-taking is a validated method for cognitive offloading and knowledge transfer across contexts (Moos, 2009), effectively bridging the gap between immersive and follow-up activities. To address Gap 3, Study III proposes a gaze-aware note-taking tool designed with cross-device interaction, allowing contextual information with attention heatmaps captured in the HMD to be retrieved on mobile devices (tablets or smartphones). This leads to the third research question:

> **RQ3:** How to design gaze-adaptive note-taking for immersive learning to support effective cross-device and context-aware information transfer?

Aiming to extend the scope of this dissertation, I reviewed the preceding studies. The previous studies focused on immersive learning with head-mounted MR technology, in which learners can still perceive their physical surroundings during the learning experience. This doesn't represent the whole spectrum of immersive technologies. Second, both Study II and Study III targeted self-directed learning activities with active exploration, allowing learners to independently determine their learning paths within the immersive learning system. Nonetheless, there are other learning formats that define a more structured and linear learning process such as lectures in virtual classrooms. Third, Study II focused exclusively on in-situ learning, where learners received attention feedback directly within the immersive learning system. While Study III introduced cross-device interaction to support ex-situ review and learning, the ex-situ learning journey itself was neither tracked nor systematically evaluated. Addressing these limitations, Study IV investigates a different immersive learning format: immersive virtual classrooms in VR. Moreover, this study introduces a novel learning ecosystem comprising a virtual classroom, an attention-driven personalization pipeline, and the integration of gaze-adaptive post-lecture quizzes delivered through a mobile assistant. To systematically evaluate this learning ecosystem and examine the effectiveness of gaze-adaptive post-lecture quizzes for ex-situ learning, Study IV was conducted as a four-week field experiment with 36 participants. Participants not only engaged with lecture consumption in VR but also received personalized support — generated from their VR-based attention metrics — on their smartphones with a mobile learning assistant. This final study aims to address the following research question:

> **RQ4:** How to design gaze-adaptive post-lecture support for immersive learning to improve engagement, motivation, and learning outcomes?

The following section outlines the structure of this dissertation, mapping the four studies to their corresponding publications or planned submissions and illustrates how they are organized to form the cumulative body of work.

1.3 Thesis Structure

This cumulative dissertation consists of seven chapters. This first chapter provides an overview of the research motivation, identifies the research gaps and questions, and outlines the overall structure of the thesis. The identified challenges and the corresponding research questions are addressed through four studies, which constitute Chapter 2 to Chapter 5.

Chapter 2 presents Study I, a systematic literature review conducted following the methodology of Kitchenham and Charters (2007) and reported in accordance with the PRISMA guidelines (Page et al., 2021). The chapter reviews and synthesizes findings from 80 research articles on MR learning systems using HMDs in higher education. It develops a conceptual framework comprising five dimensions, which address (1) the devices and technologies used to build MR learning systems, (2) the learning paradigms and theories informing system design, (3) the fields of education, (4) design and interaction features, and (4) the research designs of the reviewed empirical studies.

The results highlight several research gaps and design implications. First, there is a need for adaptive learning support as a design feature. Second, although learning paradigms and theories have been explored and applied in digital learning (Alenezi, 2023), few immersive learning systems incorporate established paradigms and theories to guide their design. Furthermore, the study identifies an opportunity to leverage biosignal-adaptive technologies to support learners' cognitive processes. Biosignal-adaptive technologies, including eye tracking, have been used to respond to users' cognitive states through adaptive user interfaces and cognitive support (Doswell and Skinner, 2014; Schultz and Maedche, 2023); however, their application in immersive learning requires further research. Overall, this chapter answers RQ1 and addresses Gap 1 by providing a conceptual framework and design implications that inform the subsequent studies.

Chapter 3 presents Study II. In this study, I conducted co-design workshops following a participatory design approach (Muller and Kuhn, 1993) to conceptualize, implement, and evaluate an attention-feedback support tool as an integrated system component for head-mounted MR. The system was evaluated with 22 participants using a mixed-methods approach. The results indicate that attention feedback can effectively reveal learners' knowledge gaps immediately after the content exploration phase. This mitigates potential cognitive overload during initial exposure to the novel technology and learning content, and it supports self-reflection during the revisit phase of immersive learning. This study answers RQ2 and addresses Gap 2.

Chapter 4 introduces Study III. This chapter presents a novel note-taking technique that integrates gaze-aware attention heatmaps with cross-device interaction. By exploring further designs for gaze-adaptive support, this study also addresses Gap 2 and represents a first attempt to bridge in-situ and ex-situ learning contexts through cross-device note-taking, thereby answering RQ3 and addressing Gap 3.

Chapter 5 reports on Study IV. The chapter introduces *AttentiveLearn*, a novel learning ecosystem that addresses the gap of bridging in-situ and ex-situ learning (Gap 3). The ecosystem comprises three modules — a VR classroom, a personalization pipeline for gaze-adaptive quizzes, and a mobile learning assistant — that connect the immersive learning experience with adaptive post-lecture support. To systematically evaluate the ecosystem, a four-week field study was conducted. The results indicate that gaze-adaptive quizzes can improve learners' motivation, engagement, and learning outcomes throughout the learning journey, although the longitudinal impact on learning success requires further investigation. This study provides new design knowledge and insights on extending learner-centered support from the immersive environment to the ex-situ review process, answering RQ4 and addressing Gaps 2 and 3.

Chapter 6 discusses the practical and theoretical contributions of the conducted studies, reflects on the limitations of the presented research, and outlines directions for future research. Finally, **Chapter 7** provides a summary of the dissertation and serves as the conclusion.

The studies presented in this dissertation have been published in or submitted to peer-reviewed international journals and conferences. Study I has been published in the *Technology, Knowledge and Learning* journal. Study II has been published in the *International Journal of Human-Computer Studies*. Study III has been submitted to the *Interacting with Computers* journal.

Study IV has been accepted for publication at the *ACM Conference on Human Factors in Computing Systems (CHI)* at the time of writing. An overview of all publications related to this dissertation can be found in the List of Publications in the appendix.

Table 1.1: Overview of thesis structure

	Chapter
1	**Introduction**
	Gap 1: *Understanding MR Learning Systems* | ***Result:*** *An Integrated Conceptual Framework*
2	**Study I (RQ1)** Liu, S., Toreini, P., & Maedche, A. (2025). *Mixed Reality Learning Systems with Head-Mounted Displays in Higher Education: A Systematic Review.* Technology, Knowledge and Learning. https://doi.org/10.1007/s10758-025-09912-z
	Gap 2: *Gaze-Adaptive Support in Imm. Learning* | ***Artifact:*** *An Attention Feedback System for HMD-AR*
3	**Study II (RQ2)** Liu, S., Toreini, P., & Maedche, A. (2025). *AF-Mix: A gaze-aware learning system with attention feedback in mixed reality.* International Journal of Human-Computer Studies, 198, 103467. https://doi.org/10.1016/j.ijhcs.2025.103467
	Gap 2: *Gaze-Adaptive Support in Immersive Learning* | ***Gap 3:*** *Disconnected Imm. and Ex-Situ Learning* | ***Artifact:*** *An Gaze-Aware Note-Taking System for HMD-MR*
4	**Study III (RQ3)** Liu, S., Toreini, P., & Maedche, A. (n.d.). *GazeNote: Designing Note-taking Support for Immersive Learning with Gaze-Aware and Cross-Device Interaction.* (Under Review at Interacting with Computers)
	Gap 2: *Gaze-Adaptive Support in Immersive Learning* | ***Gap 3:*** *Disconnected Imm. and Ex-Situ Learning* | ***Artifact:*** *Cross-Device Gaze-Adaptive Quizzes for Immersive VR*
5	**Study IV (RQ4)** Liu, S., Feick, M., Bierhoff, L., & Maedche, A. (2026). AttentiveLearn: Personalized Post-Lecture Support for Gaze-Aware Immersive Learning. Proceedings of the CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3772318.3790667
6	**Discussion**
7	**Conclusion**

2 Mixed Reality Learning Systems with Head-Mounted Displays in Higher Education: A Systematic Review (Study I)

Abstract: The advent of mixed reality (MR) technology has sparked emerging adoption of MR learning systems in higher education. In this systematic literature review, we examine the state of the art of MR systems in this field. We highlight the diverse applications and benefits of MR technology, including design features and specific fields of higher education addressed in existing studies. Despite its potential, our review also identifies research gaps, particularly regarding effective evaluation methods for learning outcomes. We identify the needs to provide advanced learning support and to better integrate established learning theories to guide more effective system design. Furthermore, we discover that leveraging biosignals to design biosignal-adaptive MR learning systems providing personalized learning support is currently underutilized. Through this systematic review, we aim to offer design implications for the application of MR in higher education, identify potential research gaps, and provide insights for future work in this rapidly evolving field.

Keywords: Mixed reality, Human-computer interaction, Systematic literature review, Higher education

Published as: Liu, S., Toreini, P., & Maedche, A. (2025). Mixed Reality Learning Systems with Head-Mounted Displays in Higher Education: A Systematic Review. Technology, Knowledge and Learning. https://doi.org/10.1007/s10758-025-09912-z

2.1 Introduction

Educational technology has witnessed rapid development, with researchers exploring innovative tools to enhance learning experiences (Mitsuhara and Shishibori, 2015; Perez and Torres-Delgado, 2023). Among these advancements, Mixed Reality (MR) using head-mounted displays (HMDs) has emerged as a new frontier, offering immersive experiences that merge the virtual and physical worlds (Milgram and Fumio, 1994; Speicher et al., 2019). Although not yet widely adopted across all educational contexts, HMD-based MR is increasingly recognized for its unique capabilities. Its application has been explored in diverse application fields, from virtual physics laboratories to immersive anatomy lessons (Grad et al., 2023; Laumann et al., 2024), often demonstrating benefits over traditional learning formats (Mohammadhossein et al., 2022; Su et al., 2025). Moreover, MR technology is considered foundational for the future of collaborative learning in the Metaverse (Mystakidis, 2022).

This growing interest arises from MR's unique affordances in addressing pedagogical challenges that are difficult to overcome with other immersive technologies. Unlike Virtual Reality (VR), which isolates learners from their physical environment, MR enables simultaneous interaction with real-world objects, making it particularly suitable for procedural skills training. For instance, a medical student can use an MR headset to practice a surgical procedure on a physical manikin while holographic anatomical models are overlaid directly onto it - a task that is challenging in the fully occluded environment of VR (Zhao et al., 2025). Moreover, MR's head-mounted, hands-free interaction provides a distinct advantage over Augmented Reality (AR) on mobile phones or tablets. Systems such as *SkillAR*, for example, can teach complex two-handed motor skills by delivering real-time holographic feedback on body and hand movements (Diller et al., 2025) - a task incompatible with mobile AR, which requires one hand to hold the device. These examples are not isolated cases but illustrate a broader class of learning tasks in which MR's capacity to merge the virtual and physical worlds offers distinct pedagogical potential.

Recognizing this potential, researchers have explored the use of MR technology across many fields of education (Lungu et al., 2021). In engineering education, MR supports the visualization of additive manufacturing by enabling students to interact with computer-aided design (CAD) models, thereby fostering a deeper understanding of engineering processes (Melo et al., 2024). The technology has also shown promise in vocational training, where realistic simulations of construction sites contribute to hands-on skill development and safer practice opportunities (Nguyen et al., 2021).

While existing studies have examined MR across different fields of education, they also span a wide range of education levels, from preschool to higher education. Within higher education, MR learning applications have shown promising results (Richards, 2023). However, knowledge about MR learning systems in this context remains fragmented. Other immersive technologies, such as VR, have already been the subject of systematic literature reviews (Hamilton et al., 2021; Radianti et al., 2020), yet the specific strengths, advantages, and design features of MR in higher education remain largely underexplored. Existing reviews on MR technology, such as those by Lungu et al. (2021) and Maas and Hughes (2020), often focus on different education levels (e.g., K-12 (Maas and Hughes, 2020)) or narrow domains such as language learning (Li and Wong, 2021) and medical education (Barteit et al., 2021).

Thus, a comprehensive understanding of MR learning systems specifically designed for higher education is still lacking. In the absence of a systematic overview, researchers and practitioners face challenges in building on prior work, identifying best practices, and synthesizing evidence on learning outcomes. By concentrating on higher education, our work provides a focused perspective that differs from broader surveys covering multiple educational levels or reviews restricted to single disciplines. To address this gap, we present a systematic literature review of the state-of-the-art research on MR learning systems in higher education. We aim to summarize existing knowledge and provide a clear overview by answering the following five research questions:

1. Which types of head-mounted MR technologies have been applied in higher education across different fields?

2. Which learning paradigms and theories have guided the design of MR learning systems in higher education?

3. In which fields of higher education have MR learning systems been implemented and studied?

4. What design features are commonly integrated into MR learning systems for higher education?

5. What research methods and study designs have been used to evaluate the learning outcomes of MR learning systems?

In addressing these research questions, this paper contributes an integrative conceptual framework that synthesizes the state of the art of MR learning systems in higher education. The framework organizes existing evidence across five key dimensions—devices and technologies, learning paradigms and theories, fields of education, design features, and study design—offering a structured lens for understanding current developments. By doing so, we aim to provide a foundation that supports researchers, designers, and educators in leveraging MR learning systems for higher education.

2.2 Related Work

2.2.1 Mixed Reality Learning Systems

The term Mixed Reality (MR) was originally defined by Milgram and Fumio (1994) as a continuum spanning from AR to augmented virtuality. With the recent advancement of immersive technology and the wide adoption of the term, a shift and polarization of its definition has been observed (Speicher et al., 2019). In the scope of this literature review, we adopt the definition used by researchers including Yue et al. (2017) as well as Roo and Hachet (2017). In this interpretation, MR is defined as an advanced form of AR that enables interaction with virtual overlays, typically realized through HMDs. This definition of "strong AR", as reported in an overview by Speicher et al. (2019), has also been adopted by the industry, including companies such as Microsoft (2023). Compared to traditional AR experience with mobile phones or tablets, MR systems allow dynamic and interactive virtual elements to blend in the environments, instead of being static on-screen overlays on top of the physical surroundings. Therefore, this review focuses on MR learning systems implemented on HMDs. This definition distinguishes MR from other immersive technologies: VR systems, for instance, are excluded as they fully occlude the user's perception of the physical environment. Meanwhile, AR on mobile devices with 2D displays, such as smartphones or tablets, also falls outside our scope; these systems are typically limited to displaying virtual information and lack the capacity for direct spatial interaction between physical and virtual content.

MR has been explored across different fields of education. For instance, it has been applied in medical education to train specific anatomical techniques and surgical procedures, leading to effective knowledge retention (Moro et al., 2017). It has also been used in language learning to support the acquisition of sign languages by promoting motor learning (Shao et al., 2020). Other prominent applications of MR learning systems can be found in STEM education (Ibáñez and Delgado-Kloos, 2018). These systems have been shown to facilitate learners' understanding of complex content, reduce cognitive load, and increase motivation (Mohammadhossein et al., 2022). For example, Khan et al. (2019) evaluated an MR system designed to visualize abstract concepts in physics and found significant improvements in students' comprehension of the subject matter. In programming education, studies such as Hennerley et al. (2017) likewise demonstrated that MR systems can enhance students' motivation and conceptual understanding. Further benefits include the promotion of immersion and presence, both of which can positively influence learning outcomes (Mystakidis et al., 2022; Ryan et al., 2022). Moreover, MR can foster kinesthetic and motor learning by encouraging learners to integrate physical activities into their learning experience, thereby enhancing motivation (Iqbal et al., 2019). Similar benefits have also been identified in higher education contexts (Tang et al., 2020).

Nonetheless, despite the growing body of research on MR learning systems, a systematic overview of their design characteristics and outcomes is still lacking. Several studies have pointed out potential limitations and disadvantages. For example, Vovk et al. (2018) reported that MR learning systems may increase cognitive load. Likewise, a systematic review by Buchner et al. (2022) indicated that some studies observed undesirable cognitive overload, whereas others did not report such effects. Overall, these findings complicate earlier assumptions that MR is mostly beneficial for learning (Mohammadhossein et al., 2022). Therefore, a research gap remains in developing comprehensive understanding of MR learning systems across different contexts, including higher education, which to the best of our knowledge has not yet been addressed in educational technology research.

2.2.2 Learning Paradigms in Higher Education

In the field of educational research, various learning paradigms and theories have been proposed and applied to designing learning systems. Based on the classification of Schunk (2012), common learning paradigms include the following: behaviorism, cognitivism, and constructivism. The behaviorist learning paradigm has its roots in the well-established psychological theory of conditioning model (Thorndike, 1932). Here, learners are believed to passively receive the teaching from educators and need to be disciplined to better accept the knowledge, e.g., by forming habits, introducing sequential curricula, and giving instructions. One of the exemplary learning theories in the scope of behaviorism is the principle of reinforcement, which stresses the effectiveness of positive feedback and rewards for desired learning behavior and punishment for unwanted behavior (Skinner, 1961). With the further development of psychology and the increased criticism of behaviorism, cognitivism, and constructivism have gained popularity over the past decades. Both paradigms acknowledge learning as an active process executed by learners to foster knowledge. From a cognitivism perspective, learning involves different perceptual and cognitive processes that allow learners to acquire information, foster understanding, solve problems, and encode the information in the memory (Bruner, 1960; Shuell, 1986). Therefore, following cognitivism, the goal of designing an effective learning experience is to optimize the process with support, including providing social cognitive support (Benight and Bandura, 2004) or optimized instructions for information processing (Mayer, 1996). Lastly, constructivism underlines the active learning and knowledge-fostering process of learners. In the framework of constructivist learning, learners need to use their cognitive skills to construct new connections and structures of knowledge acquired in the external world and extend their knowledge base by interacting with the external environment (Schunk and Greene, 2017).

In higher education, these learning paradigms have been applied to create effective learning experiences and improve learning outcomes. For example, cognitivism has been highlighted in higher education with a focus on training the cognitive skills of students, including critical thinking, making abstracts, emotion regulation (Evans et al., 2010; Sprague, 1981). Meanwhile, the constructivist paradigm has been applied in higher education to promote active learning and creative thinking for students (O'Connor, 2022). Research has also been conducted to understand the impact of e-learning systems from a constructivist perspective in higher education (Bognar et al., 2015). A focus on constructivism in higher education is to strike a balance between providing effective guidance and providing self-directed learning opportunities to promote active learning (Kirschner et al., 2006). For the behaviorism paradigm, despite criticism over the past decades, its implications on designing learning systems can still be found in the practice of higher education (Weegar and Pacis, 2012).

Apart from these three major learning paradigms, we can identify various works that leveraged other learning paradigms or theories in their research. For example, experiential learning and inquiry-based learning have been applied to encourage learners to actively explore the learning content and seek answers to questions they propose during the learning process (Huang, 2019; Maaß and Artigue, 2013). These learning methods are sometimes seen as synonyms along with other terms including discovery learning, often considered as a specific example of constructivism (Kirschner et al., 2006). Furthermore, there is research that does not connect to any of the above-mentioned learning paradigms and follows alternative paradigms. Examples of these paradigms include simulation-based learning (Viglialoro et al., 2021), collaborative learning (Ali et al., 2019). Therefore, it is worth mentioning that there has not been an exhaustive list of learning paradigms and theories, as many are intertwined with each other and often defined from different perspectives. Due to the complexity of the topic, our work will not target providing a comprehensive classification of learning paradigms and theories. Instead, we will use the existing classification in existing literature reviews as a foundation and then develop a conceptual framework inductively.

2.3 Method

We conducted our systematic literature review based on the framework proposed by Kitchenham and Charters (2007), and we report our findings following the PRISMA 2020 guidelines (Page et al., 2021). The review process began with a preliminary scoping analysis to situate our work within the existing reviews and identify a clear research gap.

2.3.1 Scoping Analysis

First, aiming to better understand the related work and current research gaps in the state-of-the-art research, we used the following search string to find existing reviews focusing on learning systems, including MR as well as other immersive technologies including VR and AR:

```
("review") AND ("learning" OR "education") AND ("Extended Reality" OR
"XR" OR "Augmented Reality" OR "AR" OR "Virtual Reality"
OR "VR" OR "Mixed Reality" OR "MR")
```

We used Scopus as our main database for this scoping analysis. In total, 354 reviews were identified using the search string. To narrow this to the most relevant publications within the educational technology research, we filtered the results to include only major research outlets. These venues were selected based on their Google Scholar ranking and h-index, resulting in a focus on the following key journals and conferences: *Computers & Education*, *Education and Information Technologies*, *British Journal of Educational Technology* (BJET), *International Journal of Educational Technology in Higher Education* (ETHE), *Educational Technology Research and Development* (ETR&D), *Journal of Interactive Learning Environments* (JILR), *Journal of Computer Assisted Language Learning* (JCAL), and *International Journal of Emerging Technologies in Learning*.

In these outlets, literature reviews on MR systems have been conducted (Rokhsaritalemi et al., 2020). However, our analysis revealed that they often have different focuses, such as revealing challenges (Moser et al., 2019), trends (Rokhsaritalemi et al., 2020), interaction techniques (Papadopoulos et al., 2021; Plopski et al., 2022), or specific application domains outside of education (De Guzman et al., 2019). After further examination, we identified seven highly relevant literature reviews from these outlets (summarized in Table 7.5 in Appendix).

Our analysis of the seven selected reviews indicates diverse adopted methodologies. Some leveraged established frameworks from the Information Systems (IS) community (Kitchenham and Charters, 2007), a research field that focuses on systems that collect and process information and analyzes their impacts on individuals, organizations and societies. Meanwhile, other reviews did not specify a particular methodology. Regarding research objectives, many reviews concentrate on specific fields of education - such as language learning (Li and Wong, 2021), surgical simulation (Lungu et al., 2021), or STEM (Ibáñez and Delgado-Kloos, 2018). Five of the seven reviews did not concentrate on a specific education level. Instead, their focus was primarily on the technology itself (Jensen and Konradsen, 2018) or a particular learning strategy like game-based learning (Yu et al., 2022). For the two reviews that did address the higher education context, one targeted VR technology (Radianti et al., 2020) and the other focused on AR within STEM education (Hidayat and Wardat, 2023). Other reviews, particularly from the Human-Computer Interaction (HCI) community, where researchers design and evaluate interactive systems using human-centered methods to fulfill user needs, have examined the use of specific sensors in immersive learning systems, such as eye-tracking (Shadiev and Li, 2023).

Therefore, we identified a clear research gap after the scoping analysis: none of the existing literature reviews provides a comprehensive investigation of MR learning systems specifically within the higher education context. Our study is designed to address this gap.

2.3.2 Conceptual Framework

Aiming to answer our research questions and systematically analyze the relevant literature, we propose a conceptual framework with multiple dimensions. The development process for each dimension is detailed below.

For the **device and technology** dimension (RQ1), we follow the work of Rolland and Fuchs (2000) and distinguish between *video see-through* devices and *optical see-through* devices. Video see-through devices present the physical surroundings as a video stream, which is rendered together with the virtual elements and presented in the display. Therefore, VR headsets are adopted as video see-through devices by integrating outward-facing cameras to capture the surroundings. Common video see-through devices used in the research include Oculus/Meta Quest, HTC Vive (Barteit et al., 2021). In contrast, optical see-through devices have a transparent display, allowing users to directly perceive the physical surroundings with naked eyes. The virtual content is then rendered on the transparent display to align with the physical surroundings. Exemplary devices in this category include Microsoft HoloLens (Barteit et al., 2021). A detailed explanation is provided in Table 7.1 in the Appendix.

For the **learning paradigms and theories** (RQ2), existing reviews have demonstrated the difficulty of creating an exhaustive taxonomy (Radianti et al., 2020). We therefore established a foundational framework based on the primary paradigms in educational research - *behaviorism*, *cognitivism*, and *constructivism* - as described by Schunk (2012). We acknowledge the following challenges of creating this taxonomy: First, there is an unclear relationship between the different learning paradigms. For example, experiential learning is sometimes seen as a synonym for exploratory learning, but often recognized as an independent paradigm of experientialism as well (Chen et al., 2024; Seaman et al., 2017). Second, other learning theories are also frequently referenced (Goldie, 2016). In some cases, they are treated as subcategories of the three paradigms, while in others they appear as independent approaches, such as self-regulated learning, active learning, collaborative learning, and gamification. Accordingly, the initial set of categories presented in Table 7.2 was designed to be extensible, allowing additional theories to be incorporated inductively during the review process.

For the **fields of education** dimension (RQ3), we adopted a bottom-up approach. Rather than using a predefined list, we categorized the disciplines during the data analysis, a common and effective method in educational technology reviews.

For the **design features** dimension (RQ4), we refer to the existing literature reviews of VR systems. In the two reviews focusing on higher education by Radianti et al. (2020) and Wohlgenannt et al. (2019), they have defined various design features including different interaction levels, instructions, user-generated content. Meanwhile, in the work of Won et al. (2023), design features are classified into sensory, actional, narrative, and social elements. To create a comprehensive overview, we therefore synthesized these varied approaches to form our initial analytical framework (see Table 7.3).

The **study design** dimension (RQ5) encompasses the overall research methodology, specific data collection techniques, and the outcomes measured. We classified the research methods as qualitative, quantitative, or mixed-methods, drawing on established classification (Creswell, 2009; Edmonds and Kennedy, 2016). We acknowledge the existence of concept-driven research methods, such as the approach proposed by Stolterman and Wiberg (2010). However, our review focuses exclusively on empirical studies. Consequently, conceptual research was not included in the conceptual framework (see Table 7.4). To address the absence of a comprehensive overview of assessment tools for MR learning systems, we introduced a data collection sub-dimension (see Table 7.6). The initial list of methods was synthesized from prior work (Cairns and Cox, 2008; Radianti et al., 2020) and may be extended during the coding process based on the reviewed articles. In addition, the learning outcomes evaluated in each study (e.g., motivation, usability, knowledge gain) will be systematically coded as part of this process.

2.3.3 Search Strategy

Based on the guideline of Kitchenham and Charters (2007), we created a review protocol describing our search strategy, including the search strings, databases, and inclusion/exclusion criteria. We selected the following databases for identifying relevant literature: *ACM Digital Library*, *Web of Science*, *Scopus*, *IEEE Xplore*, *EBSCO Host*, and *ProQuest*. The selection of databases reflects our focus on educational technology and was informed by our initial scoping analysis. We applied the following search string to each database:

```
("head-mounted" OR "headset*" OR "glass*" OR "smartglass*" OR "gog-
gle*" OR "HoloLens" OR "Magic Leap" OR "Quest")
AND ("Augmented Reality" OR "Mixed Reality" OR "Extended Reality" OR
"XR" OR "AR" OR "MR" OR "immersive")
AND ("educat*" OR "learn*" OR "teach*")
```

The search string was designed to specify the MR technology and the educational context. To ensure the search results aligned with our focus on HMD-based systems, we added the keywords including “head-mounted” as well as common device names including “HoloLens”. This approach of including device names has been established and applied in existing reviews on immersive learning systems (Radianti et al., 2020).

In addition to the search string, we defined several inclusion and exclusion criteria. First, only peer-reviewed articles were included to ensure methodological quality. Second, conceptual or technical papers without empirical studies were excluded. Third, articles had to be situated within a higher education context. Finally, we limited our search to articles published between 2013 and 2025. Even though both the concept of MR and pioneering HMD devices emerged a few decades ago (Milgram and Fumio, 1994), this timeframe captures the recent rapid development and application of modern MR devices. A complete summary of the inclusion and exclusion criteria can be found in Table 7.7.

2.3.4 Data Extraction

We identified an initial list of articles by applying our search string to the selected databases. The first and second authors then independently screened these articles in two phases: first by title and abstract, and then by full text for the remaining articles. Disagreements were resolved through discussion. Following this screening, we extracted data from the final articles into a concept matrix to provide a concept-centric overview of the topic (Webster and Watson, 2002). The complete matrix is available as supplementary material.

Our initial search across six databases identified 7,014 records. Scopus yielded the largest number of results (4,000 articles), partly due to its overlap with other databases such as the ACM Digital Library and IEEE Xplore. After removing 3,138 duplicates using the reference

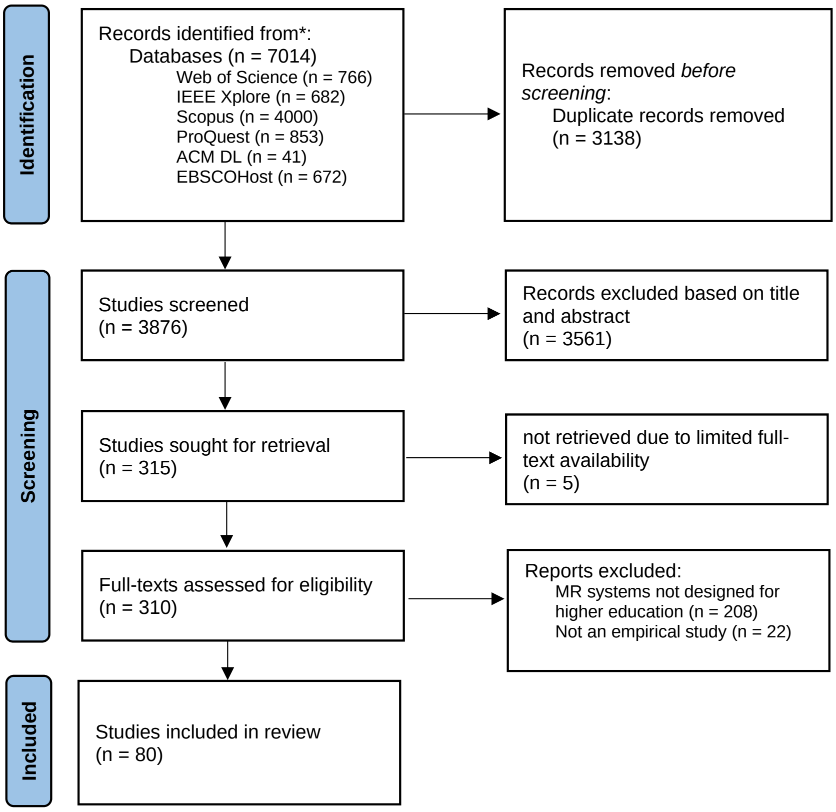

Figure 2.1: Flow chart of the data extraction process in the literature review

management tool Zotero[1], 3,876 unique articles remained. In the first screening phase, 3561 articles were excluded. Records were excluded if they were set in non-educational contexts, did not conform to our operational definition of MR, or focused on other educational levels (such as K-12). The remaining 315 records proceeded to the full-text review phase.

Of the 315 articles assessed at the full-text level, 80 were found to meet the criteria and thus included in the final review. To verify the higher education setting, we first searched the full text for the following keywords: “higher education”, “university”, or “college”. If these keywords were absent, we examined the participant demographics. Studies conducted exclusively with university students were included, a method similar to that of Radianti et al. (2020). The entire literature selection process is illustrated in the PRISMA flowchart (Fig. 2.1).

2.4 Results

We present here the descriptive analysis based on our research questions and the conceptual framework. Each subsection focuses on one particular dimension of the conceptual framework,

[1] `https://www.zotero.org/`

which matches one specific research question. Fig. 2.3 shows the results in a morphological box based on the conceptual framework. Overall, the 80 reviewed articles were published from 2013 to 2025. Fig. 2.2 shows the distribution of published articles throughout the years. Since 2018, there has been a significant increase in the number of published articles, potentially due to the increased availability of widely used MR devices such as Microsoft HoloLens. A drop in publication numbers can be identified in 2021, possibly due to the global impact of the COVID-19 pandemic on research activities.

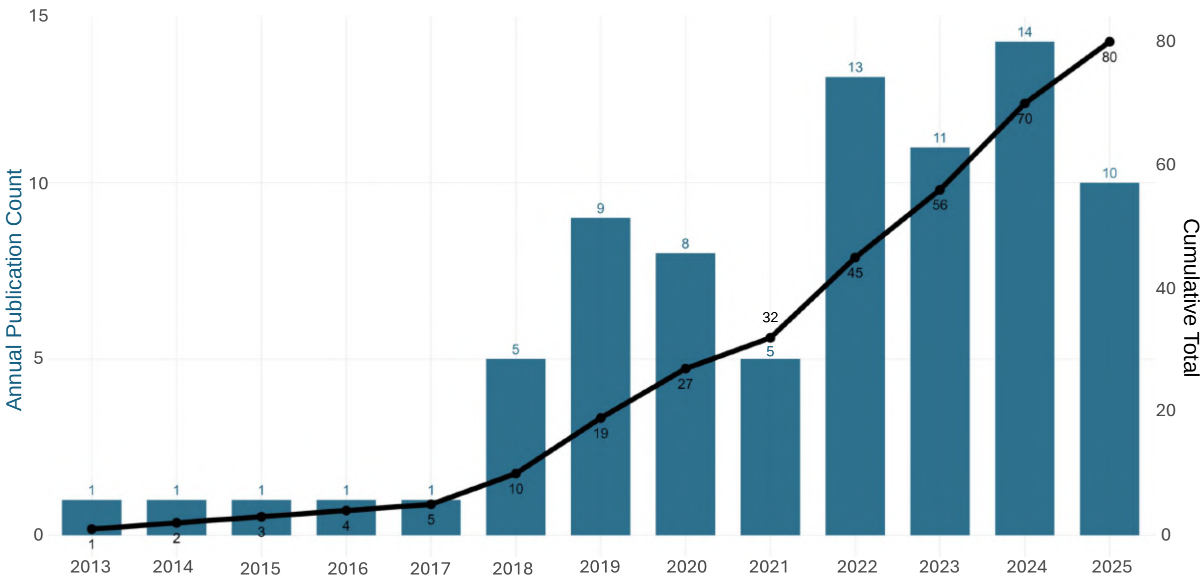

Figure 2.2: Annual and cumulative publication counts over time. The bar chart displays the annual counts of reviewed articles; the overlaid line graph shows the corresponding cumulative total.

2.4.1 MR Devices and Technology

Based on our conceptual framework, we categorize MR technology into two dimensions: video see-through and optical see-through devices. Each dimension comprises different devices, including Microsoft HoloLens and Meta Quest. The results indicate that most researchers employed optical see-through devices such as HoloLens to implement their learning systems. HoloLens has been the most popular option, with 52 articles (65%) reporting the use of either HoloLens 1 or 2. Other optical see-through devices, including Google Glass, were frequently adopted between 2014 and 2017, as well as Magic Leap used in eight articles.

The reviewed studies highlight several benefits of specific devices. For learning systems with high interactivity, HoloLens helps sustain the "spatial contiguity" of tasks through its sensors

Device and Technology	
Video See-through	Meta Quest 2/Pro *(5%)*
	Others *(5%)*
Optical See-through	HoloLens *(65%)*
	Google Glass *(9%)*
	Magic Leap *(10%)*
	Others *(6%)*

Paradigms and Theories
Experiential Learning *(9%)*
Embodied Learning *(5%)*
Simulated Learning *(15%)*
Case-based Learning *(8%)*
Collaborative Learning *(6%)*
Others *(57%)*

Fields of Education
Medical Education *(34%)*
Engineering *(15%)*
Physics *(8%)*
Learning Support *(11%)*
Computer Science *(6%)*
Others *(26%)*

Design Features	
Interaction Level	Passive Observation *(29%)*
	Basic Interaction *(38%)*
	High Interactivity *(33%)*
Social Interaction *(23%)*	
Contextual Instruction *(43%)*	
Feedback *(31%)*	
Embodied Movement *(11%)*	
Role-Playing *(11%)*	

Study Design			
Research Method	Qualitative *(10%)*	Type of Study	Lab Experiment *(89%)*
	Quantitative *(49%)*		Field Study *(11%)*
	Mixed Methods *(41%)*		
Measured Outcome	Usability *(41%)*	Data Collection	Interview *(25%)*
	Performance *(23%)*		Survey *(78%)*
	Knowledge & Understanding *(25%)*		Knowledge Test *(29%)*
	Satisfaction *(10%)*		Interaction Log *(19%)*
	Engagement *(8%)*		Biosignal Data *(6%)*
	Others *(50%)*		

Figure 2.3: A conceptual framework categorizing the reviewed literature across five key dimensions. Percentages indicate the distribution of the 80 selected articles.

and high quality optical see-through (Laumann et al., 2024). For instance, Laumann et al. (2024) employed HoloLens to enhance the learning of optical polarization in a physics laboratory. In this context, it was crucial that the physical environment (optical equipment) remained clear, while the virtual overlays (instructions and sensor readings) were precisely aligned in the physical space. Consequently, HoloLens was identified as the most suitable device. Similarly, Magic Leap offers comparable qualities and has been used as an alternative in several studies (Gießer et al., 2023; Zhao et al., 2025).

A second category of optical see-through devices supports only basic virtual information display without spatial awareness and high interactivity between virtual elements and the physical environment. Their key advantage lies in being lightweight and unobtrusive, making them suitable for tasks that merely require the display of instructions or additional information (Asaumi et al., 2025). Google Glass is a notable example. For instance, Zarraonandia et al. (2019) developed a feedback system that allowed teachers to receive synchronous student feedback displayed directly on the display. Since this scenario required a lightweight device that could be worn throughout a lecture, and the task only involved displaying feedback rather than interaction, Google Glass was considered the most appropriate solution.

Overall, the majority of studies employed optical see-through devices, while only eight used

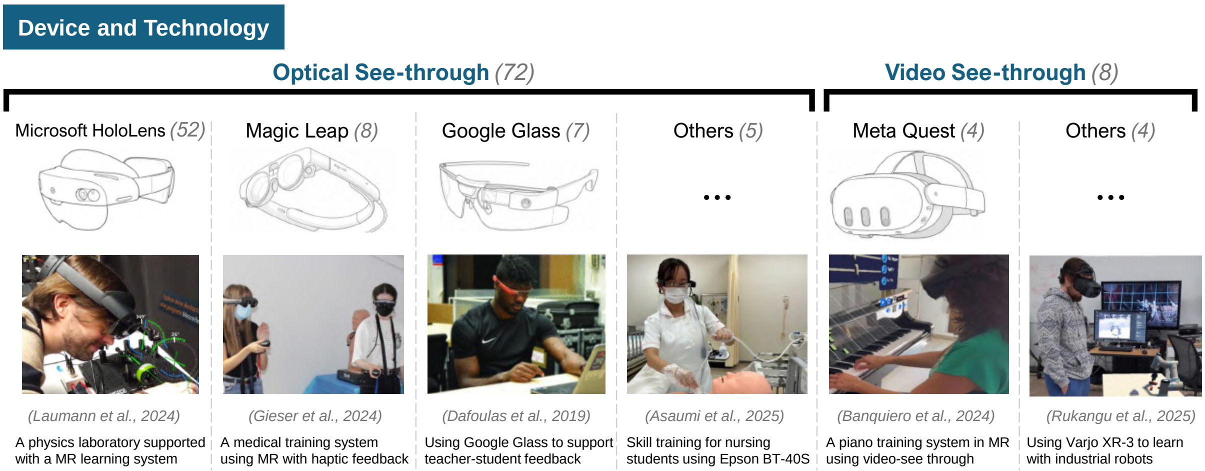

Figure 2.4: An illustration of various MR devices and examples from the reviewed articles

video see-through solutions. Of these, four adapted existing VR headsets such as Meta Quest 2 and Pro, using the integrated cameras to create MR experiences—an increasingly common approach in MR learning research (Ali et al., 2019). However, these studies also noted limitations of video see-through devices. For example, Banquiero et al. (2024) observed a "distorted" effect on Meta Quest devices that hindered effective piano training. By contrast, devices with more advanced sensors and cameras, such as the Varjo XR-3 used by Rukangu et al. (2025), produced more promising video see-through results without distortion of the physical environment. Yet, such devices often require a cable connection to an external PC, which can make the learning experience more cumbersome. The remaining four studies either did not explicitly mention the device model or built their own devices, such as the MR learning system presented by Yang and Liao (2014).

These findings illustrate the dominance of optical see-through devices over the past decade in the field of MR learning systems, with HoloLens to some extent representing MR technology. At the same time, ambiguities remain in the definition of MR, as noted in prior research (Speicher et al., 2019). While authors of some articles also described their systems as AR, others rejected the AR label (often associated with mobile phones or tablets) and instead referred to their works as MR learning systems.

2.4.2 Learning Paradigms and Theories

In the identified articles, various learning paradigms and theories have been identified. Nonetheless, compared to the previously defined theories in our conceptual framework, few of those

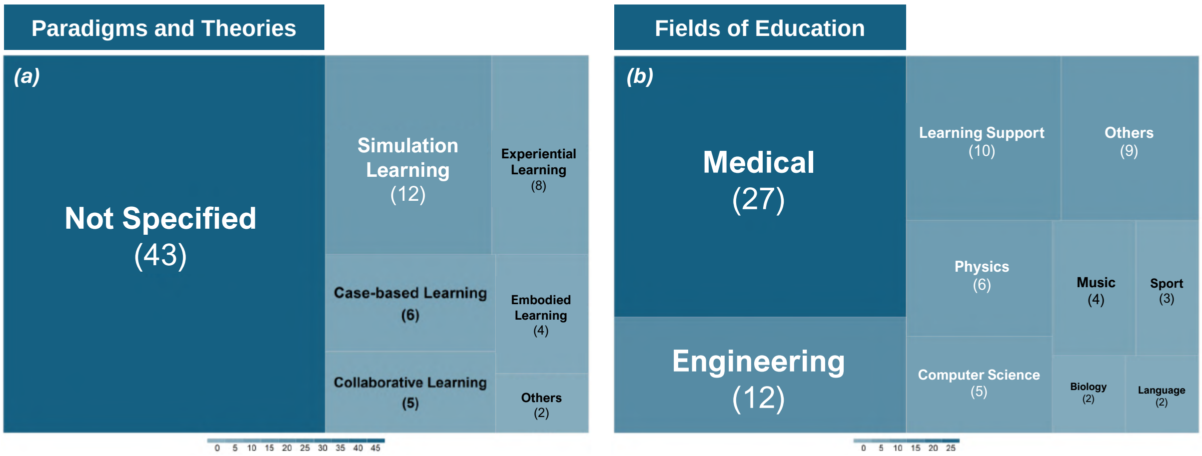

Figure 2.5: Treemaps illustrating the distribution of the reviewed articles. (a) Distribution by the learning paradigms and theories employed; (b) Distribution by fields of education.

learning paradigms have been applied in the articles. More than half of the reviewed articles (43 out of 80) do not explicitly mention a specific learning paradigm or theory that guided their studies.

Because the hierarchies and definitions of learning paradigms are often ambiguous and overlapping, many articles only indirectly referenced their theoretical grounding. Among those that explicitly stated a paradigm, simulation-based learning was the most common, identified in 12 articles. Simulation-based learning is particularly applied in contexts involving risk or cost efficiency, such as medical training, engineering, or physics (Chernikova et al., 2020; Lateef, 2010). For example, Asaumi et al. (2025) implemented an MR learning system for surgical training, where endotracheal intubation was practiced in a controlled MR environment, leading to improved procedural accuracy. Similarly, MR-based tutorials that enhanced clinical instruction in medical education have been designed and evaluated with medical students (Connolly et al., 2024; Gieser et al., 2024). These works also confirm earlier evidence that simulation-based MR environments can foster performance and empathy in learners (Nakazawa et al., 2023; Wunder et al., 2020).

Experiential learning has also been explicitly employed. For example, a "multimodal drumming education tool" designed by Pinkl et al. (2024) leverages MR for embodied rhythm training through first-person action observation and motor engagement, aligning with both experiential and embodied learning paradigms. Similarly, MR-based anatomy teaching reported in Zhao et al. (2025) emphasizes hands-on interaction with 3D neuroanatomical models, allowing learners to actively experience spatial relations that are otherwise difficult to grasp.

Related paradigms found in the literature include case-based learning (Kapp et al., 2020; Schoeb et al., 2020) and embodied learning (Campo et al., 2023), which has gained renewed attention with recent systems such as *SkillAR* that deliver in-situ feedback for skill acquisition in physical training tasks (Diller et al., 2025). These developments extend traditional motor learning by integrating real-time multimodal sensing and context-aware feedback, deepening the link between MR technology and theories of embodied cognition.

Notably, among the typical paradigms discussed in educational theory (Schunk and Greene, 2017), neither behaviorism nor cognitivism was directly referenced, nor were their characteristics identifiable in the implicit designs of the reviewed works. By contrast, constructivist elements were recurrently emphasized, even when not explicitly labeled. For instance, several articles applied inquiry-based or problem-oriented approaches, such as *CuriosityXR*, which fosters exploratory engagement in context-aware MR environments (Vaze et al., 2024), or MR-supported physics experiments that encourage students to discover principles through guided exploration (Laumann et al., 2024). These approaches align with inquiry-based learning, itself considered a constructivist methodology (Maaß and Artigue, 2013).

Taken together, these findings reveal that while explicit references to learning paradigms remain limited, simulation-based and experiential learning dominate the applied approaches, particularly in medical training and engineering. At the same time, recent developments demonstrate an increasing tendency toward embodied, exploratory, and inquiry-based methods that resonate with constructivist ideals, even when not explicitly stated. This trend suggests a growing consensus that MR technologies are particularly well suited to paradigms emphasizing interactivity, situated practice, and active learner engagement.

2.4.3 Fields of Education

All articles clearly stated the fields of education for their implemented MR learning systems. Not only are the fields very diverse, but some systems are designed to be field-dependent while others allow or are designed to be transferred in different fields of education.

Among the identified fields, medical education is the most common one. Of the 27 articles that focused on this topic, most of them support visualizing anatomy structure or transferring procedural knowledge with interactive instructions (Kim et al., 2025; Zhao et al., 2025). Nonetheless, these articles also show different approaches. For example, some have a more technological focus, such as comparing the impacts of different XR technologies (Serrano Vergel et al., 2020), while others focus on specific subdomains such as nursing education (Kim et al., 2020). Within this body of work, recent studies continue to broaden the scope and fidelity of clinical teaching with MR. For instance, remote bedside tutorials implemented with HoloLens 2 showed substantial gains in students' topic knowledge by broadcasting a tutor's point-of-view to a co-located class (Connolly et al., 2024). Other recent works cover the fields of vascular anastomosis simulation (Stoner et al., 2024), neuroanatomy seminars which students rated as more engaging than conventional sessions (Zhao et al., 2025), and nursing interventions in which AR smart glasses improved procedural skill scores and motivation in randomized or quasi-experimental designs (Kim et al., 2025). Among these articles, the design features are also heterogeneous, as some are single-user systems, while others have collaborative features such as telemonitoring (Rojas-Munõz et al., 2020).

In engineering and production, MR frequently supports interactive instructions, spatial reasoning, and design-for-manufacture. Extending earlier systems such as the digital fabrication system by Stemasov et al. (2023), more recent work uses MR to teach abstract engineering drawing with three-view projections, showing higher performance than desktop or mobile AR and improving interest and perceived intuitiveness (Yuan et al., 2024). Similarly, MR is leveraged to teach design for additive manufacturing and to scaffold complex design decisions in authentic workshop contexts (Melo et al., 2024). These studies showcase MR's value in connecting intangible representations (drawings, toolpaths) to the tangible workspace, providing effective visualization and scaffolding in the learning process.

Enhancing computer science education with MR has also been explored in existing literature. HCI researchers have introduced studies on collaborative robot programming (Radu et al., 2021a) and MR-based pair-programming with conversational avatars (Manfredi et al., 2023). Furthermore, other articles extend MR into AI education for "non-technical students" (Schulz et al., 2024). In an exploratory study with business students, cross-device XR activities (mobile AR, MR, and VR) were perceived to make abstract concepts (e.g., k-means clustering) more concrete (Schulz et al., 2024). These findings suggest an expanding niche for MR as a conceptual bridge in computer science education beyond traditional learning formats.

MR is also gaining attention in domains where sensorimotor timing and embodied interaction are central, such as in music education. Banquiero et al. (2024) introduced an MR piano tutor which improved students' performance over a popular desktop benchmark while maintaining high presence and low adverse effects. Also in the field of music education, a multimodal MR tool for drumming improved rhythmic accuracy compared to video-based practice, underscoring that first-person MR embodiment and feedback can support complex motor learning (Pinkl et al., 2024). These add to prior work on violin gesture imitation using HoloLens avatars in higher music education (Campo et al., 2023).

Apart from the specific fields of education, a few reviewed articles focus on providing educational tools that can be applied in various fields. For these articles, we classify them as learning support. Some of these articles focus on providing authoring tools for lecturers. For example, Rajaram and Nebeling (2022) designed an immersive authoring tool to help instructors create interactive instructional systems without programming requirements. Meanwhile, Faridan et al. (2023) support remote collaboration with multi-device interaction using a tablet with an MR headset to create a co-present guidance experience. These learning support tools are diverse examples of how learning experience can be optimized with MR technology. Furthermore, gaze-aware feedback in MR uses eye-tracking to help learners improve attention management during the learning process and retrieve missed content (Liu et al., 2025b). Other research explores context-aware mini-lessons with conversational AI to motivate inquiry and retention in self-directed learning (Vaze et al., 2024), as well as co-located MR classrooms at scale across courses (Chandran et al., 2024).

2.4.4 Design Features

The articles we reviewed propose different design features to support the learning experience. Based on the conceptual framework, we classified these features. During the screening process, we noticed that many MR learning systems presented in the articles are single-user systems, while others feature collaborative settings. Therefore, we first examine whether social interaction is present in the proposed systems. Out of the 80 articles, 18 include social interaction as a design feature. Some of these articles present a learning experience that involves tutoring or mentoring (Rojas-Munõz et al., 2020; Thanyadit et al., 2023), while others focus on a collaborative learning experience with peer students (Kim et al., 2022; Manfredi et al., 2023). For example, Connolly et al. (2024) showed how MR-enabled clinical tutorials can be delivered synchronously to groups of medical students, where tutors provide guidance in real time. As an example of peer collaboration, Radu et al. (2021a) investigated the learning experiences and

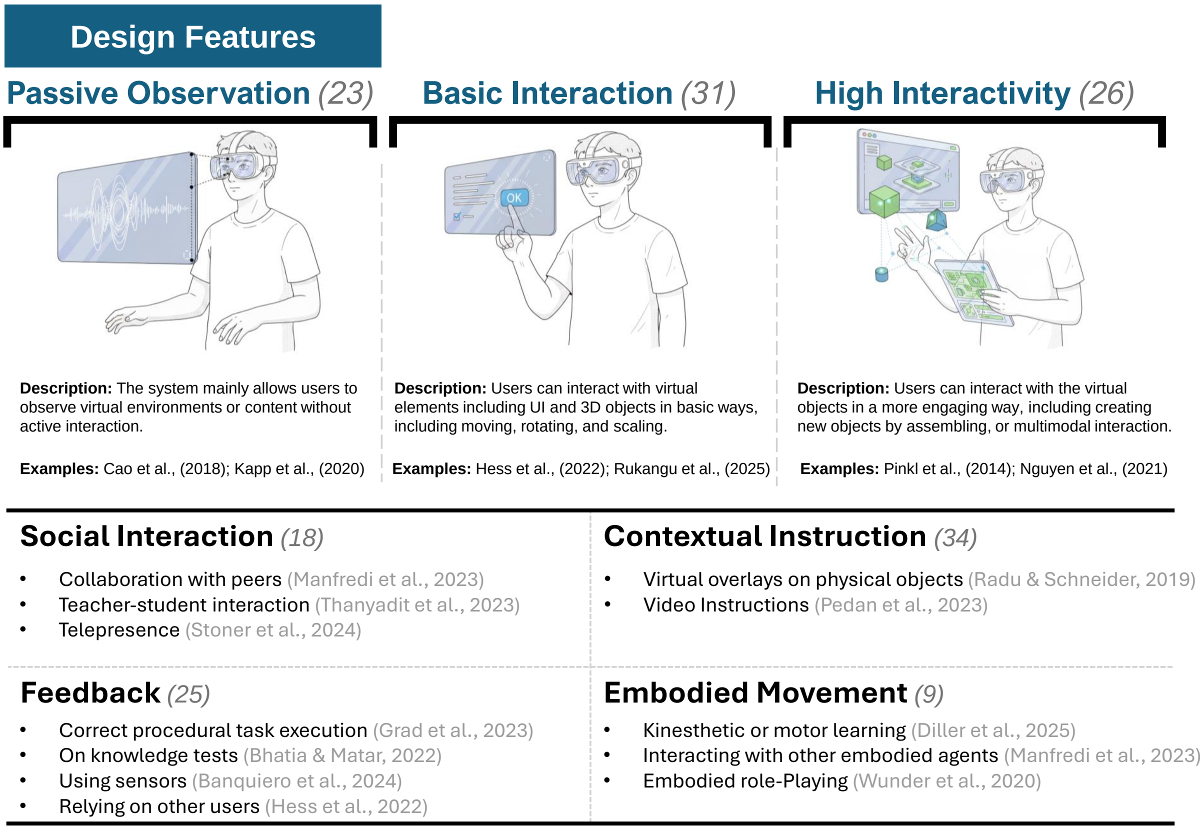

Figure 2.6: Different interaction levels and common design features identified in the reviewed articles

outcomes of a collaborative robot programming course supported with an MR learning system.

Interaction levels varied significantly across the reviewed systems. Twenty-three articles (29%) describe systems that offered only limited interactivity, restricting users to passive observation within the learning environment. The reasons for mainly providing a low level of interactivity in these systems are heterogeneous. For example, the guidance system designed by Cao et al. (2018) aimed to guide users in completing different steps in an experiment. Therefore, users only need basic information overlay on how to conduct the experiment and interact with the physical object, instead of interacting with virtual elements in the MR system. Similarly, the MR experimental support system by Kapp et al. (2020) was mainly designed to visualize the parameters of the laboratory devices. Therefore, they only integrate passive observation in the system to fulfill the design requirement. In other cases, the lack of interaction is due to technical limitation. For example, Yu et al. (2019) designed an MR system that raised situation awareness of users in a museum visit. However, they were only able to evaluate the concept with a Google Cardboard prototype for better scaling, which limited the interactivity of the MR learning system.

Basic interaction is characterized by presenting an MR system that allows simple operations to move, scale, and rotate virtual objects or to interact with the user interface in MR. Thirty-one articles identified in the review provide such basic interaction. For most articles, this seems sufficient for the learning task, as it can effectively support interactive instruction (Huang, 2019), multi-device interaction (Yang and Liao, 2014) and simulation-based learning (Hess et al., 2022). Nonetheless, it has also been suggested that a higher level of interaction may further improve the system. For example, Banky and Blicblau (2017) discuss the possibility of enhancing the experimental experience with haptic feedback to minimize the use of physical experiment equipment. Among the reviewed articles, 26 of them achieved high interactivity in the system. Here, articles such as the one by Nguyen et al. (2021) leverage multi-modal interaction, including gaze and speech interaction. Besides, more articles integrate high interactivity to support additional interactive features such as annotation (Elor et al., 2022; Zoghi et al., 2018). Recent studies go further by investigating the impact of different modalities. For example, Yuan et al. (2024) evaluated head-mounted MR against desktop and mobile modalities, demonstrating how gesture-based interaction improved learners' spatial reasoning.

The most common design feature included in the reviewed articles is contextual instruction. Nearly half of the articles (34 out of 80) have some types of context-aware instruction to guide the learning experience. The contextual instructions often provide information on how to execute certain steps in the learning experience and deliver procedural knowledge. Some MR systems also utilized spatial information detected by the devices to deliver information more accurately. For example, Radu and Schneider (2019) implemented an MR learning system that interplays with the physical environment, by placing instructions and tooltips around the inquiry-based learning objects. Other systems adopted a simplified approach without spatial awareness, such as using video instructions in the work of Peden et al. (2016). Furthermore, contextual instructions have been applied into many fields of education. For example, Gieser et al. (2024) embedded MR-based contextual feedback in medical simulation labs, while Su et al. (2025) integrated MR with mobile devices in industrial ergonomics training to deliver task-relevant guidance directly aligned with real-world assembly contexts.

Another design feature that has been implemented in MR learning systems is user feedback. In 25 of the reviewed articles, feedback was part of the system design. Out of these articles, 20 of them provide feedback on whether the expected operation has been performed correctly by participants. For example, Grad et al. (2023) compared the performed anatomical actions with the predefined baseline to give feedback to users. Other feedback methods include integrating quiz questions to test users' understanding (Bhatia and Matar, 2022). Moreover, some collaborative systems rely on the tutor or the facilitator to give feedback to students, instead of generating feedback in the system (Hess et al., 2022). Recent works, however, increasingly rely on automated and adaptive feedback. For instance, Diller et al. (2025) introduced an MR framework providing omnipresent in-situ feedback for motor skills acquisition, showing measurable improvements in procedural learning. Similarly, Banquiero et al. (2024) presented a video see-through MR system that uses real-time visual feedback to increase task accuracy in complex environments. Haptic feedback using smart gloves has also been implemented by Gieser et al. (2024).

Other design features identified in the articles include agents as learning assistants (one article), embodied or motor learning (nine articles), and role-playing (nine articles). Some design features that are commonly found in Massive Open Online Courses (MOOCs) and gamified systems such as rewards and achievement are not found in the MR systems we reviewed.

2.4.5 Research Methods and Study Design

For the overarching research method, we excluded articles without an empirical evaluation. Therefore, all articles are classified to be one of the following three types of research: qualitative, quantitative, and mixed-methods research. Almost half of the articles (39 out of 80) adopted a quantitative approach to evaluate their systems, while eight articles opted for a qualitative evaluation and 33 followed a mixed approach.

For the data collection, interviews have been commonly used to collect qualitative data (20 articles). Apart from interviews, researchers have also used other methods to collect qualitative data for the evaluation. For example, in the work of Banky and Blicblau (2017), they recorded the session and coded the activities performed by users of MR systems and analyzed it with the Kikan-Shido framework (O'Keefe et al., 2006). For the quantitative and mixed-method research, most articles (62 out of 80) used surveys to collect data from participants. The survey items were designed to measure different outcomes, which will be discussed in the following paragraphs. Other data sources include biosignal data, e.g. eye-tracking and facial expression data (Nakazawa et al., 2023). For example, Liu et al. (2025b) relied on gaze-tracking logs to evaluate attention allocation. Furthermore, 15 articles used interaction and log data, including task completion time, to measure different outcomes such as efficiency (Tang et al., 2020). For instance, Yuan et al. (2024) compared multimodal interaction logs across devices to investigate the acquisition of skills in learning engineering drawing. These approaches illustrate a growing methodological sophistication in analyzing learners' behavioral traces in MR environments.

In most studies (71 out of 80), the user evaluation has been conducted in controlled experiments, and in nine articles the authors evaluated their systems in the wild. The sample size of the studies also varies, with a mean sample size of 52.79 ($SD = 58.44$). In the articles, different dependent variables are measured to evaluate the learning system. A commonly measured outcome is usability, with 33 articles using different techniques to measure the usability of their proposed systems. Some including Magyar et al. (2020) used established measurement tools such as the system usability scale (SUS) to measure the usability, while others included a few questions in the survey to get an overall impression of the usability. Since MR systems have been researched to demonstrate the advantages of increasing motivation (Khan et al., 2019), supporting self-efficacy (Essmiller et al., 2020), reducing cognitive load (Mohammadhossein et al., 2022), these cognitive constructs have also been measured in a few studies. However, the total number of articles that directly measured these cognitive constructs remains relatively low: six articles measured cognitive load, six articles measured motivation, and seven articles measured self-efficacy. Similarly, even though engagement and immersion have been highlighted as a key factor in improving learning outcomes for MR learning systems (Dengel and Mägdefrau, 2018), only a few studies measured the engagement (six articles), presence (eight articles), and immersion (one article) level in the evaluation. For systems that involve collaborative learning experience, metrics including class dynamic, and collaborative contribution are also measured to evaluate how effective the MR system helps foster an efficient collaboration between users (Chandran et al., 2024; Kim et al., 2022).

Another commonly measured outcome is performance and knowledge. With 18 articles investigating the user performance and 20 measuring the knowledge acquisition or understanding. The constructs of performance, knowledge and understanding are defined and measured differently depending on the fields of education. For example, Campo et al. (2023) measured the performance of the violinists by recording the motion of playing the violin and comparing it with a standardized baseline. Similarly, Diller et al. (2025) used motion-tracking to quantify motor learning progression. In the study of Stemasov et al. (2023), performance was measured by the time and success rate of a personal fabrication task. Meanwhile, Gieser et al. (2024)

followed a similar approach by measuring procedural accuracy using interaction logs collected with body sensors. Meanwhile, knowledge and understanding are commonly measured with knowledge tests and quizzes in surveys. In total, 15 articles implemented a knowledge test to analyze users' learning performance, and ten of them integrated the knowledge test into the survey. For example, Su et al. (2025) used an established knowledge acquisition test to measure whether their training system improved participants' awareness on neck-related health conditions. These examples illustrate a trend toward domain-specific, fine-grained performance and knowledge measures.

2.5 Discussion

2.5.1 Interpretation of Results

The results indicate that the application of MR learning systems in higher education is diverse across several dimensions, including technological choices, fields of education, system design features, and research methods. This multidimensional diversity highlights the flexibility of MR as a learning technology, but it also complicates efforts to draw consistent conclusions about its pedagogical effectiveness.

In terms of technology, optical see-through devices are more commonly adopted than video see-through systems. Their advantage lies in the ability to perceive the physical surroundings directly through a transparent display, whereas video see-through systems rely on video streams where latency and image quality may negatively affect the learning experience. At the same time, several studies have experimented with cross-device setups or developed systems compatible with different MR hardware, demonstrating the adaptability of these technologies to varied educational settings. However, the implications of these technological choices for actual learning outcomes remain underexplored.

For the fields of education, the articles identified in the review demonstrate the potential of MR in various fields. Apart from STEM education and medical education, which commonly adopt immersive technologies, disciplines such as music education have also been explored integrating MR systems in teaching and learning activities. In addition, some articles reported on the use of MR authoring tools and support mechanisms that operate independently of specific fields of education, further emphasizing the field-independent potential of MR in higher education.

Taken together, these examples illustrate that the use of MR is not confined to traditionally technology-oriented disciplines but is increasingly relevant across a broad range of fields.

Moreover, the design features implemented in the reviewed articles are diverse as well. While some systems focus on delivering contextual instructions and allowing users to passively observe the training procedure, other systems are more immersive and strive for a higher level of interaction. A recurring design element is the provision of feedback and contextual instructions, but the effectiveness of these features in shaping learning outcomes is often overlooked in empirical evaluations. Compared to other digital learning formats, MR learning systems in higher education rarely incorporate learning assistants or gamified learning elements. This restrained approach may reflect a desire to avoid cognitive overload or distraction, yet it also suggests an underutilization of potentially effective design features. In general, through the reviewed articles, there is a lack of systematic investigation into how specific design features influence learning outcomes.

Finally, considerable variation exists in evaluation practices. Although all reviewed articles included empirical evaluation, the methodological rigor differed substantially. Usability was a common focus, but some measured it only through a single survey question, while others relied on standardized instruments. Similar inconsistencies were found in qualitative evaluations: while some articles employed a thematic analysis approach, others appeared to use the data only to form a general impression of users' opinions. A more comprehensive and standardized toolkit for evaluating MR learning systems could strengthen study design and enhance methodological rigor in future research.

2.5.2 Implications for Research and Practice

The Need for Advanced Learning Support as Design Feature

The state-of-the-art research in applying MR learning systems in higher education shows a variety of systems in terms of design features. First, the results of the review show various levels of interactivity among the MR systems. There are many systems with a high level of interaction, allowing users to immerse themselves in simulation-based learning (Wunder et al., 2020), real-time collaboration (Kim et al., 2022), human-robot interaction (Radu et al., 2021a). However, this doesn't suggest that the systems with a lower level of interactivity are inferior. Sometimes the systems offer only limited interactivity and even only allow passive observation. While this to some extent doesn't exploit the unique advantages of MR technology to its maximum, it is sometimes intended. For instance, this would allow more flexibility to transfer the design and the system to other platforms, including mobile AR or VR, where the interactivity level can be limited due to the lack of sensors, as many articles including Dafoulas et al. (2019) propose a multi-device system. In this case, it is intended to limit the interactivity to make sure the system is compatible with different devices and technologies. Therefore, the level of interactivity should be decided based on the user requirements and the use case scenarios.

Furthermore, most design features included in the systems still focus on the level of interactivity and immersion. More effective and realistic visualization is the focus of articles across different fields. For example, the system implemented by Janeras et al. (2022) mainly supports the learners with a 3D visualization of the geo-information. During the learning process, guidance and adaptive support are often missing in such systems, the goal of improving the learning outcome relies solely on the benefits of the visualization. In many cases, visualization is beneficial to the increase of learning outcomes. However, the overall focus on the visualization might create a tunnel vision on utilizing MR for better learning support.

Among the reviewed articles, only a few integrated support mechanisms for students to better navigate through the learning experience. For example, providing feedback on the learning progress and outcome is regarded as an effective way to achieve self-regulation and reflection (Johnson and Davies, 2014). In our review, less than half of the articles provided an MR system that gives feedback to users. These systems mostly focus on checking if the task and procedure executed by the user match the predefined baseline. For example, Grad et al. (2023) examines if a dental operation of filling a tooth is performed well by comparing the performed result in MR with a 3D printed model. Other systems rely on the intervention of a human facilitator to give feedback. Therefore, feedback as learning support is not fully implemented in MR systems for higher education.

Apart from feedback, the potential of other learning support techniques has also been largely ignored. For example, note-taking and annotation techniques are only supported in three of the identified articles. However, these articles also did not evaluate the efficacy of the introduced learning support mechanism. Future research can investigate additional design features to provide more effective learning support. One method is to seek inspiration from other learning platforms, such as MOOCs. Research in MOOCs has revealed several learning elements as design features that improve learning outcomes. For example, Julia et al. (2021) categorized various learning elements in student-student interaction, student-content interaction, and student-teacher interaction dimensions. These dimensions open up the design possibilities of learning support techniques, including online quizzes, automated feedback, discussion forums. While not all design features in MOOCs may be suitable to transfer to MR learning systems, it is worth considering as references by researchers to provide better learning support and design features in MR learning systems.

Effective Evaluation Methods for Learning Outcomes

All articles included in the review have evaluated their systems with an empirical study. The study design and research methods also show heterogeneity. While many articles motivate the introduction of MR systems by improving learning motivation and engagement of the learners, only a few have systematically evaluated whether the intervention indeed helps to improve these outcomes for learners. Most studies only evaluate these constructs with one or two survey items instead of resorting to an established measurement tool. For example, even though improving motivation has been mentioned as a huge benefit of using MR learning systems, only three studies measured the motivation level of the participants. Only in the article of George et al. (2023), the authors used an established questionnaire of intrinsic motivation inventory (IMI) to measure the motivation level of learners (Ryan et al., 1983). A similar situation can be found for other cognitive constructs such as engagement and confidence. Similarly, immersion has been mentioned as one of the benefits of immersive learning experiences, such as using MR learning systems. Nonetheless, only the article of George et al. (2023) measured the immersion level of participants in their study. The evaluation problem has been previously highlighted by Nebeling et al. (2020), as they pointed out, the lack of effective evaluation methods dedicated to MR systems. Therefore, future research can be devoted to better understanding the effects and outcome of MR learning systems on various cognitive constructs with better evaluation methods, including leveraging biosignal data with different sensors.

Furthermore, some articles in the review focus on comparing the use of an MR system with other digital learning formats, such as VR and mobile AR (Kapp et al., 2020). These studies offer valuable insights into the impact of the choice of technology on the learning outcome. Nonetheless, the results of these studies might not be applicable in other fields of education. Therefore, aiming to better understand the unique advantages and characteristics of all XR technologies, more comparative studies and evaluation are needed. Specifically, comparing mobile AR and MR can be insightful for research and real-world applications, as mobile AR offers a more affordable and flexible learning option, while MR focuses on better fidelity and interactivity. Moreover, comparing VR and MR can also reveal design implications, as VR has the potential to offer a more immersive experience, while MR allows the perception of the physical environment. Understanding the nuances of the technology can help educators decide which technology to use for which specific fields and educational tasks.

Lastly, the goal of MR learning systems is first and foremost to improve the learning outcome of learners in the higher education context. The reviewed articles show different approaches for the evaluation of different variables, including qualitative, quantitative, and mixed-methods. However, specifically for the learning outcome, more articles use quantitative methods with surveys or knowledge tests. Here, only a few studies use standardized tests in educational research to test the performance or knowledge retention to measure the learning outcome, e.g., the use of Individual Procedure Score (IPS) for performance measurement in Rojas-Munõz et al. (2020). Instead, most studies design their own knowledge tests with a few relevant questions. Therefore, the results may suffer from validity concerns. Even for those with a clear evaluation strategy for learning outcomes, the evaluation is often conducted within a short time. Nonetheless, the learning outcome, especially in higher education, is often observed through long-term studies. Yet, longitudinal studies are missing in the identified studies. Furthermore, the sample size of different studies varies as well, as can be seen from the relatively great standard deviation regarding the sample size. It seems that a consensus on an appropriate sample size for evaluating MR learning systems has not been reached so far, which can be further explored in future research.

Explicit Integration of Learning Paradigms and Theories in MR Learning Systems

Investigating which learning paradigms and theories have been applied in MR learning systems is one of the research questions of our paper (RQ2). Yet, most articles don't explicitly specify their adopted learning paradigms, and the design decisions are usually not supported by a specific learning theory. Therefore, we suggest that the established learning paradigms are not well transferred into the practice of designing MR learning systems for higher education. Compared to other educational technology research, MR learning systems might benefit from adopting established learning theories and paradigms to guide their design decisions to ensure a better learning outcome. This opens up further interdisciplinary collaboration opportunities between researchers from the education community with HCI as well as IS researchers.

Even though many articles don't clearly outline a learning theory or paradigm for their research. It is also implied through the system design that an exploratory or inquiry-based approach has been adopted. For those works, concrete learning theories may guide the design and evaluation of these systems more effectively. Instead of assuming exploratory learning is most suitable for MR learning systems, researchers should also be aware of the risk that an exploratory design without efficient guidance and support can deteriorate the learning experience (Mayer, 2004). A good balance between exploration and instruction needs to be achieved in the system design and evaluated to make sure that the user doesn't get lost during the learning experience. While the evaluation of contextual instructions has been missing in the identified articles in this review, the effort of integrating contextual instruction has been seen in some articles. This attempt to strike a balance between encouraging active learning and providing users with sufficient guidance. This strategy has been highlighted in educational research, that pure exploratory learning may not work effectively as intended (Mayer, 2004). Furthermore, MR is a novel device for many users, which may introduce usability challenges. In this case, sufficient instruction as guidance is even more necessary. Nevertheless, it's not yet confirmed that a perfect balance between active learning and instruction has been guaranteed in state-of-the-art research. More than half of the articles don't explicitly design contextual instructions in their systems, and even the articles that integrated the instructions rarely measured the efficacy of their integrated instruction with a dedicated measurement tool.

Besides the commonly adopted self-directed learning approach that encourages users to explore the learning material independently, some articles allow different learning approaches such as collaboration, simulation-based learning, tutoring, etc. This shows the awareness of tailoring the technology to specific learning tasks and fields of education. Future research can further explore other learning paradigms such as behaviorism or cognitivism, which have also been identified as effective learning strategies in various fields (Schunk and Greene, 2017).

In the review process, we identified a total of 18 different fields of education for applying MR learning systems in higher education. This shows the potential of MR systems to adapt to different fields of education and support various topics. Nevertheless, medical training remains the most commonly researched field of education and the advantages of MR have been demonstrated well in these studies. In the field of medical training and laboratory training, established learning methods have been transferred well in MR learning systems, including operation training, simulation-based learning, etc. Nonetheless, in other fields, traditional learning techniques are often overlooked as design features, including note-taking, quizzes, and discussion opportunities. Future research can explore the feasibility of transferring existing effective methods to MR learning systems.

Designing Biosignal-Adaptive MR Learning Experiences

As the definition of MR suggests, MR learning systems provide more advanced and interactive experiences than traditional AR on mobile phones or tablets. This is largely due to the wide range of built-in sensors on MR headsets, including eye-tracking, motion-tracking, and hand-tracking. While some studies have begun to explore the use of these sensors to capture biosignals in educational contexts (Nguyen et al., 2021), most work continues to rely on them primarily for interaction purposes. A recent survey on eye-tracking in MR confirms this pattern, showing that the majority of studies use eye-tracking primarily for gaze-based interaction (Plopski et al., 2022). We argue that these sensing capabilities should not be treated as peripheral: beyond enabling additional interaction modalities, they offer rich opportunities for learning analytics and for designing adaptive learning experiences.

Evidence from desktop-based e-learning shows that biosignal data can be used to create adaptive systems that respond dynamically to learners' cognitive states (Hu and Kuo, 2017; Kennel, 2022). However, such systems typically require additional external sensors, making them difficult to scale. By contrast, MR headsets already integrate many of these sensing modalities, lowering the technical barrier to implementing biosignal-adaptive approaches. Beyond immersive learning, researchers have already demonstrated the feasibility of biosignal-adaptive systems that adjust interfaces in response to users' cognitive states (Doswell and Skinner, 2014; Schultz and Maedche, 2023). These examples suggest that MR learning systems are uniquely positioned to combine immersive interaction with continuous biosignal monitoring in ways that traditional desktop or mobile learning systems cannot.

The potential of biosignal-adaptive MR becomes clearer when considered alongside the design features identified in our review. Many existing MR learning systems emphasize contextual instructions and feedback, yet their effectiveness in shaping outcomes is often under-evaluated. Similarly, levels of interactivity vary widely: while some systems allow only passive observation, others integrate high interactivity through multimodal input such as gaze and speech (Nguyen et al., 2021; Yuan et al., 2024). In both cases, biosignal data could serve as an additional layer of adaptation. For example, attention patterns captured through gaze tracking could be used to tailor contextual instructions more precisely, while indicators of cognitive load could inform when feedback should be delivered or withheld Liu et al. (2025b). This is particularly relevant for embodied learning tasks, which are common in MR environments and benefit from real-time responsiveness to learners' physical and cognitive states. The optimism toward biosignal-adaptive MR does not imply that other design features and evaluation methods lack value. Surveys, knowledge tests, and qualitative interviews remain indispensable for capturing learners' perceptions and outcomes across contexts. However, these methods typically provide retrospective or aggregated data, whereas biosignals enable real-time, in-situ understanding of users during the learning process. In this sense, biosignal-adaptive approaches complement rather than replace traditional designs, offering a path toward more responsive and personalized MR learning experiences.

2.5.3 Limitation

This systematic review has several limitations. First, our focus is on MR learning systems in higher education, and many of the identified articles explicitly highlight this target group in their design rationales. However, some reviewed articles did not clearly state that their systems were designed for higher education. These articles were nevertheless included, as their evaluations involved only university students. It remains unclear, though, whether the exclusive use of university students reflects a deliberate focus on higher education or merely a coincidental demographic choice without a dedicated target context. This ambiguity may introduce potential bias into the results.

Second, although our search strategy did not explicitly filter by language, the use of an English-language search string means that non-English articles would only be found if they included an English title and abstract. Consequently, relevant research published entirely in other languages was not captured by our search strategy. Nonetheless, we successfully retrieved 61 non-English articles with titles and abstracts translated in English. These articles were screened in the same process with other articles.

Furthermore, the results show a variety of different fields of education. However, due to the focus on higher education, we couldn't dive into each of these fields and discuss the characteristics of each field of education in detail. For example, each field has its unique subcategories, which may require different design solutions. This can only be explored with a dedicated literature review on the specific field. Therefore, future research can further explore MR learning systems in a specific field of education and compare the results with the findings of our work.

Lastly, our review follows mainly a qualitative approach, with descriptive analysis and a conceptual framework illustrating the result. Nonetheless, the review can be further extended with a quantitative approach in a meta-analysis. For example, we summarized measured outcomes and the study design of the articles. However, we didn't calculate and report the effect size of the measured outcomes among the articles. This insight can benefit further research and provide additional guidelines for designing MR learning systems.

2.6 Conclusion

This paper presents a systematic literature review of MR learning systems in the higher education sector. Our findings indicate that MR learning systems have been adopted across different fields of education. Through implementing different design features and learning elements, existing studies emphasize the integration of interactive and immersive experiences that enhance learning outcomes in higher education.

Despite the promising applications of MR in higher education, our review identified several areas that warrant further investigation. A primary need is the development of more comprehensive evaluation methods to assess the impact of MR systems on key cognitive and affective constructs, such as motivation, engagement, and immersion. Additionally, an area largely overlooked in current research is the incorporation of advanced cognitive and metacognitive support mechanisms, e.g., adaptive feedback, into the design of MR systems. Furthermore, the integration of established learning paradigms and theories can further guide the design and implementation of MR learning systems to deliver effective learning experiences.

Future research can also explore leveraging the integrated biosignal sensors in MR headsets to create user-adaptive learning environments. This approach holds potential for providing personalized learning support and enhancing the overall efficacy of MR-based instruction. Moreover, comparative studies between MR and other immersive technologies, such as VR and mobile AR, are needed to elucidate the distinct advantages and optimal use cases for each modality. In conclusion, while MR technology holds significant potential for transforming higher education, future research can address the identified gaps and challenges to fully realize its advantages.

Supplementary materials

The supplementary material of this systematic literature review contains the following data: (a) An Excel spreadsheet with a table of the identified existing reviews from the scoping analysis, a concept matrix serving as a coding table for reviewed articles, and the search strings used across all databases during the search process. (b) Data analysis and visualization script with JASP. We gratefully acknowledge Brudy et al. (2019), whose clear and effective visualization style provided the inspiration for our figures. The data is openly accessible here: `https://osf.io/duvqs/?view_only=6492fd17b9cb4af9a72c7d8681233438`

3 AF-Mix: A Gaze-Aware Learning System with Attention Feedback in Mixed Reality (Study II)

Abstract: Mixed Reality (MR) has demonstrated its potential in various learning contexts. MR-based learning environments empower users to actively explore learning content visualized in multiple formats, such as 3D models, videos, and images. Nonetheless, the sophisticated visualizations in MR learning environments may result in potential visual overload, posing a challenge for users in efficiently allocating their attention. In this paper, we present AF-Mix, a learning support system that leverages eye tracking sensors in Microsoft HoloLens 2 to offer attention feedback for learners. Aiming to design AF-Mix, we conducted a participatory design study and integrated the attention feedback into our system, following users' needs and suggestions. Furthermore, we evaluated AF-Mix in an evaluation study (n = 22) following a quantitative analysis of users' visual behavior, as well as a qualitative analysis of interview transcripts. Our findings show that providing feedback to support the learning process can be achieved effectively with eye tracking. In specific, attention feedback assists learners in retrieving previously missed information and encourages learners to reallocate their attention in the review process. Moreover, providing personalized feedback based on previous attention allocation is more effective in supporting users than a self-review approach without gaze-aware assistance in MR. Such feedback facilitates users in managing their limited attentional resources better and supports the reflection of their learning journey more effectively.

Keywords: Mixed reality, Eye-tracking, Attention feedback, Learning

Published as: Liu, S., Toreini, P., & Maedche, A. (2025). AF-Mix: A gaze-aware learning system with attention feedback in mixed reality. International Journal of Human-Computer Studies, 198, 103467. https://doi.org/10.1016/j.ijhcs.2025.103467

3.1 Introduction

Mixed Reality (MR) for learning has been intensively explored in recent years (Arici et al., 2019; Maas and Hughes, 2020). For example, it has been applied in medical education to help students better understand anatomy structure (Huang et al., 2018), in math education to create a playful learning scenario (Khan et al., 2018), or in simulation learning to present immersive environments (Ibáñez et al., 2016). In general, MR learning systems come with several advantages, including their potential to visually present and engage learners with abstract concepts to promote a more comprehensive understanding (Khan et al., 2018; Radu and Schneider, 2019). Consequently, MR learning environments can better motivate learners and improve their overall learning performance (Khan et al., 2019).

Nonetheless, MR learning environments also come with significant limitations, including challenges for learners in attention management and overcoming cognitive overload (Knierim et al., 2018; Vovk et al., 2018). Meanwhile, most MR learning environments follow a constructivist approach, requiring active learning (Hanid et al., 2020), and such an approach also requires sufficient guidance and assistance (Mayer, 2004). Therefore, a need for more effective support for learners in MR environments has been recognized (Thoravi Kumaravel et al., 2019). Besides, users sometimes encounter visual overload and are challenged to manage their attention during the learning process appropriately (Knierim et al., 2018). In physical classrooms, this is typically the task of the teacher to structure the learning materials and provide feedback, which is essential if learners are expected to establish a comprehensive understanding of the learning topic (Quintana et al., 2004).

While some existing MR learning systems are designed for tutoring or collaborative learning with multiple users (Huang et al., 2021; Radu et al., 2021a), other systems expect users to learn independently without help from fellow learners or teachers in the MR environment (Akçayır and Akçayır, 2017b; Mohammadhossein et al., 2022). Successful learning outcomes in such scenarios require effective self-regulated learning skills. In the framework of self-regulated learning, designing an effective learning system requires consideration and support of various cognitive and metacognitive activities, such as goal-setting, planning, monitoring, etc. (Panadero, 2017). Among these activities, providing feedback to students is crucial for cognitive support during the learning process (Butler and Winne, 1995). While feedback can be achieved in various ways, we focus specifically on providing attention feedback for MR learning systems in our study. This is due to the reported challenges of attention management in the MR learning environment. Embedding attention feedback to support learning in MR environments to the best of our knowledge has been not investigated so far.

We provide attention feedback in MR learning systems to support users in managing their attention allocation and promote a more efficient review process of their learning outcomes. Therefore, we name our system *AF-Mix*. Our human-centered design process focused on collecting user needs and potential design solutions in participatory workshops. After implementing the system in Unity and deploying it on HoloLens 2, we evaluated the impact of attention feedback on the review process with a user evaluation, analyzing eye tracking and interview data collected from 22 participants in a between-subject experiment. The data analysis, using a Bayesian t-test framework, suggests that *AF-Mix* enhanced awareness by helping learners retrieve previously missed information while having no significant impact on learners' processing time during the review. Furthermore, participants perceived *AF-Mix* as supportive and effective for learning in MR. Participants also reported that such a system can benefit learners of different learning styles and encourage learners to optimize their review strategy.

Our main contribution in this paper is a new approach for presenting attention feedback using eye tracking in MR learning environments. The results of our user evaluation suggest that *AF-Mix* supported participants in obtaining an overview of their learning progress and successfully helped them to better allocate their attention in the review process. By conducting user evaluation and synthesizing user feedback, we contribute to the HCI community with design recommendations and guidelines for gaze-aware learning support in MR systems, as well as insights on supporting self-monitoring and self-reflection in MR learning environments.

3.2 Related Work

The term MR was originally coined to indicate a spectrum of virtual environments, including Augmented Reality (AR) (Milgram and Fumio, 1994). However, researchers have interpreted the term differently over the past decade (Speicher et al., 2019). One attempt is to define MR as "strong AR", as reported by Speicher et al. (2019), and it has been adopted in works of Yue et al. (2017) and Maas and Hughes (2020). According to this definition, MR environments should create an enhanced user experience compared to their mobile-AR counterparts, constructing a blended and interactive environment of physical surroundings and virtual elements, not just static virtual overlays on top of the physical environment. In this paper, MR specifically refers to using head-mounted displays (HMD) to construct immersive AR environments with extended features such as eye tracking.

Using MR for education and learning has been explored in many fields, such as medical education (Bianchi et al., 2020; Moro et al., 2021), science education (Arici et al., 2019; Radu and Schneider, 2019), language learning (Shao et al., 2020), K-12 education (Maas and Hughes, 2020), collaborative programming (Radu et al., 2021a), etc. Research has shown that using MR for educational purposes has the potential to improve students' motivation and learning outcomes by making abstract ideas and concepts more tangible (Khan et al., 2019). With MR, students can interact with the virtual learning material naturally, as they would do when learning with physical material. For example, Radu and Schneider (2019) implemented a collaborative learning environment with HoloLens 1, where abstract concepts in physics are visualized in MR (e.g., magnetic fields). Students can easily explore the "dynamic nature of relationships between important variables" in such learning environments (Radu and Schneider, 2019). The importance of giving and receiving feedback in such MR environments is also addressed in their work, specifically the feedback given by other collaborators during the learning process. As another example, Vazquez et al. (2017) proposed language learning by using MR. Compared to traditional learning formats such as phrasebooks, serendipitous language learning is achieved in MR with extensive contexts. For instance, when users discover a physical object in the environment, the MR system not only displays the corresponding word for the object but also augments the physical object with additional contextual information by "embedding 3D models" or a movie clip to the object (Vazquez et al., 2017). Similar to the previous example, Hennerley et al. (2017) implemented an MR application that supports the understanding of abstract concepts in programming with 3D visualizations. Therefore, one major advantage of learning in MR environments is the following: MR systems support students in understanding intangible concepts by embedding 3D visualizations and various formats of learning material in the learning environment. Besides, MR devices motivate kinesthetic learning, a learning strategy that by definition requires physical activities to actively engage in the learning task (Iqbal et al., 2019), whereas other learning formats such as mobile-AR often hinder it (Oh et al., 2018; Radu and Schneider, 2019).

Even though MR systems have advantages over other learning formats in some contexts, researchers have pointed out the limitations of MR learning systems as well. Optimizing cognitive workload caused by information overload, addressing usability issues, and reducing motion sickness are challenges for designing MR learning environments (Vergel et al., 2020; Vovk et al., 2018). In response, some research has explored providing learning support in MR systems. For instance, Giraudeau et al. (2019) claim that intuitive interactions in MR can reduce cognitive overload during the learning process. Ibáñez et al. (2016) adopt scaffolding in AR simulation learning systems to guide users in exploratory learning. However, little research has been done to provide feedback that increases users' self-awareness and supports them in optimizing their review strategies. We believe that this research gap prevents learners from fully utilizing the advantages of MR learning environments.

Compared to traditional learning environments, MR systems often present learners with a self-directed and constructivist scenario, including ones developed by Radu and Schneider (2019) and Khan et al. (2018). Constructivism focuses on motivating learners to actively engage in learning activities and construct their understanding of the knowledge proactively (Schunk, 2012). Examples of constructivist learning strategies include inquiry-based learning, experiential learning, kinesthetic learning, etc. These learning strategies have been commonly adopted for designing MR or AR learning environments (Huang et al., 2016; Radu et al., 2021a; Radu and Schneider, 2019). Meanwhile, education research suggests the necessity of providing feedback in constructivist learning environments to promote better self-regulation (Kirschner et al., 2006). Such feedback can increase users' self-awareness and support in improving their learning performance. Receiving feedback and adjusting learning strategies are also seen as an integral part of self-regulated learning (Butler and Winne, 1995).

Nonetheless, most MR learning systems have so far focused on creating innovative interaction and experience during the learning journey, while integrating effective self-awareness feedback for learning in MR has been overlooked. This research gap has been recognized by Thoravi Kumaravel et al. (2019), as they criticize the lack of feedback in most MR learning environments. As a response, they integrate learning feedback in their system by implementing "bidirectional mixed-reality telepresence" tutoring (Thoravi Kumaravel et al., 2019), allowing tutors to give asynchronous feedback on learners' performance. Furthermore, another study conducted by Shao et al. (2020) focuses on American sign language education, which also emphasizes the importance of providing feedback during the learning process, especially in the context of motion learning. The feedback discussed in these studies focuses on addressing whether a user action is correctly performed or not, instead of examining users' attention distribution and promoting self-awareness in the learning process. Moreover, these systems are typically designed for multi-user learning environments. We argue that there is a research gap in providing feedback on users' attention during the learning process in self-directed and single-user MR learning environments.

Attention plays an important role in learning, and managing attention in constructivist learning environments is crucial for learning success (Schunk, 2012). Existing studies have leveraged eye tracking to provide attention feedback in non-MR learning contexts (D'Mello et al., 2012; Sharma et al., 2016). In a review of studies that used eye tracking for learning support, Lai et al. (2013) suggest that eye tracking is a "promising channel for educational researchers to connect learning outcomes to cognitive processes". To the best of our knowledge, providing attention feedback has not been well-studied in MR settings. Therefore, we aim to fill this research gap by designing a gaze-aware system that presents attention feedback to users in MR learning environments.

3.3 Method

Our study aimed to introduce attention feedback into MR learning systems. For this goal, we followed an iterative research process. Overall, we organized our research process in three phases. Each phase contributes to the development and evaluation of the attention feedback system, denoted as *AF-Mix*. The first phase focused on unveiling user needs and establishing design goals. We conducted a participatory design study and organized three design workshops. The goal of the workshops was to elicit ideas and potential design solutions for *AF-Mix*.

The second phase revolved around the practical implementation of *AF-Mix* based on design solutions derived from the preceding phase. This resulted in the extension of an existing MR learning system with gaze-aware learning support, i.e. attention feedback. The feedback included an analytical overview of learners' attention distribution (Figure 3.6-1) and color-coded flashcards (Figure 3.6-3) reflecting learners' prior attention allocation.

The final phase comprised a comprehensive evaluation of *AF-Mix* through a dedicated user study. The primary objective was to investigate the impact of attention feedback on learning strategy and outcomes, particularly its influence on the review process during the learning experience. Employing mixed methods, we analyzed eye-tracking data and interview transcripts.

For the evaluation, we recruited 22 participants and divided them into two groups: Participants in the treatment group received attention feedback after exploring the learning material, while participants in the control group didn't.

For the data analysis, we adopt the Bayesian t-test framework for the quantitative eye tracking data (Rouder et al., 2009). Specifically, we aim to determine whether the attention feedback enhances learners' awareness of content previously overlooked during the learning process. Additionally, we explore whether the introduction of attention feedback influences the processing time during the subsequent review phase. Complementing our quantitative analysis, we also conducted interviews to gain insights into the user experience. The interviews had been transcribed and coded by two independent coders in an opening coding process before we conducted a thematic analysis based on the guidelines proposed by Braun and Clarke (2006a).

3.4 Participatory Design

Aiming to understand user needs and define our design goals, we conducted a participatory design study. For the participatory design, we organized three design workshops and recruited 15 students (6 females) to generate ideas and design solutions. Students were recruited voluntarily from a Human-Computer Interaction (HCI) lecture at the local university. Participation in the workshops did not influence the outcomes or grades for the lecture. The recruitment process received approval through an Institutional Review Board (IRB) review process. The inclusion and exclusion criteria for the workshops include the following:

Inclusion Criteria	**Exclusion Criteria**
Willing to test and provide feedback in group discussions	Severe motion sickness, epilepsy, or other medical conditions
Enrolled as a full-time university student	Limited English proficiency
Proficient in English (CEFR B2)	Unable to participate due to constraints, e.g. COVID-19

Table 3.1: Inclusion and exclusion criteria for workshop participants

We used the 6-3-5 brainwriting method to collect ideas and sketches from participants. After the workshops, we summarized user needs and design goals.

Based on the previous work of Liu et al. (2022), we used their MR learning system for HCI education as the foundation of our participatory design. In their MR learning system, users can explore the established HCI concept of Model Human Processor (MHP) proposed by Card (1981), using multimodal interaction such as gesture and eye-gaze-based interaction. In the end, the system will present users with a review option, which allows users to rediscover previously missed learning content. Nonetheless, the system has not been systematically evaluated and the review option requires a more effective design. Apart from this system, we also presented various other MR learning systems to the participants, including HoloPatient[1], HoloAnatomy (Wish–Baratz et al., 2019). The goal is to reduce the priming effect of one particular MR learning system on participants during the design workshops.

The goal of our participatory design is to integrate effective attention feedback in such an MR learning system based on users' suggestions. For the participatory design study, 15 participants (6 females) were recruited voluntarily and divided evenly into three sessions. All participants were undergraduate students taking an HCI lecture and had a basic understanding of HCI-related learning content and MR technology. Each session lasted around one hour and was organized in a meeting room at the local university. Before the workshops, participants were first informed about the study procedures and gave their consent.

The workshops consist of three parts. First, participants received a brief introduction to MR technology and the HoloLens 2 device. They were also invited to use the device to get a better understanding of the gesture and eye-gaze-based interaction techniques with this hands-on experience. Next, we presented the participants with the challenge of MR learning environments by inviting them to use some existing MR learning systems. Then, we briefly described the basic idea of providing attention feedback as potential support for learning in MR.

Lastly, participants generated and exchanged ideas on how to design attention feedback using the 6-3-5 brainwriting method. The 6-3-5 brainwriting method is a classic ideation technique originally proposed for product innovation and later adopted in interdisciplinary research (Heslin, 2009; VanGundy, 1984). In HCI research, it has been used as an effective participatory method (Boy, 1997; Muller and Kuhn, 1993), and researchers have revealed its advantages for group elicitation and preventing the ideation process from being dominated by verbally dominant participants (Wilson, 2013). For our design workshop, we followed the guidelines proposed in prior research on participatory design (Hutt et al., 2021; Muller and Kuhn, 1993).

[1] HoloPatient is an MR learning system developed by GIGXR: https://www.gigxr.com/holopatient/

Figure 3.1: One session of the participatory design workshops

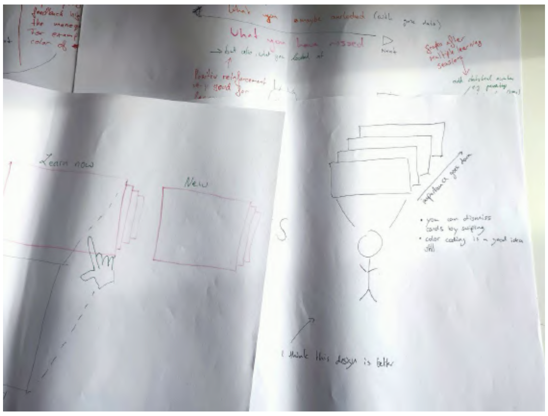

Figure 3.2: Brainwriting worksheets produced by participants

In our workshop, a brainwriting session typically lasted around 30 minutes. In this session, participants formulated their ideas by making sketches or writing them down on a worksheet. The session followed a repetitive process consisting of five rounds. In the first round, each participant contributed three ideas on his or her worksheet within five minutes. After the first round, the worksheets were passed around to the participants so that all participants could review others' ideas and engage in this additive process. The session finished when every participant had reviewed the worksheets of all other participants. After the brainwriting session, participants discussed the final results and suggested potential design solutions. Figure 3.1 shows participants exchanging design ideas using the brainwriting technique. After organizing three participatory design workshop sessions, the results were collected and synthesized from the worksheets.

3.4.1 Findings

We coded and clustered all ideas generated by the participants into three themes, representing the main user needs:

1. Attention feedback should be presented in a simple and straightforward way in the review process, without causing additional cognitive overload.
2. The review process should provide a different learning approach from the previous explorative learning process.
3. The review process including attention feedback should help learners to effectively process previously missed information.

Many ideas attempt to reduce users' cognitive overload. The first two themes emphasize this user need. Participants commonly recognized a difficulty users may face in MR environments: cognitive overload. After reviewing the learning system developed by Liu et al. (2022), participants assumed that the initial exploratory learning process requires a high mental workload. Therefore, the review process and attention feedback should feature a simple design, which doesn't require much effort. Besides, a clear distinction between the exploration and the review process should be clarified as affordance in the system design, assisting learners to adjust their learning strategies accordingly. For example, the initial learning process may allow users to explore learning material in an arbitrary sequence, making the learning process more autonomous and flexible. Thereafter, the review process can present a predefined learning path that guides users in covering content they might have missed previously.

Overall, the collected ideas suggest a unique design to present attention feedback in a simple format. As one participant wrote in her worksheet, attention feedback should "reduce users' mental workload with a simple interface" instead of further overwhelming users. Additionally, the third theme suggests that participants wanted more than just receiving information on their attention. Particularly, participants wished for additional support for self-reflection in the review process after receiving the attention feedback. One participant explained this during the discussion:

> In my opinion, the goal of getting feedback is to know more about your weakness in the process. If the system is already capable of telling me what I missed, is it possible to help me review and memorize what I missed? (...) Fixing your weakness is more important than finding it out. (P4)

Sketches (see Figure 3.2) produced by the participants also corroborate the clustered themes. In the design sketches produced by the participants, one idea attempts to address the user needs in all clustered themes. The idea is to present attention feedback and review options as a deck of virtual flashcards. The MR learning system can potentially display one flashcard at a time. Each flashcard should be color-coded based on users' previous attention. Users can go through the flashcards one by one to review their learning progress and the learning material at their own pace. Figure 3.2 shows this idea from the worksheets. This design was favored in the group discussion during the participatory design sessions. Participants from the workshop sessions proposed several advantages of this design. First, this design reminds users of familiar experiences in using a deck of physical flashcards, thus potentially reducing their cognitive workload. Moreover, participants suggested that the new design has the potential to help users reflect on their learning results and even assist them in fostering knowledge. In conclusion, the design of attention feedback for our system was finalized based on the results of the participatory design study: The feedback scene of our system is designed as a set of color-coded flashcards. Details of the design and implementation of the attention feedback are discussed in the next section.

3.5 Design and Development of *AF-Mix*

Our system *AF-Mix* integrates **A**ttention **F**eedback in **Mix**ed Reality. *AF-Mix* consists of three scenes: a) introduction; b) exploration; c) feedback. The introduction scene assists users in onboarding and presents essential information to help them understand the overall concept to be learned. The exploration scene allows users to explore the learning material with eye-gaze-based interaction and learn the concept at their own pace, The learning material is presented as content items, including 3D models, images, texts, and videos. Lastly, attention feedback is provided in the feedback scene based on users' previous gaze attention.

We selected the established HCI concept MHP as the learning content, based on the previous study of Liu et al. (2022). The concept is well-established and was initially proposed by Card (1981) to integrate Human Information Processing (HIP) knowledge in HCI education and research. The user requirements suggest an engaging and active learning experience. Users should be able to freely explore the learning content and activate content items based on their attention. Therefore, we arranged the content with multimedia material, including 3D models, videos, images, and texts as content items. Users can trigger the content items and explore the information with their gaze, and *AF-Mix* presents users with attention feedback based on the collected eye tracking data. Section 3.5.2 describes the user interaction in more detail.

Before implementing *AF-Mix*, we created low-fidelity prototypes and made design decisions. Figure 3.3 shows the feedback scene with the design of presenting attention feedback as color-coded flashcards, inspired by the idea that emerged from the participatory design study (See Figure 3.2). The final design of attention feedback in our system is the following: After exploring the content in the exploration scene, users would receive a summary of how many content items they have explored or missed. Next, users have the opportunity to review all items displayed as a deck of flashcards in a randomized order and go through them one after another.

Figure 3.3: Low-fidelity prototype of the feedback scene

Each flashcard is also color-coded to indicate the level of attention from the previous scene.

The following sections describe the system architecture and each scene with more details, including how users explore the learning content and how attention feedback is presented in *AF-Mix*.

3.5.1 System Architecture

For implementing *AF-Mix* as a HoloLens 2 application, we used Unity as our engine and imported the foundation package of Mixed Reality Toolkit (MRTK) 2.7.0 in Unity. In Unity, our workflow consists of two parts: a) creating 3D models, content items, and the user interface; b) compiling C# scripts to connect the frontend and backend. The first part was achieved with the open-source Computer-Aided Design (CAD) software Blender[2]. In Blender, five 3D models (human body, eye, ear, neuron cells, and brain) were created to represent the main components of the MHP: a human user, visual perception, auditory perception, motor system, cognition, and memory.

After importing the 3D models in Unity, the frontend and backend of *AF-Mix* were implemented. Prefabs from the MRTK library were used for the frontend interface, including prefabs of information panels that hold images, texts, and videos as content items as well as UI buttons. The backend collects eye tracking data when users interact with the items. Based on the collected data, the system categorizes the content items according to users' gaze duration and presents attention feedback accordingly. Additionally, the collected eye tracking data were also used for quantitative analysis in the evaluation study.

[2]https://www.blender.org/

3.5.2 Introduction and Exploration Scene

Attempting to avoid usability issues, we include a short introductory scene in the final system. The introduction scene provides interactive examples of how to use eye-gaze-based interaction to explore elements in MR.

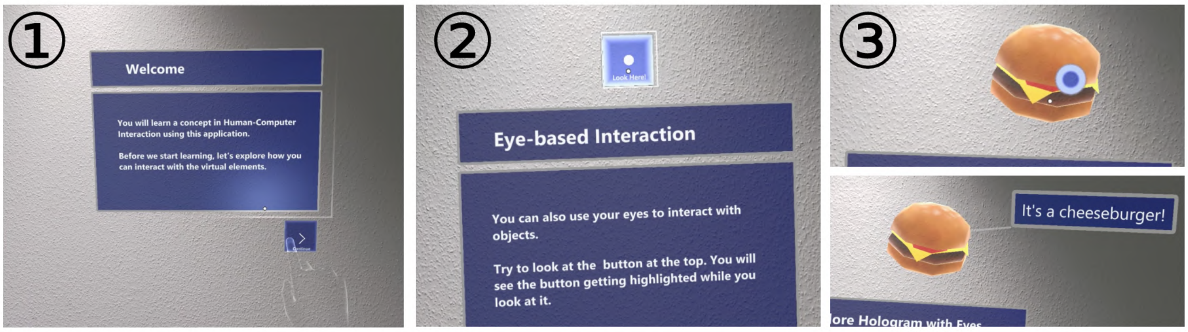

Figure 3.4: Guidance on using gesture (1) and eye-gaze-based interaction (2,3) in the introduction scene

In the introduction scene, *AF-Mix* introduces eye-gaze-based interaction to users (See Figure 3.4-2, Figure 3.4-3). At this stage, users are instructed to toggle a button with their gaze. If users successfully toggle the button by gazing at the button for more than two seconds (default setting for button prefabs in MRTK for HoloLens 2), they can proceed to the next page. On the next page, a 3D model of a cheeseburger will appear on top of the information panel. Here, the introduction guides users to look at the model and activate a tooltip before they proceed to the last page (See Figure 3.4-3). This interaction technique is used throughout the exploration scene. The last page of the introduction scene presents basic information on the MHP concept. It is designed to give users a general impression of what they will learn later and reduce their cognitive workload when they enter the exploration scene.

The exploration scene contains learning material users can explore, including 3D models and content items. Upon entering the scene, users encounter the main 3D model of a human user interacting with a laptop. Besides, a notification message is fixed above the 3D model, reminding users to employ the previously introduced eye-gaze-based interaction to explore the content (Figure 3.5-1). The self-directed learning process follows a drill-down approach, where users explore content arranged in three layers. Figure 3.5 illustrates the process of exploring the three layers of content. The main 3D model is the first layer of content: It is static and displayed throughout the scene. When users gaze at particular parts of the main model, including eyes, ears, the brain, and hands, the second layer will be activated. Each second layer comprises another 3D model (secondary model) and a short text explaining one of the following components: visual perception, auditory perception, motor system, cognition, and memory. When exploring the second layer, users can gaze at the secondary models and activate the third layer. The third layer contains content items that elaborate on sub-parts of the MHP concept in detail. For example, the auditory perceptual processor (ears) is part of the second layer, consisting of three sub-parts: inner, middle, and outer ear. Each sub-part includes two content items: one information panel with texts and another with either images or videos.

Throughout the exploration scene, a total of 20 content items are presented. As detailed in the system architecture, each item is associated with a script that logs whether users have fixated on the item and records the gaze duration for each item. This functionality is crucial for providing attention feedback in the subsequent scene. In the exploration scene, users can find a 'Finish Learning' button located in front of the main hologram, enabling them to proceed to the feedback scene when they decide to conclude their content exploration.

3.5.3 Feedback Scene

Attention feedback is presented as a deck of color-coded flashcards in the last scene in *AF-Mix*. After finishing the explorative learning in the previous scene, users are presented with a brief analytics of their learning outcome with three numbers (See Figure 3.6-1): the number of completely missed items, the number of briefly explored items, and the number of items that have been examined with great attention. The criteria for such three categories are defined based on users' gaze duration. If users' gaze duration on a content item is longer than a predefined

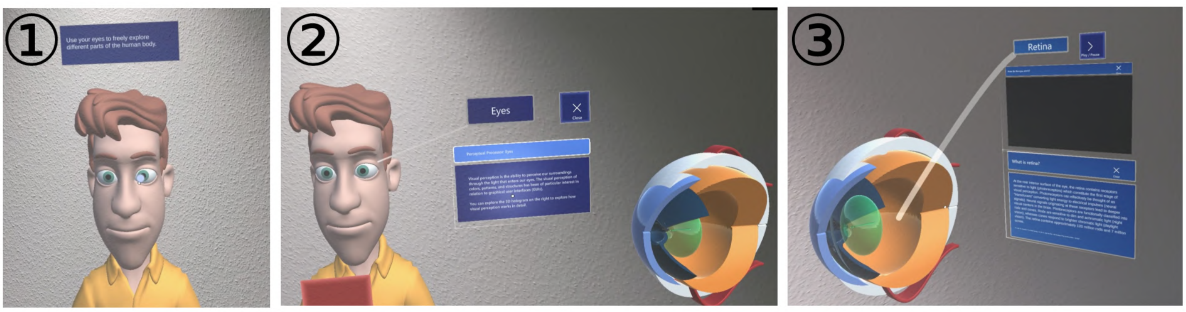

Figure 3.5: First, users use eye-gaze-based interaction to explore the main 3D model (1) and activate the second layer of content (2). Next, users gaze at the secondary models (2) to activate and explore the content items (3).

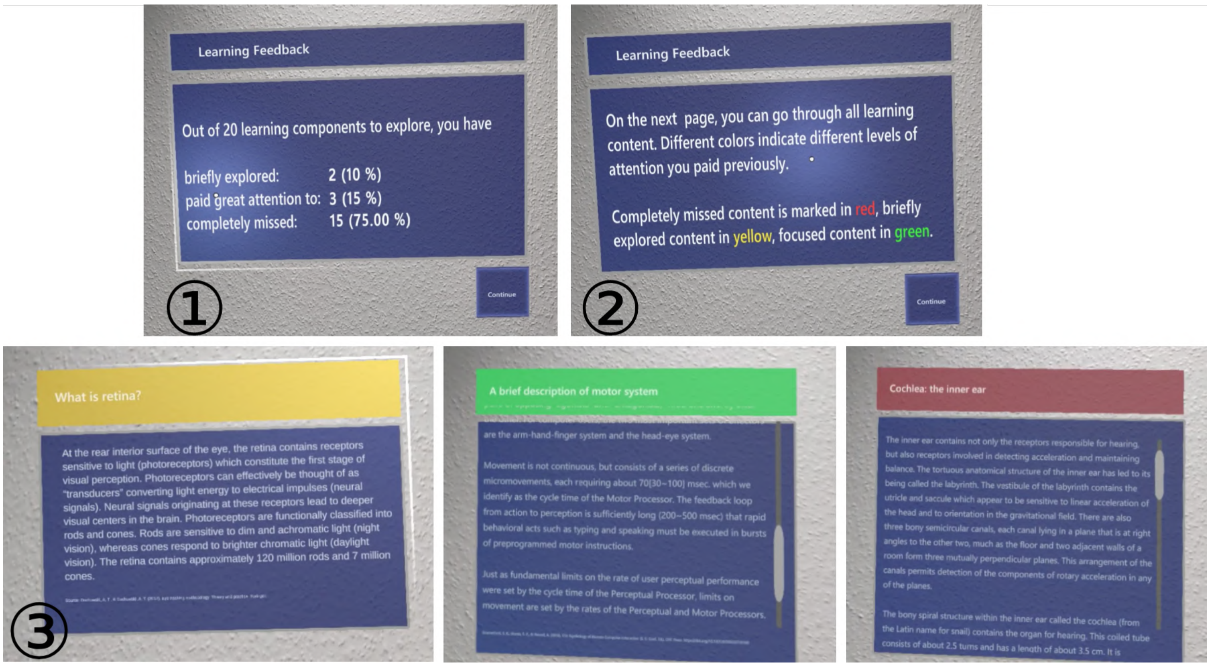

Figure 3.6: Attention feedback in our system includes basic learning analytics on users' attention (1) and color-coded flashcards for reviewing the content (2,3)

threshold set based on reading rate research (Brysbaert, 2019), the component is labeled as "explored with great attention". On the contrary, if the gaze duration is lower than the threshold, the item is labeled as "briefly explored". Other items are marked as "completely missed". Thereby, we categorize the content items in the feedback scene with three levels of attention.

The analytics is presented in an information panel consisting of the aforementioned three numbers. After reviewing the analytics, the system presents all content items as a deck of flashcards to users (See Figure 3.6-3). The order of the cards is randomized, and the system displays one card at a time. Here, each card is labeled in one of three colors: red, green, and yellow, indicating the three levels of attention. Users can navigate through the cards by selecting either the "back" or "forward" button, arranged on two separate sides of the card. Through presenting the flashcards, *AF-Mix* provides users with individual attention feedback along with a review process. During the review process, users are allowed to check their previous attention to each content item based on the color coding at their own pace. At any time during the process, they can finish the review process by selecting a "Finish Review" button at the bottom of the card deck.

Our design follows the basic idea collected in the participatory design workshop, as most participants were in favor of using flashcards to review the learning content in MR. Aiming to alleviate potential cognitive overload as suggested by the participants, all 3D models from the exploration scene are hidden in the feedback scene. Consequently, only the content items remain, serving as flashcards for attention feedback in the review process.

3.6 Evaluation Study

3.6.1 Data Collection

The evaluation of *AF-Mix* was conducted as a between-subject laboratory experiment with students from the local university in a co-learning space. The learning space had been previously designed for immersive learning experiences and self-study at the university. The evaluation received approval from the university IRB. In the evaluation study, 22 student participants (8 female / 14 male) were recruited, with an average age of 24.09 years (*SD* = 2.69). They were evenly divided into a control group and a treatment group, each comprising 11 participants. Participants who completed the study were incentivized with a monetary remuneration of 12 euros.

The entire evaluation consists of 22 sessions, including 11 sessions for the treatment group and 11 for the control group. Each session was designed to last approximately 60 minutes. In each session, one of the 22 participants was randomly selected and invited to immerse themselves in the MR learning environment with HoloLens 2. The integration of the eye tracking sensor on HoloLens 2 facilitated the collection of eye tracking data to track users' attention allocation with the *AF-Mix* system. This data collection process involved tracking participants' gaze focus on each content item, determining whether an item was examined in both the exploration and feedback scenes, and capturing the gaze duration on each reviewed item in both scenes.

Figure 3.7: A user testing the system

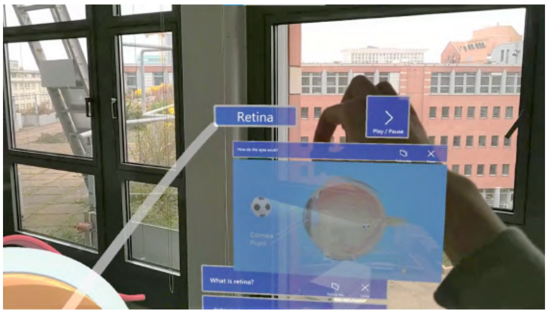

Figure 3.8: Screenshot during evaluation

Additionally, qualitative data were collected through semi-structured interviews conducted afterward. The goal of the interview is to collect feedback as qualitative data for a thematic analysis. In the coding process for thematic analysis, two coders independently reviewed and coded the interview transcripts to improve inter-coder reliability and validity. An achieved Cohen's kappa of 0.82 suggests a good agreement level through the coding process.

3.6.2 Task and Procedure

The primary objective of this study is to investigate the efficacy of providing attention feedback for MR learning environments. Thus, we introduced two separate review options within the MR learning app: one integrated with attention feedback, as described in Section 3.5 (assigned to the treatment group), and another without such feedback (assigned to the control group). The introduction and exploration scenes for the two options are identical, only the feedback scene has been modified for the control group. Participants in the control group received flashcards without color coding or analytics, unlike the treatment group, who received attention feedback along with insights into their learning outcomes, as depicted in Figure 3.6. This deliberate design meant that participants in the control group were left uninformed about which specific items they may have missed during the initial exploration. Consequently, they had to rely on their memory, while this information was presented explicitly in the feedback scene to the treatment group.

With these design considerations, we organized 22 sessions. Each session began with an onboarding process, presenting participants with study background information, explaining the experimental procedure, and clarifying their rights. After giving informed consent, participants were instructed to use the HoloLens 2 and calibrate the device for eye tracking. This concludes the onboarding process.

After the onboarding process, each participant in both groups was given the same task: They should explore the learning content freely without time constraints in the exploration scene and review their learning progress by going through the flashcards in the feedback scene with *AF-Mix*. During the process, eye tracking data were collected as described in the previous Section 3.6.1. Upon completing the learning task, each participant was invited to a semi-structured interview. The interviews were designed to take around 10-15 minutes to complete, addressing the following topics: a) overall impression; b) potential usability challenges; c) opinions on the attention feedback; and d) suggestions for further improvement. The interviews were recorded with the consent of the participants. After the interviews, participants completed the evaluation study and claimed the remuneration.

3.6.3 Participants

We recruited 22 participants (8 females). The age of the participants ranged from 19-28 years old (M = 24.09, SD = 2.69). The recruitment process was done through the online recruitment platform of the Karlsruhe Design and Decision Lab (KD2Lab) at Karlsruhe Institute of Technology[3]. The KD2Lab has a standard subject pool comprising university students from diverse educational backgrounds and with different academic degrees. The inclusion and exclusion criteria remain the same as for the design workshops. Through this process, we managed to recruit participants that represent the target user demographic for *AF-Mix*.

During the onboarding process, we inquired about participants' previous experience with immersive technologies and their familiarity with the presented topic. None of the participants had prior experience with HoloLens 2, and none were familiar with the MHP concept presented in *AF-Mix*.

3.7 Data Analysis and Results

3.7.1 Eye Tracking Data Analysis

We collected eye tracking data to assess gaze duration on each content item throughout the evaluation. This allowed for the measurement of the variables NewFeedbackItems (NFI) and TimeNewFeedbackItems (TNFI). NFI represents the total number of content items that were not initially viewed in the exploration scene but were later **discovered and examined** in the feedback scene. TNFI indicates the total gaze duration, measured in seconds, spent examining the newly **discovered and examined** content items in the feedback scene. Note that, to be classified as discovered and examined, a content item must accumulate a gaze duration exceeding 5 seconds in the feedback scene. We set this minimum threshold based on the assumption that participants, with a fast reading rate of 337 words per minute, would likely need to read at least three lines of text presented on content items (with an average length of 27.6 words) to grasp

[3]More information on the recruitment platform can be found on the website of the KD2Lab: https://www.kd2lab.kit.edu/

an overview of the content. We use prior research from Rayner et al. (2010) as a reference for the reading speed, while the average text length is calculated from the presented content items in our design. This approach aims to provide a thoughtful consideration of participants' interactions with the presented content.

Considering our between-subject design and data collection method, we opted for the Bayesian t-test framework to analyze the gathered data, as described by Rouder et al. (2009). Our research goal is to explore the impact of introduced attention feedback in the MR learning environment, and as such, we formulated relevant hypotheses. Given that the attention feedback offers learners a comprehensive overview of their prior learning outcomes, providing a clear indication of what they missed in the exploration phase through color coding, we outline the rival hypotheses regarding the NFI variable as follows: The null hypothesis ($H_0\_NFI : \delta = 0$) assumes that attention feedback does not affect the number of newly discovered and examined items in the feedback scene. In contrast, the one-sided alternative hypothesis ($H_+\_NFI : \delta = 0$) suggests that providing attention feedback results in a higher number of such items for participants in the treatment group. The δ value represents the standardized effect size.

For the TNFI variable, we propose the following rival hypotheses: The null hypothesis ($H_0\_TNFI : \delta = 0$) proposes that attention feedback does not affect the total gaze duration on newly discovered and examined items in the feedback scene. In contrast, the one-sided alternative hypothesis ($H_+\_TNFI : \delta > 0$) suggests that attention feedback leads to a longer gaze duration due to the increased amount of information (i.e., the color coding on each flashcard). Due to limited prior knowledge of the research topic, a Cauchy prior distribution ($r = 1/\sqrt{2}$, truncated at zero) has been assigned for both hypotheses. This choice ensures a conservative and unbiased approach to the Bayesian analysis.

Notably, before delving into the examination of the two variables and associated hypotheses mentioned earlier, we investigated the number of items examined during the exploration scene in a pre-analysis. The results from a t-test (BF_{10} = 0.481, BF_{01} = 2.079, $\delta =$ -0.228, 95% CI: [-0.987, 0.457]) reveal no significant difference between the two groups. This supplementary pre-analysis is crucial as it helps verify that any potential disparities observed in the feedback scene (e.g., differences concerning NFI and TNFI) are likely attributed to the design intervention rather than variations in participants' learning behavior between the two groups.

Regarding the NFI variable, the mean item count was 7.909 for the treatment group and 4.091

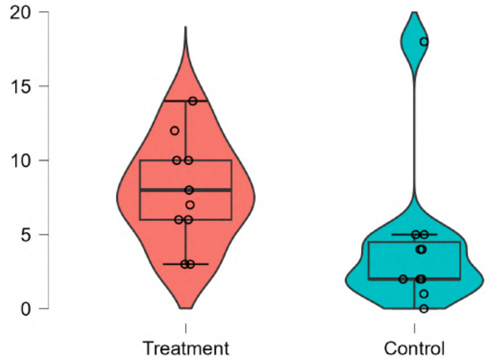

Figure 3.9: Number of newly discovered and examined content items in review phase (NFI) for the treatment group (*M* = 7.909, *SD* = 3.448) and the control group (*M* = 4.091, *SD* = 4.888)

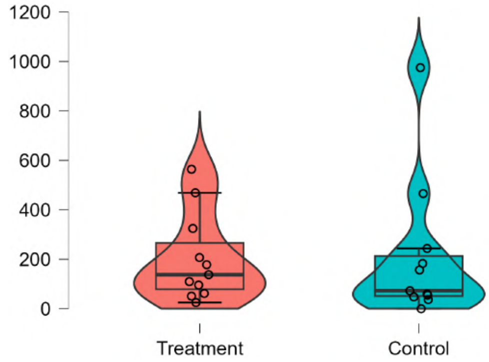

Figure 3.10: Total gaze duration (in seconds) on newly discovered and examined items in the review phase (TNFI) for the treatment group (*M* = 202.273, *SD* = 177.648) and the control group (*M* = 208.909, *SD* = 286.533)

for the control group, with standard deviations of 3.448 and 4.888 respectively. Therefore, the higher value of NFI in the treatment suggests the efficacy of attention feedback in helping learners retrieve previously missed information. Regarding the TNFI variable, the mean gaze duration was 202.273 seconds for the treatment group and 208.909 seconds for the control group, accompanied by standard deviations of 177.648 and 286.533, respectively. Upon plotting the data (See Figure 3.9 and 3.10), we observed non-normality in both variables, accompanied by the presence of outliers. These observations may be attributed to the relatively low sample size and considerable individual differences in learning strategies (Schunk, 2012). To address model misspecification arising from non-normality and outliers, we implemented a log transformation for both variables. Additionally, a Mann–Whitney U test was conducted alongside the two-sample t-test, as it remains unaffected by the log transformation. This dual-method approach has been applied in existing studies to ensure the robustness of a Bayesian t-test analysis (Doorn et al., 2020). For a comprehensive overview of the collected data, descriptive analyses, and the Bayesian analysis results, please refer to the accompanying .jasp file, accessible at https://osf.io/5t7vf/.

The findings reveal strong evidence supporting H_{+}_NFI as the Bayes factor (BF_{+0}) is 11.757, indicative of a great likelihood for the alternative hypothesis. Consistently, the Mann–Whitney U analysis yields a BF_{+0} of 5.171, providing moderate evidence in favor of the alternative hypothesis. This finding underscores the effectiveness of attention feedback in helping learners retrieve previously missed information. On the contrary, for the gaze duration on newly exam-

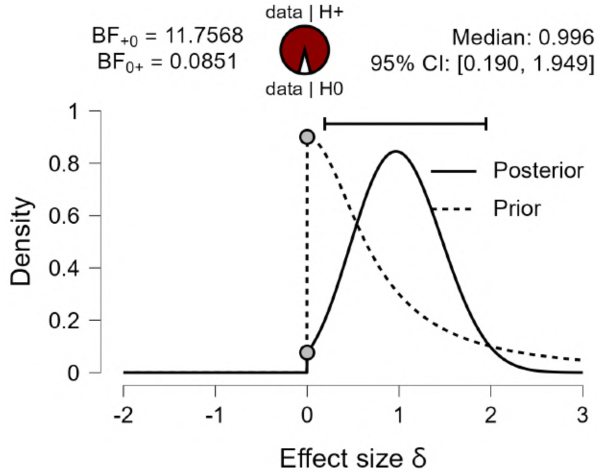

Figure 3.11: Hypothesis testing with one-sided Bayesian independent samples t-test for $H_+\_NFI : \delta > 0$

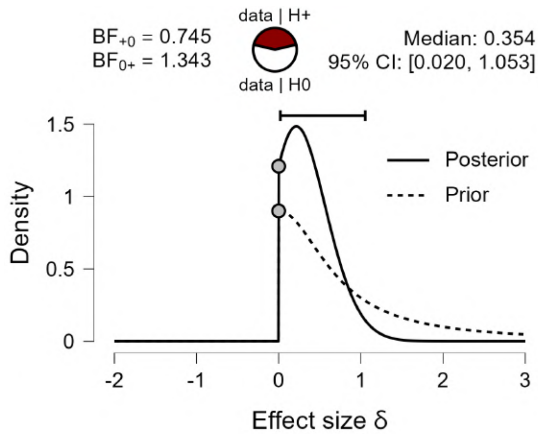

Figure 3.12: Hypothesis testing with one-sided Bayesian independent samples t-test for $H_+\_TNFI : \delta > 0$

ined items in the feedback scene, no decisive evidence is shown for H_+_TNFI, with BF_{+0} = 0.754 and BF_{0+} = 1.343. The error percentage in both cases is $< 0.001\%$. Further exploration and analysis may be needed to better understand the impact of attention feedback on the time spent examining newly discovered items.

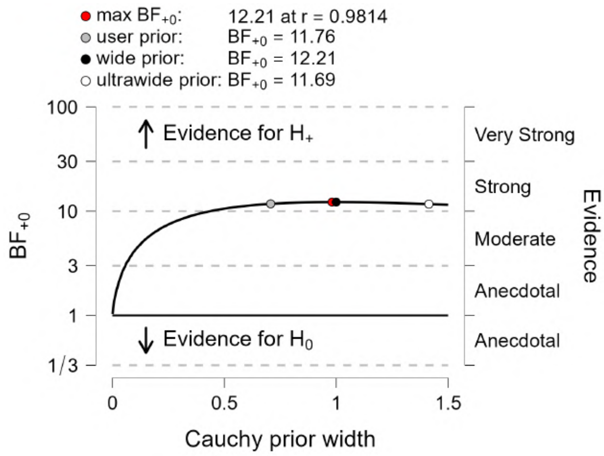

Figure 3.13: The Bayes factor robustness plot for BF_{+0} concerning NFI. The maximum BF_{+0} = 12.21 was reached at a prior width r of 0.98. The evidence for the alternative hypothesis is stable throughout different prior distributions.

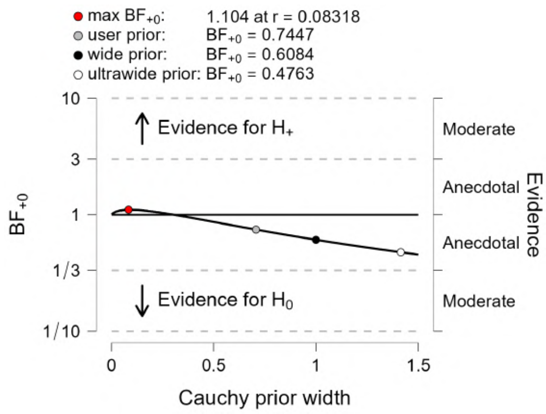

Figure 3.14: The Bayes factor robustness plot for BF_{+0} concerning TNFI. The evidence for both the null and alternative hypothesis is rather weak across the range of prior distributions, with slightly more evidence for H_0 as the prior width increases.

To ensure the robustness of our findings, we conducted additional analyses. Figure 3.13 and Figure 3.14 depict the Bayes factors concerning the prior width (r). For BF_{+0} regarding variable NFI, the Bayes factor exhibits great stability, ranging from 3 to 12 across a broad range of prior widths from 0.2 to 1.5. In the case of the second pair of hypotheses, the BF_{+0} value attained its peak at the outset (1.104 at r = 0.083) before gradually declining, eventually leaning towards evidence for the null alternative. Lastly, we investigate the parameter estimation for the first hypotheses pair on variable NFI, as the previous Bayesian t-test indicates strong evidence for H_{+}_NFI. By conducting a two-sided t-test, we observe a posterior distribution of δ = 0.99 with a 95% CI ranging from 0.128 to 1.946.

3.7.2 Thematic Analysis

For the collected qualitative data, we conducted a thematic analysis of the interview transcripts. 18 out of 22 participants provided consent for interview recordings, which were later transcribed for coding. The remaining four participants agreed to the interview but chose not to be recorded. Instead, we opted for taking notes during their interviews as an alternative to transcripts. The interviews had an average duration of 10.1 minutes (*SD* = 4.342), and the average length of each transcript is 1082.1 words (*SD* = 559.49).

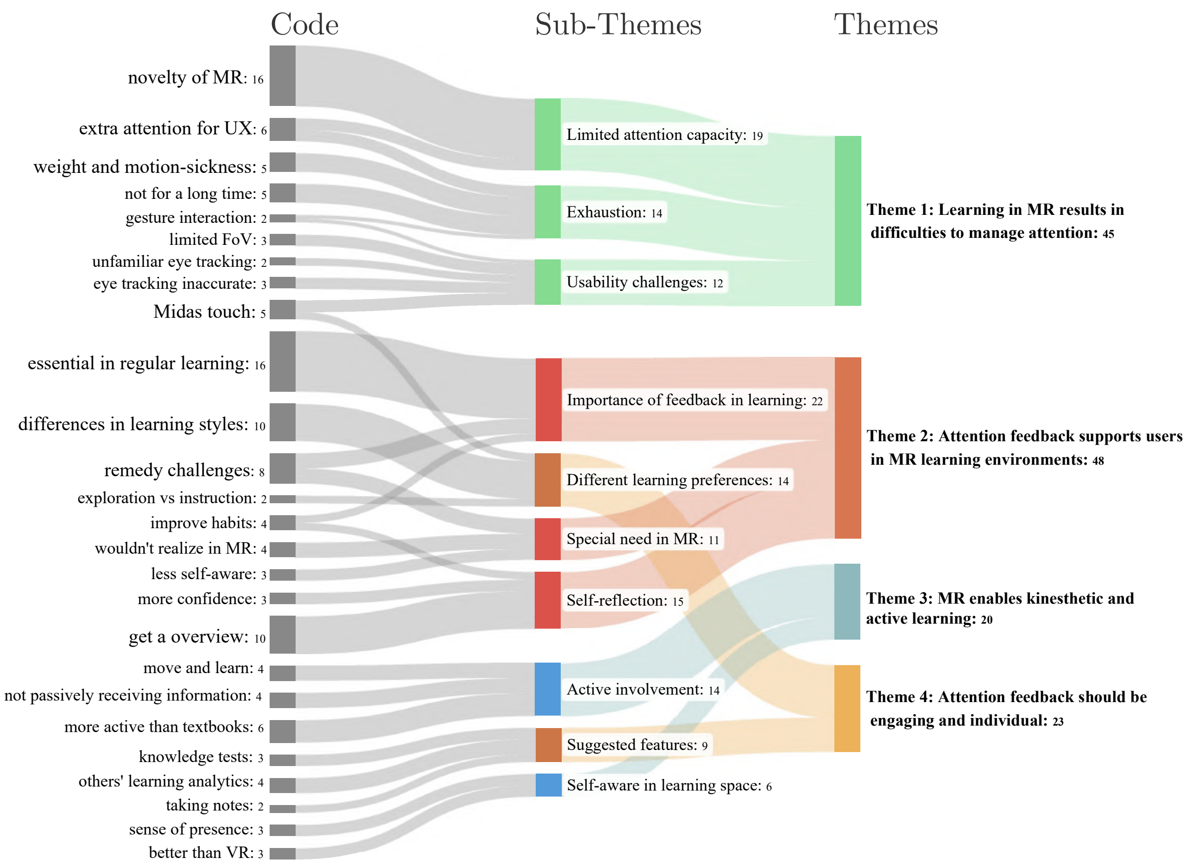

Figure 3.15: Visualization of the thematic analysis with a Sankey diagram. The numerical values next to each code suggest its occurrence frequency within the transcripts

After collecting the qualitative data, we conducted a thematic analysis, resulting in a set of themes that represent participants' opinions on *AF-Mix*. In the process of coding and generating themes, we followed the guidelines proposed by Braun and Clarke (2021). In line with guidelines, we embraced the concept of recognizing themes as "stories we tell about our data" (Braun and Clarke, 2021). The inductive coding process, involving two independent coders (the first and second authors, both HCI researchers with experience in qualitative analysis), produced 92 codes. Subsequently, through collaborative efforts, these codes were consolidated into 27 overarching codes. A group discussion led to the identification of 10 sub-themes from the example codes. These sub-themes were then examined against the transcript data for accuracy and validity. Ultimately, four main themes emerged from the thematic analysis.

To ensure the reliability of our analysis, we cross-referenced the identified themes with the original codes from the transcripts. Additionally, Cohen's Kappa coefficient was calculated during the thematic analysis, with a result of 0.82 for the coding process and 0.88 for theme generation, indicating a satisfactory level of agreement between coders.

The resulting themes from our qualitative analysis are described in the following section. Quotes are presented with participant information enclosed in brackets, denoting their respective groups: C for the control group and T for the treatment group. For instance, PT1 indicates participant number one, who was in the treatment group.

Theme 1: Learning in MR results in difficulties to manage attention

Among the 18 participants who gave consent to the recording and transcribing of the interviews, 16 participants acknowledged the novelty of using MR for self-directed learning. One participant reported that using HoloLens 2 is "full of wow factors" (PC2). Nonetheless, the learning experience in MR was somewhat overwhelming. As a result, participants couldn't manage to concentrate on the learning content as intended. Instead, they paid more attention to the proficiency of interaction methods because they "didn't use anything like this before" (PT17).

Most of our participants (16 out of 22) hadn't used any MR devices before the evaluation study. For the six participants who had experience with MR, they had only used video-see-through (VST) devices such as HTC Vive. Therefore, all participants were first-time users of HoloLens 2. Common problems identified from the qualitative analysis include usability challenges and motion sickness. Three participants from both the control groups (PC2) and the treatment group (PT17, PT15) complained about the limited field of view (FoV) of the device. Two other participants (PC4, PT7) reported the difficulty of using their fingers to press the buttons in MR environments, while other participants suggested that the eye-gaze-based interaction was not easy to use, because “the gaze indicator is always 1-3 centimeters off from my real gaze position” (PT1). In the interviews, five participants explicitly suggest that the experience tends to “get very exhausting after a while” (PT15) and they “couldn’t use it for a long time” (PC8). As a consequence, participants found it tricky to pay full attention to the learning content as they got accustomed to the learning environment.

Theme 2: Attention feedback supports users in MR learning environments

All participants in the treatment group expressed positive opinions on receiving the feedback information. Attention feedback in MR seems to remedy difficulties in the learning process and “optimize the whole learning process” (PT5). Many participants (16 out of 22) admit that receiving feedback is an essential process in their regular learning strategy, so integrating feedback into the learning process in MR made them more confident. Besides, the simple design and interaction method also contributed to supporting users in focusing on reviewing their learning progress. Overall, participants who reviewed their learning progress with attention feedback suggested the following significant advantages of having attention feedback in MR: First, attention feedback helped them find the content they had missed. Reviewing the content with attention feedback helped participants to “**see the big picture**” (PT3). Furthermore, attention feedback supports users in reflecting on their learning journey. The feedback provokes participants to think about “how I (they) have read and learned the content” (PT1) and “improve my habits of learning” (PT5). Eventually, participants reported that *AF-Mix* helped reduce their cognitive workload and improved their confidence after the learning journey. One participant explained this as follows:

> I really liked that part (feedback scene). Because this is my first time using this device (HoloLens 2), I wasn’t sure whether I was doing the right thing or whether I have checked out everything. With the feedback, I know what I missed and get a better sense of the topic. (PT9)

In the control group, participants also suggested the need for such feedback. They recognized the learning support provided in their group by the regular flashcards (not color-coded) because they "gave me (them) the opportunity to review everything" (PC18). However, five participants admitted that they needed more support in reviewing their learning process. Using regular flashcards presented in the control group, some participants (PC2, PC4, PC6, PC18) still indicated that "I (they) wouldn't have even realized I've missed something" (PC2). To those participants, we proposed the design of adding color coding based on eye tracking data to the flashcards. All of them agreed that such a design would support them even better. One participant claimed that this design can "help you make sure that you really learn the whole thing and not just some information" (PC6).

In both groups, even participants who don't integrate feedback as a necessary part of their learning routine also appreciated having attention feedback in MR. Participants in our evaluation have reported different attitudes toward feedback for their regular learning experience (e.g., exam preparation). Some participants revealed that "feedback is not part of my learning strategy at all" (PC14), while others claimed that it is a crucial part of exam preparation for them (PT1, PC6, PC16). One participant justified her neglect of feedback as follows: "When I study for exams, instead of finding a way to get feedback, I might as well just keep studying" (PC14). Nonetheless, we found that attention feedback in MR environments is desired by all participants. Even for participants who normally don't value feedback, attention feedback is also desired in MR for two reasons. First, some participants explained that receiving attention feedback, at last, helped them build confidence, as a new learning environment in MR makes them (PC2 and PT9) wonder: "Have I really learned everything?" Secondly, participants expressed a desire to "know more about myself" (PT7), and receiving attention feedback fulfills such a need.

Theme 3: MR enables kinesthetic and active learning

Despite the usability challenges faced in MR learning environments, participants still acknowledged the unique advantages of the learning experience. Four participants from different groups (PT1, PC2, PT5, PT13) pointed out that users need to "move and learn" in MR environments. Even though a few participants suggested that it can be "demanding" for specific learning contexts, others assumed such a learning method may help users retrieve information and foster knowledge by nudging them to create a connection between spatial contexts and the learning content. One participant gave a concrete example of writing an exam:

> I noticed different information is arranged at different places in the space. So you can imagine when you're taking a quiz (about the MHP concept), you may still remember where the ear was (in the MR environment). This can help you recall relevant information. (PT5)

In general, participants needed to actively explore the learning content in MR by moving in the room and interacting with different elements, instead of passively receiving information as students would do when they "learn something from a textbook" (PT5). Previous research also suggests the advantage of motivating kinesthetic learning in MR environments (Knierim et al., 2020), and our thematic analysis supports this assumption. Apart from comparing our system with traditional learning formats such as textbooks, several participants (PC2, PC6, PC16) also compared it with the learning experience using VR-HMDs:

> Especially when I'm in this room (learning space at the university), I prefer using HoloLens instead of VR, (because) when I see the surrounding environment, I realize this is a space for learning and I should concentrate. In VR, I'm totally immersed in a virtual space (PC6).

Theme 4: Attention feedback should be an engaging and individual

Participants in our study exhibited different learning skills and strategies. For some participants, exploring the learning content actively in the exploration scene is an engaging experience. For others, self-directed learning as well as content exploration can be confusing, and they would like to "see something like a table of content(s)" and learn everything "with instructions and perhaps with someone else" (PT15). Such comments reflect individual differences in learning styles and strategies. Therefore, diverse and sometimes contradicting suggestions for improvements were proposed by participants.

First, participants reported a contrast between the exploration scene and the feedback scene. The exploration scene requires users to actively activate content items with eye-gaze-based interaction and move around the physical space, while the feedback scene presents all items one by one to users. While some participants recognized this as an effort to reduce cognitive workload in the feedback scene, others found it less engaging than the exploration process. Another emerging issue is the so-called Midas touch problem (Jacob, 1995; Velichkovsky et al., 1997). Since the intents behind visual attention remain unclear to our gaze-aware system, participants sometimes find the feedback frustrating. One participant elaborated on this during the interview as follows:

> I already knew everything about certain parts, so I skipped it. However, in the end, the application suggested I need to pay more attention. On the other hand, I looked at one part because I was confused, and it was hard to understand; but in the end, I was told I already fully explored it. This inaccuracy demotivates me a little bit. (PC2)

Participants also proposed potential solutions and attempted to make the process of receiving attention feedback more engaging. Four participants (PC4, PT7, PT15, PC16) suggested that integrating knowledge tests in MR can be supplementary to the current design. The knowledge tests can also contribute to determining the reason behind visual attention. For example, if a user could answer all questions regarding a certain item correctly but hadn't paid attention to it during the learning process, *AF-Mix* could assume that users had skipped it because of their familiarity with the content. Another idea to make the feedback and review process more engaging is presenting statistics of fellow users. This idea was elaborated by one participant as follows:

> I don't mean that I care about other people who are studying the same thing, but seeing the learning analytics of everybody else who is learning the same thing can really motivate me in reflecting on my own learning behavior. (PT1)

3.8 Discussion

In general, participants of the evaluation study reviewed our system as an effective approach to helping them better allocate their attention during the learning journey. The analysis of the eye tracking data also suggests that providing attention feedback helps users discover previously overlooked content more efficiently. Notably, this didn't require additional time within the feedback scene on the newly discovered items. This observation suggests that the implemented design features may not have induced cognitive overload, preventing a scenario where participants would require additional time to process supplementary information, such as color coding. This confirms that our designed learning system is aligned with what users suggested in the participatory study.

Overall, the eye-tracking data seem to suggest the following pattern introduced by the attention feedback: While all participants exhibited similar learning progress and outcomes during the exploration scene (validated by the pre-analysis), and they all had the option to examine all 20 content items in the feedback scene, participants in the treatment group demonstrated in the feedback scene a different review strategy. They effectively directed more attention toward previously overlooked items, resulting in a higher count of NFI (it's worth noting that a minimum gaze duration of 5 seconds is required for an item to be counted as NFI). Simultaneously, aided by color coding, they examined the items they had missed more efficiently in the review process, spending less time on each item than the control group. This resulted in a TNFI value comparable to that of the control group. This nuanced shift in attention allocation and review strategy highlights the potential impact of the introduced design intervention on participants' interaction with the content in the MR learning environment.

3.8.1 Design Implications

Meanwhile, the analysis of qualitative data reveals diverse opinions among users regarding attention feedback. In summary, we outline the following design implications:

Facilitating Effective Learning Through Attention Feedback

Exploring learning content in MR was reported by participants to be an engaging, innovative, and unique experience, and it promotes active learning. Nonetheless, they often felt overwhelmed, as usability challenges (e.g., limited FoV, inaccurate calibration) disrupted their attention in the learning process. These challenges may affect users' autonomy and self-confidence in using MR systems. Concerning the learning support presented in our system, participants in both groups reported that the flashcards had reduced their cognitive workload after the exploration, as presenting flashcards in the feedback scene allowed participants to see all learning content and reflect on their learning progress. Under these conditions, participants in the treatment group emphasized that the gaze-aware attention feedback allowed them to reflect on their learning journey and allocate their attention more effectively during the review process. Consequently, attention feedback in MR helped them "see the big picture" (PT3). In addition, participants in the control group suggested that adding such gaze-aware support could have supported them further in the review process. Therefore, when comparing the results of the two groups, we consider that the color-coded flashcards with attention feedback provided participants with more effective learning support.

Even participants who did not typically integrate feedback into their regular learning routine considered receiving attention feedback in MR environments as necessary and beneficial. We attempt to interpret this opinion as follows: When participants explore content presented in traditional learning formats (e.g., textbooks), they have an established strategy to approach the content from their accumulated experience. Having much experience in efficiently using the given formats, feedback may not be deemed as quintessential for learning success by some participants. However, if participants are exposed to a new learning environment such as MR, with no prior experience and success, they may need more learning support such as attention feedback. Further research can explore the hypothesis that there is a correlation between the level of prior experience in MR and the need for attention feedback as a form of learning support. Lastly, our results suggest that attention feedback in MR should be designed compactly, assisting users in reducing cognitive overload and managing their attention.

Tailoring Attention Feedback to Individual Learning Styles

While participants found *AF-Mix* to be a practical support mechanism in using the MR learning system, our interpretation of the evaluation results revealed variations in participants' learning styles and preferences concerning attention feedback. Due to this variation, alternative designs for attention feedback were proposed. These include two ideas: First, integrating learning analytics of fellow students to promote extrinsic motivation. Second, making attention feedback more interactive with a quiz based on users' previous attention.

The first alternative design may benefit users who prefer peer learning over self-directed learning environments, information on the learning progress of others may motivate them to review their own learning outcomes. In contrast, the second alternative design might appeal to users favoring inquiry-based learning. Unlike our current design, which presents attention feedback with flashcards, the alternative design with quiz questions may better provoke active learning in the feedback process.

Having alternative design ideas proposed by participants doesn't necessarily suggest the inferiority of our current design. Rather, these alternative ideas illustrate the diversity of learning styles among users. Both our findings of the evaluation study and existing studies suggest that people embrace different learning strategies (Schunk, 2012). As a result, it's impossible to suggest a universal method to support learning in MR with feedback. One type of feedback may be effective in some contexts for some users but not for others. Besides, our relatively small sample size and predefined learning topic can not represent all learning scenarios in MR. We argued against a one-size-fits-all design of attention feedback. Our work aims to showcase the potential of attention feedback for MR learning environments and propose one effective design. More research is needed to understand how different designs of attention feedback in MR can benefit learners with different mindsets and learning styles.

3.8.2 Limitation and Future Work

Currently, *AF-Mix* only incorporates one method of providing feedback: attention feedback based on eye tracking. This was inspired by the identified research gap in providing gaze-aware attention support in MR learning systems. However, MR learning environments may benefit from other types of feedback. For instance, giving feedback on whether a physical action is performed correctly is crucial for kinesthetic and spatial learning. Existing research integrates such feedback in MR by using external sensors to collect behavioral data and compare it to a predefined baseline (Shao et al., 2020). Additionally, motivating kinesthetic and spatial learning is a major benefit of MR environments, as recognized by our participants. However, participants mentioned that the current design doesn't effectively encourage kinesthetic learning during the feedback and review process. Therefore, future work should explore additional methods to enhance *AF-Mix* with feedback for promoting kinesthetic learning.

Furthermore, understanding the intent behind users' attention is missing in our system, which is commonly known as the Midas touch problem in eye tracking research (Jacob, 1995). Participants in our study reported a lack of extrinsic motivation, when *AF-Mix* constantly assumes the missing attention as a result of poor attention management. In reality, users may ignore certain content because they've already acquired the relevant knowledge before. Meanwhile, paying excessive attention to certain content doesn't always lead to a better understanding of the content. Instead, our qualitative analysis shows that it can be correlated with confusion or frustration. Therefore, tracing users' attention and inferring user intent based solely on eye tracking data may be insufficient. Future work could explore ways to interpret individual intent behind attention, possibly incorporating additional biosignal data for more accurate attention feedback presentation.

Moreover, future research can enrich the Bayesian analysis with more data and longitudinal studies, shedding light on nuanced aspects of the introduced attention feedback. Specifically, it remains unclear whether the integration of attention feedback, including gaze-aware learning analytics and color-coded flashcards, requires learners to adapt their learning or review strategies. Presently, there is no strong evidence suggesting that attention feedback demands an increase in review time for previously missed items (see Section 3.7.1 for rival hypotheses H_0_TNFI and H_+_TNFI), a positive indication that the system does not introduce unforeseen challenges in attention management when presenting gaze-aware assistance, resulting in participants' feeling overwhelmed and needing more time. However, the absence of a significant increase in review time may also imply that the feedback does not always stimulate active self-reflection, a process that typically consumes more time. Due to the relatively low sample size for our evaluation, no conclusive interpretation can be made to clarify the results. Collecting data on cognitive load through a survey including the NASA-TLX questionnaire may also provide valuable insights into the mental processes involved. Exploring these nuances can enhance our understanding of the cognitive processes influenced by the feedback mechanism.

Lastly, our study aims to enhance the MR learning experience through support features, such as attention feedback. We did not compare the advantages of MR learning systems with traditional formats, focusing instead on refining the current MR learning experience with an existing system. We somewhat presume the benefits and advantages of MR systems over traditional learning formats as inherent in some aspects, as various advantages have been highlighted in prior research (Akçayır and Akçayır, 2017a; Mohammadhossein et al., 2022). Nonetheless, to what extent such advantages hold true and whether they apply to all contexts (including HCI education in our study), require further research. We acknowledge the need for future work to validate assumed benefits and explore the specific advantages of MR learning systems over other learning formats, particularly in under-researched areas such as HCI education.

3.9 Conclusion

In this paper, we present *AF-Mix*, a gaze-aware learning system that provides attention feedback in MR. The motivation for integrating attention feedback in MR is to assist users in managing their attention in the learning experience. We identified several challenges users typically encounter in MR learning environments, including attention management. These challenges highlight the growing need for effective support in the learning experience. Therefore, our goal is to provide learning support by assisting users in managing their attention and reviewing their learning outcomes with attention feedback in self-directed MR environments.

By adopting a human-centered design approach to develop *AF-Mix*, we first conducted a participatory design study. Through participatory design workshops, we adopted the idea of visualizing the learning content as color-coded flashcards based on users' attention. We delivered the final design as follows: *AF-Mix* first allows users to actively explore the content using eye-gaze-based interaction while collecting eye tracking data during the process. Next, *AF-Mix* notifies users of their attention distribution based on the collected data, including the number of items users have previously missed, briefly explored, and explored with great attention during the content exploration. Subsequently, *AF-Mix* displays all learning content as a deck of flashcards. Here, each flashcard is color-coded based on users' previous attention levels, allowing users to review their learning outcomes.

After designing and implementing *AF-Mix*, we evaluated our system with 22 participants. The results show that *AF-Mix* assisted participants in figuring out what content items they had missed during the content exploration and motivated them to examine these items in the review process. Eventually, participants described *AF-Mix* as efficient in guiding them to retrieve missed information and reflect on their learning progress. Thereby, they could allocate their attention more efficiently in the review process.

In summary, our work contributes to the field of HCI by highlighting the challenges associated with attention management in MR learning environments and providing effective gaze-aware support through attention feedback. Our findings suggest that attention feedback serves as an effective gaze-aware support method to improve the overall learning experience in MR. This research also offers a foundation for further exploration: The design methods and solutions proposed in our paper can also be applied to other learning contexts in MR. Future research can extend the design knowledge with evaluation in different learning scenarios, contributing further implications and innovative methods for enhancing gaze-aware support in MR learning environments. We believe that this study can contribute to unleashing the full potential of MR and provide valuable insights for ongoing research in designing human-centered MR learning systems.

4 GazeNote: Designing Note-taking Support for Immersive Learning with Gaze-Aware and Cross-Device Interaction (Study III)

Abstract: Existing immersive learning systems, often implemented with mixed reality (MR) technology, frequently lack support for note-taking. From a pedagogical standpoint, note-taking is essential for cognitive offloading, knowledge retention, and review. While MR enriches learning through 3D visualization and multimodal interaction, learners encounter challenges in managing cognitive load and retaining information. Simultaneously, developers often prioritize the representational fidelity of immersive environments and struggle to incorporate effective learning support features such as note-taking. To address this gap, we introduce GazeNote, an innovative note-taking system that facilitates seamless cross-device interaction between head-mounted displays and external mobile devices (such as tablets or smartphones). It can be integrated into existing Unity-based MR learning systems as an extension. We conducted a formative design workshop (n = 6) to derive key design goals. A system evaluation (n = 18), employing a mixed-methods approach, was conducted to assess its usability and users' perceptions. Our contributions include: 1) the design and implementation of a novel artifact, 2) empirical insights into the challenges of immersive learning with MR and note-taking, and 3) stakeholder-informed design implications and requirements for effective note-taking support. These findings can advance human-centered design of interactive systems for immersive learning, paving the way for more effective and engaging MR educational experiences.

Keywords: Mixed reality, Gaze-adaptive systems, Note-taking, Cross-device interaction

Under Review as: Liu, S., Toreini, P., & Maedche, A. (n.d.). GazeNote: Designing Note-taking Support for Immersive Learning with Gaze-Aware and Cross-Device Interaction. (Under Review at *Interacting with Computers*)

4.1 Introduction

Immersive learning leverages advanced technologies to create engaging, interactive educational experiences (Kuhail et al., 2022; Mystakidis and Lympouridis, 2023). A prominent approach is the use of Mixed Reality (MR) technology, which creates learning environments that blend real and virtual elements to enhance understanding and engagement (John et al., 2022; Tang et al., 2020). With its unique advantages and recent advancements, the adoption of MR learning has expanded, finding applications in medical training (Moro et al., 2021), social learning (Pamparău et al., 2021), language learning (Vazquez et al., 2017), and STEM education (Radu and Schneider, 2019). Research has demonstrated MR's potential to create immersive learning environments that help students grasp abstract concepts (Alzahrani, 2020), enhance motivation (Essmiller et al., 2020), and promote active learning (Mohammadhossein et al., 2022). By integrating interactive 3D visualizations and multimodal interactions, MR learning systems provide novel ways to engage students in a unique learning experience.

Despite their advantages, immersive learning systems have limitations that can impede effective knowledge acquisition (Christopoulos et al., 2024; Mayer et al., 2023; Wang et al., 2020). The immersive nature of MR, characterized by rich visualizations and interactive elements, can contribute to cognitive overload, complicating learners' ability to process and retain information (Charfi et al., 2009; Mayer et al., 2023; Wang et al., 2020). Furthermore, MR learning is often restricted to the in-situ immersive experience, lacking flexible mechanisms for revisiting and reinforcing knowledge afterward (Fernandes et al., 2023; Van Der Veer et al., 2024). At the same time, note-taking is a well-established cognitive strategy that facilitates cognitive offloading. It helps learners manage information overload during learning sessions and serves as a valuable tool for review and self-reflection afterward (Jensen and Konradsen, 2018; Moos, 2009). While existing research has explored the benefits of note-taking in educational contexts (Jensen and Konradsen, 2018), there has been limited focus on integrating note-taking as a learner-centric support mechanism within MR and immersive learning systems (Di Gioia et al., 2022; Yang and Chiu, 2014). Our work seeks to address this gap by developing a system that seamlessly incorporates note-taking into MR learning settings, thereby enhancing cognitive offloading and facilitating post-learning review and self-reflection. To address this, we conducted a participatory design workshop in which students shared their current note-taking practices and generated ideas for MR-supported note-taking. Findings underscore the importance of cross-device interaction, with participants highlighting the necessity for tablet-based note-taking synchronized with MR headsets.

Building on insights from the participatory design workshop, we present *GazeNote*. *GazeNote* extends existing MR learning systems to support cross-device note-taking. Our system facilitates such interaction by enabling learners to capture their field-of-view in MR as screenshots, which are then seamlessly transferred from MR headsets to tablets for annotation. During our participatory design workshop (n=6), participants expressed keen interest and positive anticipation regarding the potential of leveraging eye tracking to enhance note-taking in MR environments. This interest aligns with cognitive research that demonstrates a strong link between eye gaze and attention (Rayner, 1998). Eye tracking offers insights into learners' visual attention by revealing what information they are potentially processing (Just and Carpenter, 1980). Furthermore, prior work highlights the use of eye tracking to support attention management and multitasking (Langner et al., 2020; Peng et al., 2022), both of which are crucial for effective learning within complex MR environments. Therefore, *GazeNote* incorporates real-time visual heatmaps of learners' attentional distribution, generated from eye tracking data collected during MR learning sessions. These heatmaps are synchronized with the note-taking application on the mobile device, serving as a gaze-contingent reference to support reflective note-taking.

We evaluated *GazeNote* in a user study involving 18 participants to analyze its usability and learners' preferences for note-taking interaction in immersive learning experiences with MR. A mixed-methods approach was employed for the system evaluation. Quantitative survey data included NASA-TLX and UEQ scores to assess workload and user experience. In addition, we conducted a behavioral analysis of log data, examining note-taking activity (e.g., number of notes, usage time, gaze shifts) and investigating the relationship between notes taken and participants' gaze data to understand how visual attention influenced note-taking strategies. Qualitative insights were derived from a thematic analysis of semi-structured interviews. The results indicated a satisfactory user experience and revealed key design implications for supporting note-taking in MR learning systems. These implications suggest the importance of accommodating diverse note-taking strategies, balancing guidance with learner autonomy, and leveraging cross-device interaction to seamlessly bridge learning activities.

This study aims to contribute to the following areas through the design and evaluation of our cross-device note-taking system: First, we propose three distinct approaches for cross-device note-taking in MR learning settings and implement an MR support system to demonstrate these approaches. Second, through our mixed-methods user evaluation, we capture user preferences and perceptions of the proposed note-taking techniques. Third, we propose design implications and identify future research opportunities aimed at supporting note-taking for MR learning.

In summary, this study seeks to advance the HCI research in several ways: First, we present the design and implementation of the novel artifact *GazeNote*, an innovative note-taking system for MR learning environments featuring gaze-awareness and cross-device interaction. Second, through a mixed-methods evaluation, we provide empirical insights into learners' cognitive challenges when taking notes in MR. Finally, informed by our findings, we elicit requirements and derive actionable design recommendations for effectively integrating note-taking into immersive learning environments. These contributions will: 1) **empower learners** with more engaging and effective note-taking strategies to enhance their learning experiences, 2) **guide software engineers and developers** in creating more human-centered support mechanisms for immersive learning systems, and 3) **enable educators and authoring tool users** to seamlessly integrate note-taking into their pedagogical practices and MR learning content.

4.2 Related Work

4.2.1 Immersive and Mixed Reality Learning Systems

MR has emerged as a powerful technology for seamlessly blending physical and virtual environments (Milgram and Fumio, 1994). In the last decade, one definition of MR has been refined as a more interactive and engaging Augmented Reality (AR) experience, often utilizing head-mounted displays and advanced sensors for multimodal interaction (Speicher et al., 2019). As an immersive technology, MR offers unique opportunities for learners to engage with educational content and "enhances learning through multiple perspectives" (Chris Dede, 2009). Existing immersive learning experiences with MR systems enable learners to explore learning materials in innovative and interactive ways, potentially enhancing their engagement and motivation (Alzahrani, 2020; Mohammadhossein et al., 2022; Wu et al., 2013). However, it is noteworthy that prior research on MR learning systems has predominantly focused on visualization and creating immersive learning experiences (Akçayır and Akçayır, 2017a; Maas and Hughes, 2020). One exception is the work of Zhang et al. (2023), which presents a gamified design to support knowledge retention. Yet, the primary focus of their study is to compare the cognitive benefits of an immersive learning system against a physical version, rather than providing cognitive support in existing immersive systems. Compared to traditional educational formats, many effective learning strategies, such as note-taking, have not yet been thoroughly investigated or supported in MR learning systems (Azizo et al., 2022; Yang and Chiu, 2014).

4.2.2 Note-Taking as Learning Support

Note-taking has been recognized as an effective learning technique (Jansen et al., 2017; Kiewra et al., 1991; Kiewra and Frank, 1988). In learning contexts, note-taking involves the process of selecting relevant information and subsequently recording it (Piolat et al., 2005). Beyond processing existing information, note-taking also encompasses summarizing and "externalizing one's thought-processes" (Romat et al., 2022). Prior education research has identified various benefits of effective note-taking, including external-storage and encoding effects (Kiewra and Frank, 1988). During learning activities, actively taking notes can assist learners by fostering knowledge (Shrager and Mayer, 1989), aiding information retrieval (Golovchinsky et al., 1999), and facilitating the development of self-regulated learning skills (Zimmerman, 2002).

In contrast to educational research, previous studies in HCI and software engineering have primarily explored using innovative technologies to create interactive note-taking processes, rather than leveraging the cognitive benefits provided by these systems. Some studies focus on empowering users with specific note-taking methods in digital learning. conducted an extensive analysis of the sketchnote method by examining common practices and instances of its application for note-taking purposes. Similarly, Tholander and Normark (2020) investigated the bullet journal method and proposed design recommendations to support this technique. Another research trend in HCI focuses on designing and evaluating innovative note-taking interactions and systems, such as note-taking with mobile phones (Khan et al., 2021), using projected interfaces (Ren et al., 2014), and comparing different input modalities (Mitsuhara et al., 2010), and comparing different input modalities (Khan et al., 2022). Among these research streams, our contribution is positioned in the second category. Specifically, we design and evaluate a novel note-taking interaction for MR learning systems.

For HCI research focused on designing note-taking systems, the primary goal typically involves enabling users to employ note-taking techniques in digital learning environments, aiming for interactions that are both innovative and efficient (Nguyen and Liu, 2016; Porta, 2008). The work of Hailpern et al. (2015) focused on using note-taking to support hybrid learning in their METIS platform, rather than for immersive learning. Among these studies, eye tracking and gaze-based interaction have been identified as effective interaction modalities to support e-learning experiences (Porta, 2008). Specifically, note-taking with gaze-based interaction and generating annotations with eye tracking data have been investigated in prior research (Cheng et al., 2015; Nguyen and Liu, 2016).

4.2.3 Supporting Note-taking in Mixed Reality Learning Systems

Existing research on immersive and MR learning systems often overlooks note-taking as a vital learning technique, offering limited support for it. Compared to Virtual Reality (VR) devices, MR headsets allow users to maintain their perception of physical surroundings through optical see-through or pass-through technologies (Milgram and Fumio, 1994; Speicher et al., 2019). For note-taking purposes, this capability opens up possibilities to use external devices in the physical environment for note-taking and cross-device interaction. Among related studies, Denoue et al. (2003) highlight the potential of cross-device interaction for supporting note-taking. However, their work does not involve MR systems, focusing instead on multiple handheld devices. In a recent study by Di Gioia et al. (2022), the authors explore the use of Microsoft HoloLens combined with mobile devices for note-taking scenarios. Their contribution lies in leveraging MR devices and multi-device interaction as note-taking tools in classrooms and regular learning scenarios, rather than focusing on note-taking in MR learning systems. In the HoloDoc system developed by Li et al. (2019), note-taking plays a secondary by supporting interactions between physical and digital papers, not within an immersive learning context. Similar studies, such as that conducted by Azizo et al. (2022), explore MR-related technologies, like marker-based AR, yet do not address the need for note-taking support in MR learning environments. Lastly, Yang and Chiu (2014) present note-taking techniques directly within a 3D learning environment, though recent advances in MR systems and digital inking technologies suggest alternative methods to the finger-writing techniques they introduced. Therefore, we aim to conduct a design study using participatory approaches to investigate user needs and design solutions to support note-taking in MR learning environments.

4.3 Design Workshop

To investigate stakeholder needs for supporting note-taking in MR learning environments, we conducted a group design workshop aiming to derive practical and valuable design goals for note-taking support beneficial to learners, educators, and developers of immersive learning systems.

Six students (2 female, 4 male; mean age 23.5, SD = 2.57), enrolled in an HCI course, participated. The group was intentionally diverse, including a student AR developer and a lab teaching assistant, to represent varied stakeholder perspectives relevant to MR learning system design and implementation in educational settings. The number of participants was chosen to facilitate in-depth qualitative interaction and ensure diverse perspectives were heard.

The workshop was structured into three phases, with a focus on system testing, brainwriting, and focus group discussion:

Phase 1: Parallel System Testing & Initial Group Reflection (30 minutes). Participants were invited as a group and introduced to the concept of learning with mixed reality. Each participant individually explored existing immersive learning systems on a Microsoft HoloLens 2 headset (including HoloPatient (GigXR, 2025) and HoloStudy (Group, 2025)) to establish a baseline understanding of current MR learning experiences. Participants interacted with these systems in parallel within a shared space and were instructed to actively identify challenges in MR learning environments, especially potential areas where note-taking could enhance learning. During this exercise, participants were asked to think aloud, with sessions audio-recorded for accuracy and comprehensive data collection. Notes were also taken by the facilitator (first author) to capture key observations. After individual explorations, the group reconvened to share observations and reflect on note-taking needs. This shared reflection established a common experiential foundation for collaborative ideation in subsequent phases.

Phase 2: Group Brainwriting for Collaborative Idea Generation (20 minutes). Participants engaged in a group brainwriting session using the 6-3-5 method (VanGundy, 1984; Wang and Chang, 2017), a structured approach effective for idea elicitation in HCI (Boy, 1997; Muller and Kuhn, 1993). This involved three rounds of iterative, silent idea generation on worksheets passed between participants. This approach fosters inclusive discussion and diverse idea generation, accommodating various stakeholder needs.

Phase 3: Refining Perspectives in Focus Group (15 minutes). Ideas from brainwriting were explored in a facilitated discussion focusing on the perspectives of learners, educators, and developers. This exchange focused on usability, integration, feasibility, and practicality across stakeholder groups. The session was recorded, and the facilitator transcribed and conducted a thematic analysis to derive key findings. Themes were identified using a top-down approach, involving iterative review and coding to ensure robust analysis. Analysis of workshop outputs involved summarizing audio recordings and synthesizing key themes from participant discussions and brainwriting sheets.

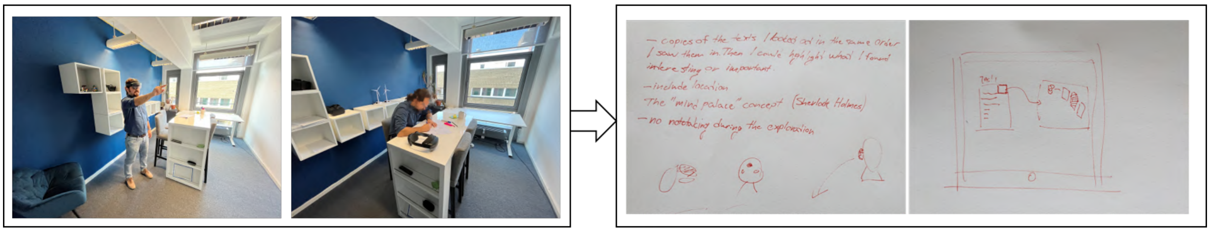

Figure 4.1: Phase 1 of the design workshop and ideas produced by participants on the brainwriting worksheets in phase 2

4.3.1 Findings

The design workshop provided valuable insights into user needs for note-taking in MR learning, revealing key themes through analysis of participant discussions and brainwriting outputs. These themes reflect not only learner preferences but also address the practical needs of educators and developers who integrate such features into immersive learning systems. Key findings include the need for intuitive and adaptable note-taking support, contextual anchoring of notes across devices, and workload and attention management in MR learning environments.

Need for Intuitive and Adaptable Note-Taking Support Participants underscored the necessity for note-taking support that feels intuitive and adapts to individual learning styles, recognizing the novelty of MR learning environments. For instance, P2 noted the overwhelming nature of “learning how to take notes while coping with interactive content exploration using the HoloLens”. This highlights the need for systems that reduce initial cognitive load and streamline the note-taking process within MR. Additionally, participants emphasized the importance of aligning MR note-taking with their existing note-taking habits. Many participants were already using digital note-taking tools: three (P2, P3, P6) used tablets with digital pens, while two (P1, P5) employed multiple devices (laptops and tablets) for organizing notes. Participant P6 also expressed a preference for the “feeling of writing on paper”. Consistently, participants suggested leveraging familiar devices like tablets for MR note-taking. This preference extends beyond individual learners; educators could benefit from solutions that integrate with students’ existing devices, promoting “a wider adoption and easy integration in regular lectures” (P6). From a developer’s perspective, supporting interactions with external mobile devices can “enhance accessibility and welcome a broader user base” (P3).

Need for Contextual Anchoring of Notes Across Devices Participants frequently indicated the importance of capturing and synchronizing contextual information from the immersive MR environment to note-taking spaces, particularly on secondary devices. Across all note-taking preferences (digital, paper-based, multi-device), the importance of context was emphasized. P3 and P4 noted that MR learning content’s interactive and engaging nature “cannot be captured easily” using traditional methods like photos or screenshots. This underscores the necessity for dynamic and interactive contextual information transfer. The workshop solutions, which often involved tablets (see Fig. 4.1 for participant ideas), implicitly addressed this need through cross-device functionality. This aligns with Kiewra and Frank (1988)’s emphasis on embedding notes within learning content. From an educator’s perspective, contextual anchoring is crucial for students to effectively review and connect notes to specific learning moments in MR. For developers, providing robust APIs for contextual data export and cross-device synchronization is key to creating valuable and usable note-taking support. The identified need hints at exporting dynamic learning material from MR to external note-taking devices, enabling learners to take context-rich notes using their preferred strategies.

Need for Workload & Attention Management in MR Learning Participants highlighted the significant mental workload associated with MR learning. As P2 stated, MR is “different and takes a while to get used to and I need to focus a lot”. Beyond the general workload, the novel format of information presentation in MR can make it difficult to track details or realize missed information, necessitating attention management support (P3, P4). Participants viewed note-taking as pivotal in managing workload. Moreover, attention-aware heatmaps could directly aid attention management by visually indicating learners’ gaze focus during MR sessions. Integrating these heatmaps with note-taking tools would enable learners to better manage workload by offering insights into their learning process and “guide the review of notes towards attention gaps" (P1). This integration also benefits educators by fostering student self-regulation and assists developers by reinforcing the importance of attention-aware feedback for creating usable MR learning experiences.

4.3.2 Articulating Design Goals

Based on these identified user needs and considering the perspectives of learners, educators, and developers, we formulated the following design goals for note-taking support in MR learning systems. These goals aim to develop solutions that are user-friendly, practically implementable, and scalable within educational contexts.

- **DG1: Cross-Device Flexibility for Personalized Note-Taking:** The system should support flexible note-taking workflows across devices, allowing users to leverage familiar tools (e.g., tablets, laptops) for note creation while interacting with MR learning content. This includes accommodating diverse note-taking preferences and existing digital workflows.
- **DG2: Context-Aware Information Transfer for Enhanced Review:** The system should enable seamless capture and transfer of contextual information from the MR learning environment to external note-taking spaces. This context-aware information transfer might include attentional information derived from eye tracking data, allowing learners to understand not only what content was presented but also what aspects captured their attention.

4.4 Development of GazeNote

Based on the identified design goals, the *GazeNote* system was developed with a modular architecture comprising three main components:

- A web application programmed with Angular, which enables learners to take notes and receive contextual information (such as screenshots from the MR environment) on mobile devices like tablets. The note-taking interface supports handwritten notes using a digital pen (e.g., Apple Pencil) alongside the contextual information.

- A Unity package that extends existing MR learning systems, installable on MR headsets (e.g., HoloLens). This component captures learners' field-of-view as screenshots and records eye tracking data to generate heatmaps of learners' attention distribution. Both the screenshots and heatmaps can be sent as image files via HTTP requests using the Unity API.

- A web connector server implemented with Node.js, which manages the connection between the web application and the MR learning system.

This modular design offers flexibility and adaptability to various setups. Figure 4.2 illustrates the system architecture and the utilities of its different components.

We designed and implemented the *GazeNote* system as described below. First, we developed a web application for note-taking. This application is platform-independent, providing flexibility to learners with different device preferences. Specifically, for our study, we utilized an Apple iPad Pro and Apple Pencil to enable handwritten note-taking via the web application. The front-end user interface presents users with a blank workspace for writing notes, which can be saved for later review (Fig. 4.3-1).

In addressing the second design goal, we implemented a synchronization function between the MR head-mounted display and the web application. Within the MR learning environment, learners can capture a screenshot of their currently viewed material using hand gestures, speech, or gaze-based interactions. Each learning item (text or images) in the MR system features a virtual button for triggering these actions in the HoloLens through multimodal interactions: gaze fixation over one second (as recommended in the MRTK[1]), near interaction with hand, and performing the air-tap gesture. Activating this button sends a screenshot of the relevant learning item to the web application on the tablet, where it appears atop the note-taking input field (See Fig. 4.3-2). This feature preserves contextual information, enhancing the note-taking experience by maintaining the connection with the learning environment.

Additionally, we integrated an eye tracking feature within the screenshot-capturing function. Prior research highlights the benefits of integrating eye tracking in MR systems, particularly for supporting attention management and coping with cognitive overload (Rappa et al., 2022). Consistent with this, we developed a feature that not only captures the field-of-view as a screenshot sent to the tablet but also enables an overlay displaying learners' visual attention distri-

[1]Mixed Reality Toolkit developed for HoloLens: https://github.com/microsoft/MixedRealityToolkit-Unity

bution as heatmaps, aggregated from eye tracking data. Figure 4.3-3 demonstrates how the heatmap is displayed in the web application to support note-taking.

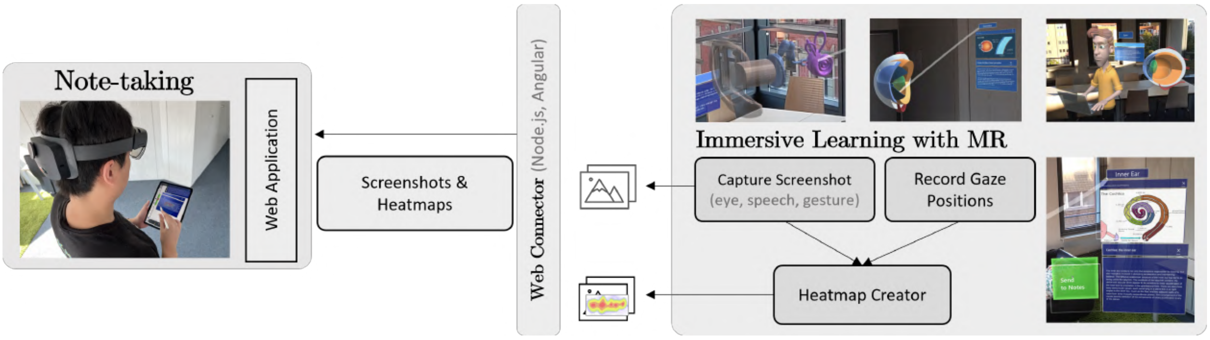

Figure 4.2: System Architecture of *GazeNote*. Diagram illustrating data flow from MR headset (screenshots captured with gaze or gesture interaction, heatmaps generated based on eye tracking data) to tablet-based note-taking application via web connector

For our study, the system was deployed as an extension to an existing MR learning system on the topic of human information processing. We hosted the web connector and web application on a local network to address data privacy concerns. Although we used Apple devices (iPad and Apple Pencil) for note-taking, *GazeNote* is compatible with other mobile devices on different operating systems.

4.5 System Evaluation

We evaluated *GazeNote* in a user study using a mixed-methods approach. The purpose of the evaluation was to investigate user experience and preferences regarding different note-taking techniques when using *GazeNote*.

4.5.1 Participant Information

We recruited 18 participants (8 female, 10 male), each participating in individual sessions. These participants were students from a local university, representing diverse academic disciplines such as industrial engineering, mechanical engineering, and biochemistry (See Appendix 7). The participants were aged between 21 and 30 years (M = 24.8, SD = 2.60). None of the participants had been involved in the initial design workshop, which ensured unbiased perspectives during the system evaluation.

Additionally, a subset of participants (7 out of 18) reported previous experience with immersive technologies, used in contexts such as personal entertainment, training, and museums. Two participants within this group had prior experience with immersive learning systems, specifically a medical training application on HoloLens. Including participants with previous immersive technology experience was valuable for mitigating the novelty effect often associated with evaluating new MR systems, thereby allowing a more valid evaluation of GazeNote's usability and impact on learning behavior.

4.5.2 Note-taking Scenarios

We developed and evaluated three distinct note-taking techniques with *GazeNote* (See Fig. 4.3). These techniques, each used in different evaluation scenarios, are detailed below:

- **Scenario 1 (S1): Handwriting Only.** Participants used an iPad and Apple Pencil to take handwritten notes on a blank webpage. Importantly, there was no direct connection or data transfer between the Mixed Reality (MR) learning environment and the tablet in this technique (Fig. 4.3-1).

- **Scenario 2 (S2): Handwriting with Screenshots.** Participants again used the iPad and Apple Pencil. However, in Scenario 2, participants could capture screenshots of the MR learning content using multimodal interaction (gaze, speech, or gesture) and transfer these screenshots to the iPad. This enabled the integration of contextual visual information directly into their notes (Fig. 4.3-2).

- **Scenario 3 (S3): Handwriting with Screenshots and Heatmaps.** This scenario built upon Scenario 2 by incorporating gaze-aware attention heatmap overlays, as detailed in Section 4.4. Participants could capture screenshots of the MR learning content with these visualizations, reflecting their gaze patterns in the MR environment. These enhanced screenshots, along with handwritten note-taking capabilities, were available on the iPad (Fig. 4.3-3).

The hardware setup for all scenarios included a HoloLens 2 MR headset for experiencing the learning content and an iPad with Apple Pencil for the note-taking interface.

Figure 4.3: Three note-taking techniques introduced during evaluation: 1) Handwriting only on a blank sheet; 2) Handwriting with screenshots; 3) Handwriting with screenshots and heatmap overlays.

4.5.3 Task and Procedure

Participants were recruited voluntarily and provided informed consent. They were invited to a dedicated learning space at the local university, the same location where the design workshop was conducted. Initially, participants were briefed on the study's goal, which was to investigate note-taking strategies in immersive learning and evaluate the proposed GazeNote system with three note-taking techniques. Participants then utilized *GazeNote* in three different scenarios.

To mitigate order effects from the within-subjects design, we applied complete counterbalancing for the presentation order of the three note-taking scenarios. With six possible orders (S1, S2, S3), each participant was randomly assigned to one of these orders, with three participants completing each. Consequently, each scenario was presented in every possible position (first, second, third) an equal number of times among the 18 participants.

For each scenario, participants engaged with a different learning module focused on human information processing. Three learning modules from the existing MR learning system were used, each covering distinct topics with comparable difficulties. To ensure that the learning content itself did not affect the results, we systematically rotated the assignment of learning modules to note-taking scenarios across the counterbalanced orders. Each module was matched with every scenario (S1, S2, S3) in two of the six orders, achieving a balanced distribution of content across conditions.

In each scenario, participants were instructed to explore the assigned learning content and take notes at their own pace using the designated technique. Before starting each scenario, the eyetracker on the HoloLens 2 was calibrated for each participant to ensure accurate gaze tracking during content exploration and note-taking. Following each scenario, and before moving to the next, participants completed the NASA-TLX and UEQ-S questionnaires to assess their subjective workload and user experience for the specific technique.

After completing the scenarios, an interruption task was introduced to simulate a realistic learning environment. Adapting the LEGO® Serious Play® methodology, participants constructed a tower using LEGO® blocks while articulating their construction strategy (Kristiansen and Rasmussen, 2014). This task, lasting an average of 19.7 minutes (SD = 4.7 minutes), created a time gap between the note-taking experience and the subsequent review phase, mirroring real-world learning where review follows a pause. A dedicated note review phase was facilitated to elicit informed and reflective feedback. Participants reviewed all notes taken across the three scenarios at their own pace using the iPad. This review, averaging 4.6 minutes (SD = 3.5 minutes), enabled participants to reassess their notes and evaluate their relevance and effectiveness for learning and recall, fostering accurate rankings and insightful feedback during the subsequent interviews.

Finally, we conducted semi-structured interviews to collect qualitative data on participants' experiences with each scenario and note-taking technique. Interviews probed opinions on each technique, elicited rankings of the three methods (See Fig. 4.8), and explored participants' typical note-taking strategies (See Appendix 7 for a detailed evaluation protocol and interview guidelines).

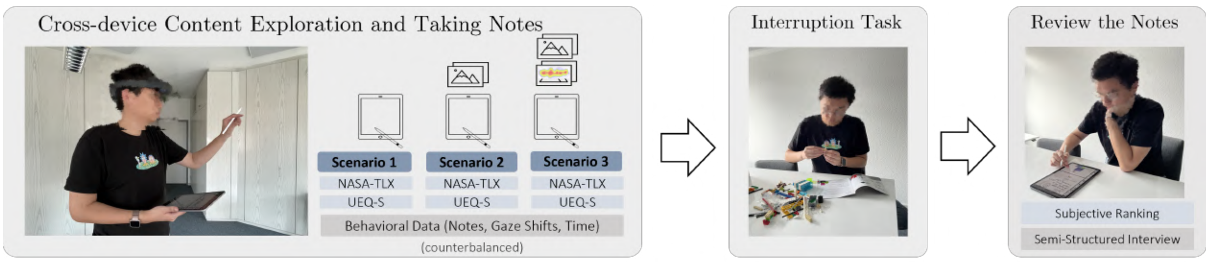

Figure 4.4: Evaluation protocol for *GazeNote*. Participants engaged in cross-device content exploration and note-taking across three scenarios, followed by interruption and review. Mixed-methods evaluation included NASA-TLX, UEQ-S, behavioral analysis, and interviews.

4.5.4 Data Collection

During the study, we evaluated mental workload and user experience using established questionnaires. Participants completed the NASA-TLX (Hart, 2006), rating six workload dimensions on 100-point scales (5-point increments). User experience was assessed with the UEQ-S, employing a 7-point Likert scale (Schrepp et al., 2017).

In addition to these subjective measures, we collected objective behavioral data for each scenario to provide a nuanced understanding of user interactions. Specifically, we quantified:

- **Number of notes taken:** For each scenario, we recorded the total number of digital notes, providing a measure of note-taking activity.
- **Word count of notes:** The notes from each scenario were analyzed, and word counts were manually tallied during data analysis to assess note richness and detail.
- **Cross-device gaze shifts:** To understand attention allocation between learning content and note-taking, we manually tracked and recorded instances when participants shifted their gaze between the HoloLens 2 display and the iPad.
- **Interaction time:** We measured the total time spent engaging with the *GazeNote* system in each scenario, reflecting engagement duration.

Finally, participants' preferences and opinions on the three note-taking scenarios were gathered through semi-structured interviews. The study was reviewed and approved by the university's institutional review board (IRB) following the ethical approval process. All collected data and analyses are available in our open data repository [2].

[2] Our OSF Open Data Repository: `https://osf.io/5he24/?view_only=3ca034beee08451d9bbcbe6ae91a50cb`

4.6 Results

4.6.1 Survey Results

Mental Workoad

A repeated-measures ANOVA was conducted to examine the effect of different note-taking techniques (S1, S2, S3) on participants' mental workload, as measured by the NASA-TLX survey. Before the analysis, the normality of the data was assessed. Shapiro-Wilk tests and visual inspection of Q-Q plots indicated that the data for each condition were approximately normally distributed and linear, satisfying the assumptions for ANOVA. Mauchly's test further confirmed that the assumption of sphericity was not violated (W = 0.922, p = .521), allowing for a reliable interpretation of the ANOVA results.

For overall workload, there was a statistically significant main effect of the note-taking technique on overall workload, as suggested by the averaged NASA-TLX scores across all six dimensions ($F(2, 34) = 7.278$, $p < 0.01$), with a small to medium effect size ($\omega^2 = 0.041$). This indicates that different note-taking techniques significantly impacted the overall workload perceived by participants.

Additionally, when each NASA-TLX dimension was examined individually, a significant effect of note-taking techniques was observed specifically in the performance dimension ($F(2, 34) = 5.563$, $p = 0.008$, $\omega^2 = 0.134$). This medium-to-large effect size indicates a meaningful difference in perceived performance depending on the note-taking technique used. Post hoc analysis with Holm-corrected pairwise comparisons revealed that performance in S3 was significantly higher than in both S1 (Mean Difference (MD) = 18.667, SE = 5.967, Cohen's d = 1.028, $p = 0.018$) and S2 (MD = 13.667, SE = 5.057, Cohen's d = 0.753, $p = 0.03$). However, the difference in perceived performance between S1 and S2 was not statistically significant (MD = 5, SE = 6.288, Cohen's d = 0.275, $p = 0.437$). No statistically significant differences were found among scenarios for the other five NASA-TLX dimensions: mental demand, physical demand, temporal demand, frustration, and effort.

Considering the limited sample size and aiming for a complementary perspective on the impact of note-taking techniques on mental workload, a Bayesian repeated-measures ANOVA was conducted following guidelines by Doorn et al. (2021). This analysis employed a multivariate

Cauchy prior for the effects. The Bayes factor for the performance dimension indicated strong evidence supporting the alternative hypothesis, with the model including note-taking scenarios being more likely to explain the data than the null model ($BF_{10} = 11.411$). This Bayesian analysis supports the frequentist ANOVA findings, suggesting a substantial influence of note-taking techniques on perceived performance.

Bayesian post hoc comparisons, corrected for multiple testing, further clarified pairwise differences between scenarios. These comparisons revealed moderate evidence for differences in perceived performance between S1 and S3 ($BF_{10,U} = 7.964$) and between S2 and S3 ($BF_{10,U} = 3.766$). In both instances, the Bayes factors suggest that the data are at least three times more likely under the alternative hypothesis of a difference compared to the null. Conversely, the comparison between S1 and S2 produced a Bayes factor of 0.321 ($BF_{10,U} = 0.321$), indicating anecdotal evidence favoring the null hypothesis. For a comprehensive assessment, we also conducted Bayesian repeated-measures ANOVAs for the remaining five dimensions of the NASA-TLX. The findings indicated that the data were better predicted by the null model for these dimensions, implying that the note-taking scenarios did not significantly affect these facets of workload.

In conclusion, the Bayesian analysis of the performance dimension robustly supports the idea that the different note-taking scenarios have a differential impact on perceived performance. Notably, Scenario 3 demonstrates a higher perceived performance compared to both S1 and S2. Although the performance comparison between S1 and S2 remains inconclusive, the overall findings highlight that the introduction of gaze-aware heatmaps in Scenario 3 is linked to enhanced perceived performance during note-taking tasks.

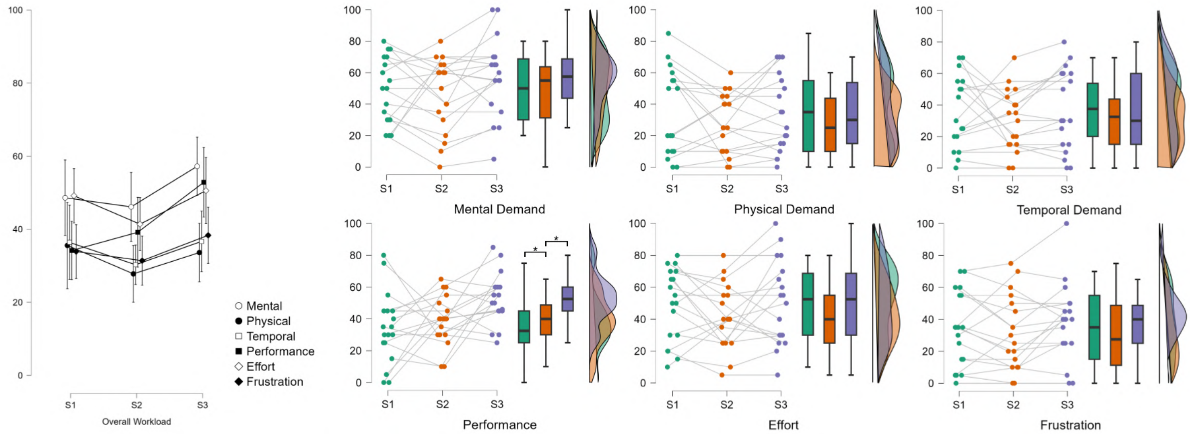

Figure 4.5: Raincloud plots of NASA-TLX sub-scale scores and overall workload for three *GazeNote* note-taking scenarios.

User Experience

The UEQ-S data provide insights into user perceptions across scenarios S1, S2, and S3, with potential trends emerging in both pragmatic and hedonic quality.

For pragmatic quality, Scenario S2 exhibited the highest mean score (M = 1.375, SD = 1.579, 95% CI [0.645, 2.105]), followed by S1 (M = 0.931, SD = 1.512, 95% CI [0.232, 1.629]), while S3 showed the lowest mean score (M = 0.708, SD = 1.935, 95% CI [-0.186, 1.602]). The overlapping confidence intervals suggest caution in interpreting these differences based on descriptive statistics alone. A repeated-measures ANOVA on pragmatic quality scores showed a statistically significant effect between note-taking scenarios ($F(2, 34) = 3.523$, $p = 0.041$).

In terms of hedonic quality, all scenarios reported satisfactory mean scores, with S1 being the highest (M = 1.806, SD = 1.042, 95% CI [1.324, 2.287]), closely followed by S2 (M = 1.597, SD = 1.164, 95% CI [1.060, 2.135]) and then S3 (M = 1.472, SD = 1.274, 95% CI [0.883, 2.061]). The considerable overlap in confidence intervals is reflected in the repeated-measures ANOVA on hedonic quality scores, which did not reveal a statistically significant effect of note-taking scenarios ($F(2, 34) = 0.821$, $p = 0.449$).

The overall UEQ-S scores suggest a trend favoring Scenario S2 (M = 1.486, SD = 1.135, 95% CI [0.962, 2.010]), with S1 slightly lower (M = 1.368, SD = 1.186, 95% CI [0.820, 1.916]) and S3 being the lowest (M = 1.090, SD = 1.373, 95% CI [0.456, 1.725]). Despite these trends, the response distribution across the UEQ-S items indicates a similar level of user experience across all note-taking techniques (see Fig. 4.6).

In summary, the descriptive analysis of the UEQ-S points to subtle variations in perceived pragmatic and hedonic qualities and overall user experience across the scenarios, with potentially lower pragmatic quality in S3. However, given the overlap in confidence intervals and lack of statistically significant differences found in the ANOVAs, any interpretations based solely on mean differences should be approached cautiously.

4.6.2 Behavioral Analysis

We conducted a behavioral analysis to objectively assess how the different note-taking techniques influenced user interactions within *GazeNote*. Specifically, we analyzed four key behavioral measures: the number of notes taken, the word count of the notes, the total interaction

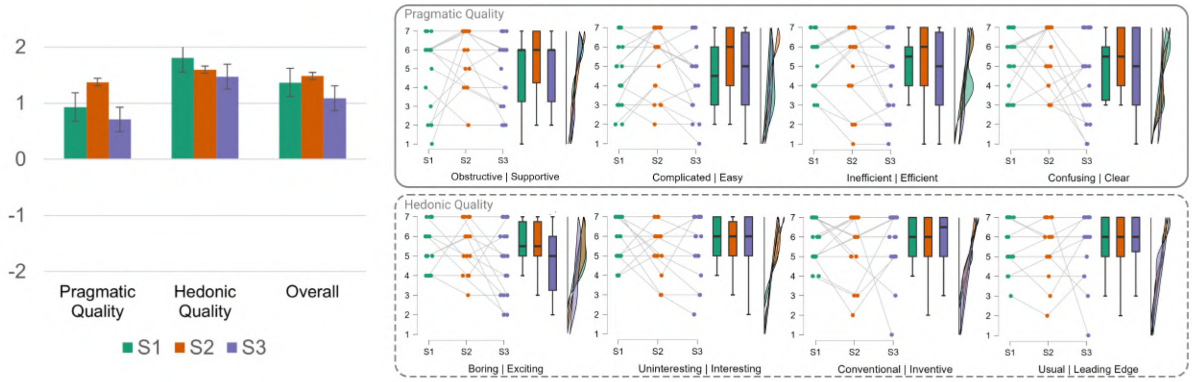

Figure 4.6: UEQ-S ratings for GazeNote note-Taking scenarios. The figure compares Pragmatic Quality, Hedonic Quality, and Overall User Experience scores across three note-taking techniques (S1, S2, S3).

time in each scenario, and the number of gaze shifts between the MR learning content and the note-taking interface (see Fig. 4.7). For each measure, both frequentist and Bayesian repeated-measures ANOVA were performed.

Number of Notes Taken

The analysis revealed no significant differences in the number of notes taken across the three note-taking scenarios, suggesting that the overall quantity of notes was consistent regardless of the technique employed.

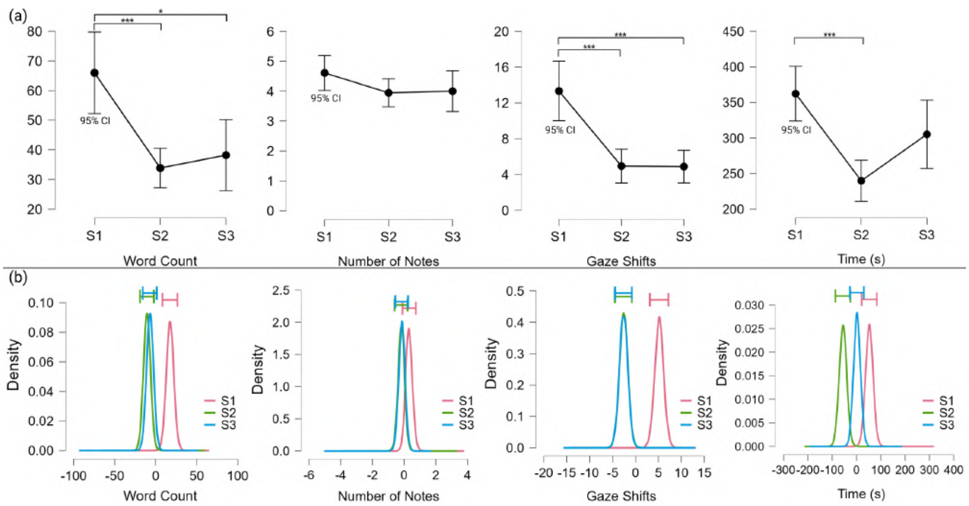

Figure 4.7: Behavioral analysis results. (a) Descriptive plots showing means and 95% Confidence Intervals for Word Count, Number of Notes, Gaze Shifts, and Time (in seconds) across the three note-taking scenarios (S1, S2, S3). Significant differences from frequentist RM ANOVA are indicated by asterisks. (b) Model-averaged posterior distributions from the Bayesian RM ANOVA for each behavioral measure and scenario.

Word Count of Notes

The word count analysis showed a significant effect of the note-taking technique ($F(1.371, 23.302) = 10.765$, $p = 0.001$, $\omega^2 = 0.150$), using Greenhouse-Geisser correction for sphericity. Post-hoc analysis indicated that participants in S2 and S3 produced notes with significantly lower word counts compared to S1, suggesting more concise note-taking when incorporating screenshots or heatmaps. However, no significant difference was found between S2 and S3.

The Bayesian analysis found that the data were 126.278 times more likely under the model including note-taking technique as a predictor for word count, compared to the null model. Post hoc comparisons revealed decisive evidence for differences in word count between S1 and S2 (posterior odds = 74.187) and moderate evidence between S1 and S3 (posterior odds = 3.031). Comparisons between S2 and S3 provided anecdotal evidence in favor of the null hypothesis (posterior odds = 0.192, $\mathrm{BF}_{10,U}$ = 0.327).

Cross-device Gaze Shifts

The number of gaze shifts between the MR content and the iPad showed significant differences across scenarios ($F(1.155, 19.628) = 17.569$, $p < .001$, $\omega^2 = 0.324$). Post-hoc analysis (Holm-corrected) revealed significantly fewer gaze shifts in S2 and S3 compared to S1 (S1 vs S2: $t(17) = 4.253$, $p_{holm} = 0.001$; S1 vs S3: $t(17) = 4.340$, $p_{holm} = 0.001$). There was no significant difference in gaze shifts between S2 and S3 ($t(17) = 0.089$, $p_{holm} = 0.930$). This indicates that screenshots, with or without heatmaps, reduce the need for frequent attention shifts.

Bayesian RM ANOVA indicated that the data were 13,613.579 times more likely under the model including note-taking technique as a predictor for gaze shifts, compared to the null model. Decisive evidence supported differences between S1 and S2 (posterior odds = 37.766) and S1 and S3 (posterior odds = 44.445). Comparisons between S2 and S3 again showed anecdotal evidence favoring the null hypothesis (posterior odds = 0.143, $\mathrm{BF}_{10,U}$ = 0.244).

Interaction Time

The total interaction time analysis also revealed significant differences ($F(2, 34) = 10.814$, $p < .001$, $\omega^2 = 0.208$). Descriptive statistics indicated that interaction time was longest in S1

(M = 362.28 seconds, SD = 114.47), followed by S3 (M = 305.22, SD = 73.01), and shortest in S2 (M = 239.89, SD = 86.61). Post-hoc analysis (Holm-corrected) showed significantly shorter interaction times in S2 compared to S1 ($t(17) = 6.577$, $p_{holm} < .001$, Cohen's $d = 1.316$). No significant difference in interaction time was found between other scenario pairs.

The Bayesian RM ANOVA indicates that the data are 231.666 times more likely under the model that includes different note-taking techniques as the predictor for interaction time, compared to the null model. Post hoc comparisons of S1 vs. S2 revealed posterior odds of 2566.462 against the null hypothesis, indicating decisive evidence in favour of the alternative hypothesis. When comparing S1 and S3, there was anecdotal evidence in favour of the null hypothesis (posterior odds = 0.538, $BF_{10,U}$ = 0.916). For the comparison of S2 and S3, there was also anecdotal evidence in favour of the null hypothesis (posterior odds = 1.403, $BF_{10,U}$ = 2.388).

4.6.3 Thematic Analysis

We analyzed the interview data using inductive thematic analysis, guided by the framework proposed by Braun and Clarke (2006b), a well-established qualitative method in HCI research known for its rigor and systematic approach (Fikar et al., 2018; Heyko and Flatla, 2021). The interviews averaged 11.53 minutes (SD = 4.14), with transcripts averaging 930.36 words (SD = 484.74).

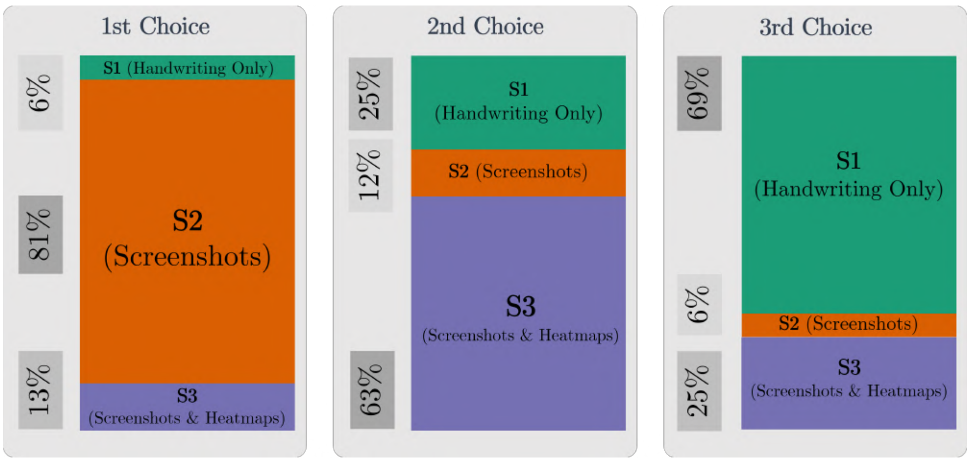

Figure 4.8: The subjective ranking of the three note-taking scenarios reported by the participants

In the initial coding phase, the first and second authors independently reviewed the transcripts, identifying initial codes to capture salient meanings and patterns within the data. This iterative process involved reading transcripts carefully to allow themes to emerge organically rather than imposing pre-conceived categories. Following this independent coding, the two coders collaboratively reviewed their initial codes, comparing interpretations of 164 selected quotes. Through discussion and refinement, coding discrepancies were addressed, leading to a consensus on 20 distinct codes.

Building on this collaboratively developed code set, we proceeded to the thematic searching and reviewing phases of Braun and Clarke's framework. The 20 codes were iteratively grouped into eight clusters through joint discussion and consensus. Subsequently, in the theme review process, these clusters were refined and synthesized, resulting in the identification of three overarching themes. The robustness of our thematic structure was ensured by rigorously reviewing the emergent themes, constantly revisiting the initial codes and the raw interview data, as recommended by Braun and Clarke (2006b, 2012). The final thematic structure is visualized in a Sankey diagram (Figure 4.9), illustrating the relationships between codes, clusters, and themes.

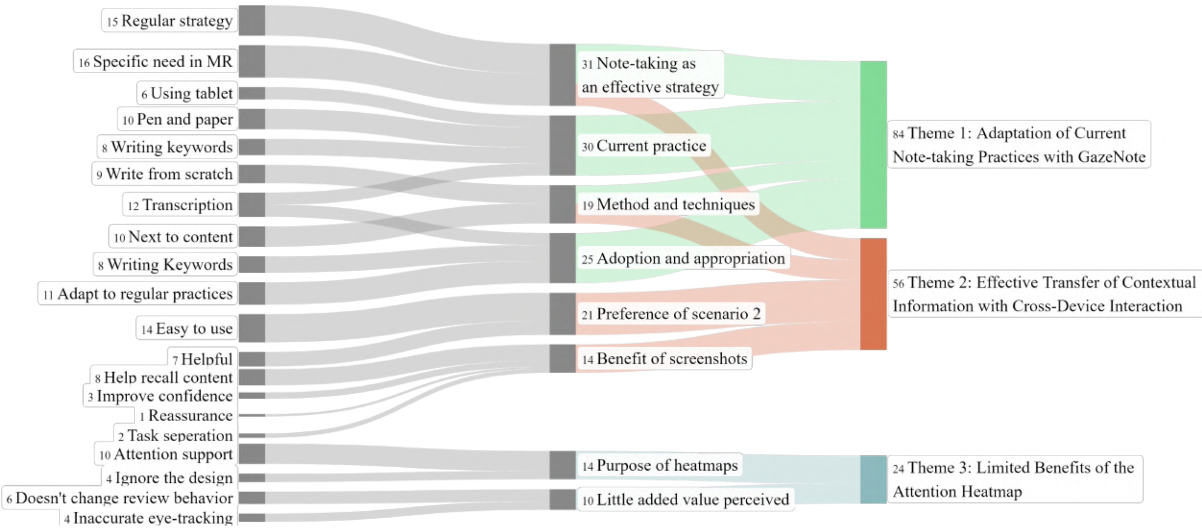

Figure 4.9: Visualization of the thematic analysis. Numerical values indicate the occurrence frequency within the transcripts.

Theme 1: Adaptation of Current Note-taking Practices with GazeNote

Most participants (15 out of 16) identified effective note-taking as essential in both regular and MR learning experiences. They reported using various devices and formats: six participants used tablets or laptops for digital note-taking, while ten preferred pen and paper for its tactile feel, as expressed by P1. Participants reported two main approaches to organizing notes: some kept notes alongside learning materials for direct linking (P4), while others preferred separate documents for categorized, structured note-taking (P11). Three participants (P11, P6, P3) combined these approaches.

Participants tended to adapt *GazeNote* to their usual note-taking habits. Six participants mentioned manually transcribing sentences from learning materials to aid memorization (P9), favoring S1's familiarity. Even in S2 and S3 scenarios, they manually rewrote content from screenshots. Conversely, some participants recorded only keywords for easier reference (P2). This led to different note-taking patterns (see Fig. 4.10). Participants appreciated that *GazeNote* accommodated their styles (P12), despite diverse methods and preferences.

Figure 4.10: Examples of notes taken by three participants in S2 on the same topic, showing three practices: 1) using only screenshots; 2) adding keywords; 3) rewriting important sentences.

Theme 2: Effective Transfer of Contextual Information with Cross-Device Interaction

Scenario S2 was preferred by most participants (13 out of 16), described as "easy to use" (P9) and "most helpful" (P7). The ability to take gaze-based screenshots and transfer them to mobile devices was viewed as efficient, saving time by focusing on personal thoughts rather than rewriting existing content. During the review process, participants also report the benefits of having screenshots, as they help learners "immediately recall the meaning of my notes" (P3). In addition, taking contextual information as screenshots improves confidence during the learning process, P11 reported that this function gives him “a kind of reassurance”, as learners would still have access to the original content in the review process.

Many participants (8 out of 16) suggested that the cross-device interaction supports a seamless transfer of contextual information as screenshots from the MR environment to tablets, allowing for a convenient review process later. P12 mentioned that “I can completely focus on what I’m looking at and use eye movement to send (screenshots) and take notes on the iPad”. The quantitative analysis of NASA-TLX data shows similar results, indicating that S2 leads to the lowest mental and physical demands among the three note-taking scenarios. Furthermore, the participants also gave strong arguments for using tablets for receiving screenshots and taking notes, instead of accomplishing all learning activities in MR. P3 and P1 mentioned a sense of “task separation” that helped them reduce cognitive overload, as they completed the more demanding content exploration task using MR headsets while taking notes separately using a more familiar device such as an iPad.

In addition to the existing screenshot capture, participants also proposed other methods of capturing contextual information for cross-device note-taking. Marking text or selecting snippets of information in MR was suggested as a possible extension of our system. This would allow learners to precisely select the most relevant information from the learning environment and transfer it to their personal devices, rather than sending a screenshot of their entire field of view.

Theme 3: Potential Benefits of the Attention Heatmap While only two participants in our evaluation study preferred S3 with the heatmap function over the other scenarios, a notable number appreciated its potential. Some participants (P2, P9, P4) recognized the purpose of the heatmap in visualizing attention but initially saw limited added value compared to regular screenshots. P2 remarked: “I would say it’s a good feature but I think it’s a little bit like a gimmick. Sure, you can see where you’ve looked at but maybe the benefits are not very impactful”. Other participants (P1, P4, P7, P8, P14) felt the heatmap might not enhance the review process significantly, because they will “review everything (personal notes and the screenshots) anyway” regardless of what the heatmap shows.

Despite varied initial reactions, over half of the participants (10 out of 16) identified the heatmap’s potential as a supportive feature for future use, particularly in promoting self-reflection during note review. P3 notably highlighted the heatmap’s utility:

> when I first read a whole paragraph (in MR), there were a few sentences that I couldn't understand. So I went back to read them again. The heatmap reflects this (reading pattern). Then, when I write down my notes, I immediately recall that this is the important part and I should make some notes.
>
> And in another aspect, when I reviewed my notes, I could see, okay, I may have read this part twice or even three times. So I could see the complexity and review more carefully. (P3)

This commentary underlines the heatmap's dual potential to enhance not only the note-taking process but also the review phase by highlighting areas that garnered significant attention. The variations in engagement with the heatmap were evident in the notes taken by participants. While some used the feature as designed, others opted not to utilize it. Although some activated the heatmap without content reading, resulting in less informative overlays, this was often a deliberate choice based on their note-taking needs.

However, limitations in the current eye tracking technology hinder full adoption of the heatmap function. Some participants noted that the accuracy of the MR headset's eye tracking sensors could be improved, as the visualized attention did not always align with actual gaze patterns. Additionally, there were concerns about the use of eye tracking in real-world learning, where students feared its potential misuse for assessing academic performance by teachers.

4.7 Discussion

4.7.1 Interpretation of Results

Our quantitative analysis presents a nuanced understanding of note-taking support in MR learning environments. The NASA-TLX results, reinforced by interview feedback, reveal a general user preference for screenshot-based note-taking techniques (S2 and S3) over manual transcription (S1). Although no significant differentiation in overall mental demand was observed across scenarios—indicating that optimized learning experiences involve inherent cognitive engagement—participants reported subjectively perceived performance improvements with S2 and S3. The gaze-aware attention heatmaps in S3 were specifically noted for aiding attentional focus and enhancing perceived performance. While these heatmaps might introduce a marginal increase in temporal demand and information processing load, they were not perceived as detrimental to the overall workload. The Bayesian analysis adds nuance to these findings, recognizing the study's limited sample size while supporting the interpretation that integrating screenshots and attention-aware heatmaps offers tangible benefits.

UEQ-S scores demonstrated satisfactory pragmatic and hedonic qualities across all scenarios, suggesting that each method delivered an acceptable user experience. The slight decrease in pragmatic quality for S3 might stem from the additional cognitive effort required to process attention heatmap information. Behavioral data further underscore the advantages of S2 and S3; S1 exhibited significantly longer interaction times and more frequent gaze shifts compared to S2 and S3, indicating the higher cognitive demand of manual transcription and device switching. Conversely, the lower word count in notes taken during S2 and S3 suggests that the contextual information provided by screenshots reduced the need for extensive written elaboration, highlighting increased note-taking efficiency.

The qualitative analysis reveals a broad need for effective note-taking in MR learning scenarios, with *GazeNote* perceived as helpful in facilitating this process. Participants provided insightful feedback on the three scenarios presented, which aligns with existing research on note-taking benefits, such as those introduced by Kiewra and Frank (1988). These benefits include cognitive offloading during learning and the support for subsequent review. Such effects were particularly evident in S2 and S3, where participants could capture contextual information conveniently. Consequently, S2 and S3 were generally preferred over S1 (see Fig. 4.8). Between the two preferred scenarios, the heatmap feature in S3, although enhancing participant performance, might have contributed to increased complexity, leading most participants to favor S2 over S3 (see Fig. 4.8). Based on these findings, we propose the following implications for future research.

4.7.2 Design Implications and Requirements Elicitations

Achieving the Balance between Support and Autonomy for Note-taking

In developing *GazeNote*, we aimed to enhance note-taking in MR learning by using cross-device interaction and gaze-aware features. Learners were supported in transferring learning content and contextual information from the MR environment to tablets using gaze, speech, or hand gestures. Additionally, attention heatmaps were offered as optional resources for note-taking and review. Our evaluation revealed nuanced user perceptions regarding these support mechanisms. While the screenshot functionality for content capture was widely appreciated for preserving context, the attention heatmap overlay feature was underutilized and received mixed feedback.

A subset of participants recognized the potential of heatmap overlays for guided note-taking. However, the others viewed them as less helpful. Participants generally found the ability to extract screenshots from the MR environment sufficient for their needs. This indicates that learners' note-taking techniques in MR are largely self-directed, and overly prescriptive or excessive support mechanisms may be perceived as intrusive and counterproductive.

Participants also critically reflected on the readily available screenshot functionality. While valuing the ease of capturing contextual information, some cautioned that reduced cognitive effort might negatively affect memorization and deeper knowledge processing. A minority of participants indicated they felt more engaged and focused when they actively transcribed all notes themselves, as seen in the unassisted handwriting scenario S1. This highlights the potential for system support to inadvertently reduce learner agency and perceived control, potentially hindering confidence and increasing frustration.

Requirement 1: Provide Foundational, Non-Intrusive Support Mechanisms Note-taking systems for immersive learning should focus on core functionalities that facilitate basic information capture and organization, such as content extraction via screenshots. Avoid incorporating overly prescriptive or complex support features that could impede learner autonomy and introduce usability challenges.

Requirement 2: Balance System Guidance with Learner Agency: Interactive systems for immersive learning should strive to balance offering support for note-taking with preserving learner control over their individual learning processes. Developers should consider making support features adaptive to avoid intrusiveness and accommodate diverse learning preferences. Furthermore, when incorporating eye tracking for learning support, it is essential to address potential privacy concerns with users. It is important to distinguish between using eye tracking for supporting self-reflection (as in *GazeNote*) and its potential use as a supervisory metric. Providing transparency and control over eye tracking features can enhance the adoption of adaptive systems.

Adaptation to Individual Note-taking Preferences and Methods

Our study underscored the diverse range of note-taking preferences among participants in MR learning environments. Participants across all three scenarios adapted and appropriated the provided note-taking techniques to different extents. In S1 – handwriting only – strategies varied from verbatim transcription to summarizing content using keywords or paraphrasing. In S2 and S3, which were screenshot-based, some participants relied solely on screenshots for review, while others used them as a foundation for personal elaboration. Notably, in S3, where heatmaps were included, some participants ignored the heatmap feature entirely, defaulting to an S2 approach, whereas others selectively used heatmaps for self-reflection and guidance.

This heterogeneity in note-taking approaches highlights the influence of pre-existing learning habits. Participants' preferred scenarios and techniques often mirrored their established note-taking practices in traditional learning contexts. For instance, those who favored S1 often reported extensive handwritten note-taking in physical classrooms. Conversely, participants who heavily utilized screenshots in S2 frequently mentioned a preference for annotating learning materials alongside side notes. These observations emphasize that prior note-taking strategies significantly inform learners' preferences and behaviors in MR settings.

Requirement 3: Support Diverse and Personalized Note-taking Styles MR learning systems should be designed to accommodate a broad spectrum of note-taking approaches, ranging from free-form handwriting to screenshot-based annotation. These systems should offer flexible and adaptable tools that empower users to tailor their note-taking processes to suit their individual learning styles and established habits. Future MR note-taking solutions should integrate features that allow for personalization and adaptation, enabling users to customize the level and type of system support according to their preferences and learning strategies. This might include options to toggle features (e.g., heatmaps) on or off, choose from various formatting styles, or adjust the degree of automated assistance provided. Such personalization will cater to a wide array of learners, fostering engagement and effectiveness in MR learning environments.

Bridging In- and Ex-situ Learning with Cross-Device Interaction

Our cross-device interaction design, which utilizes a digital pen and tablet alongside an MR headset, significantly enhances both usability and note-taking efficiency within MR learning environments. User interviews highlighted a key advantage of *GazeNote*: the explicit separation of devices corresponding to distinct learning stages, combined with the inherent portability of notes via the tablet. This separation offers substantial benefits for both learners and MR system developers.

A primary challenge in MR learning is the increased cognitive burden resulting from the technology's novelty, the dense presentation of information, and potential usability issues (Charfi et al., 2009; Kockord and Bodensiek, 2021). By strategically offloading the note-taking task to a separate, familiar tablet, *GazeNote* provides a practical solution to mitigate cognitive overload directly within current MR learning systems. Developers can exploit this architecture to allow the MR headset to focus on delivering immersive and exploratory content experiences, thus facilitating development efforts and enhancing system performance. Moreover, delegating note-taking to a conventional and user-friendly tablet leads to a more natural and less overwhelming learning experience for users. The resulting tablet-based notes afford learners critical flexibility: easy review, seamless digital export (e.g., PDF), and printing for offline study, effectively bridging the gap between in-situ MR immersion and ex-situ reflection and continued learning.

We advocate for a holistic design paradigm that explicitly integrates critical learning stages—from immersive content engagement in MR to efficient in-situ note capture on a tablet, attention-informed self-reflection, and convenient ex-situ review — into a cohesive and continuous learning journey. Crucially, this cross-device note-taking framework is designed for straightforward integration as a modular extension into existing and future MR learning platforms, offering immediate and tangible benefits to educators, developers, and learners by enhancing usability, streamlining development, and fostering more effective and complete learning workflows.

Requirement 4: Leverage Cross-Device Architecture for Workload Distribution and System Extensibility By adopting a cross-device engineering approach, developers of immersive learning systems can optimally utilize the MR headset for immersive content delivery and exploration, while strategically employing external mobile devices (e.g., tablets, laptops) for auxiliary tasks such as note-taking and review. This approach can improve usability and reduce cognitive overload, while also promoting a modular system design that facilitates easier development, maintenance, and integration of specialized learning support features.

4.7.3 Limitations and Future Work

Several limitations of our study should be acknowledged. Firstly, our study was conducted with a limited sample size, warranting future research with larger participant groups to further validate our findings and enhance generalizability. Secondly, while *GazeNote* offers a functional cross-device note-taking experience, the tablet application was intentionally kept basic, prioritizing validation of core concepts over feature richness. Future iterations could explore richer feature sets within the tablet app, such as typing input besides handwriting, enhanced formatting options (stroke styles, fonts, highlighting), mind-mapping capabilities, and bringing notes back into MR. These features can cater to diverse user preferences and support more note-taking strategies.

Furthermore, it's important to note that we evaluated *GazeNote* as an extension to a specific, openly available MR learning artifact. While this demonstrates the feasibility of our cross-device approach, future research should investigate the generalizability and integration of *GazeNote* with a wider range of MR learning systems, across different educational domains (e.g., STEM, medical education) and learning scenarios (e.g., collaborative learning). This includes exploring the potential of *GazeNote* as a flexible and modular extension that can be readily adopted and adapted by MR engineers and content creators.

Finally, while our initial findings on gaze-based attention heatmaps are promising, further investigation is needed to optimize their design and empirically validate their impact on learning outcomes (e.g, with knowledge tests). Future work should also explore other input modalities, such as speech-to-text, to enhance accessibility and hands-free note-taking within MR. Interdisciplinary research, incorporating educational perspectives, can further investigate the cognitive and pedagogical benefits of *GazeNote* in MR learning, particularly regarding cognitive offloading, encoding effects, and knowledge retention.

4.8 Conclusion

In our study, we addressed the pressing challenge of integrating effective note-taking strategies within MR learning environments. Through the development of *GazeNote*, a novel cross-device system, we demonstrated a practical approach to enhancing MR learning by seamlessly combining tablet and MR headset interactions while innovatively leveraging gaze-aware features. Our human-centered evaluation provided crucial empirical insights, emphasizing the fundamental of note-taking for effective learning in MR and revealing key preferences and challenges. We found validation for basic note-taking functionalities, such as contextual information capture, while highlighting the nuanced and individual nature of note-taking strategies and the need for adaptable support mechanisms. Furthermore, our study yielded valuable insights regarding the design of non-intrusive support and the careful application of eye tracking technologies in this context. These findings collectively demonstrate the significant potential of cross-device and gaze-aware approaches to meet the critical need for robust note-taking support in MR learning.

Beyond user-centric insights, our work offers a comprehensive perspective for stakeholders involved in engineering interactive learning systems with MR. The *GazeNote* artifact serves as a concrete example of system extension, showcasing a concise cross-device architecture for enhancing MR learning. Our empirical evaluation lays the groundwork for future research and development by illuminating learner behaviors and cognitive considerations in immersive note-taking scenarios. Importantly, the actionable design requirements and implications derived from this study offer practical guidance for developers aiming to integrate note-taking support into existing and future MR learning platforms. This encompasses empowering learners with more engaging and effective learning experiences, streamlining development efforts for software engineers, and providing educators and content creators with adaptable tools to enrich their pedagogical practices within MR [3].

By advocating for human-centered solutions in immersive learning support, this research contributes to realizing the full potential of MR in fostering active, independent, and enriched learning experiences for diverse stakeholders in the evolving landscape of immersive education.

[3]GazeNote can be adopted as an extension to existing immersive learning systems, see our Open Code repository: `https://anonymous.4open.science/r/GazeNote_IJHCS-4274`

5 AttentiveLearn: Personalized Post-Lecture Support for Gaze-Aware Immersive Learning (Study IV)

Abstract: Immersive learning environments such as virtual classrooms in Virtual Reality (VR) offer learners unique learning experiences, yet providing effective learner support remains a challenge. While prior HCI research has explored in-lecture support for immersive learning, little research has been conducted to provide post-lecture support, despite being critical for sustained motivation, engagement, and learning outcomes. To address this, we present *AttentiveLearn*, a learning ecosystem that generates personalized quizzes on a mobile learning assistant based on learners' attention distribution inferred using eye-tracking in VR lectures. We evaluated the system in a four-week field study with 36 university students attending lectures on Bayesian data analysis. *AttentiveLearn* improved learners' reported motivation and engagement, without conclusive evidence of learning gains. Meanwhile, anecdotal evidence suggested improvements in attention for certain participants over time. Based on our findings of the field study, we provide empirical insights and design implications for personalized post-lecture support for immersive learning systems.

Keywords: Virtual reality, Immersive learning, Eye-tracking, Adaptive support

Published as: Liu, S., Feick, M., Bierhoff, L., & Maedche, A. (2026). AttentiveLearn: Personalized Post-Lecture Support for Gaze-Aware Immersive Learning. Proceedings of the CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3772318.3790667

5.1 Introduction

Using immersive technologies such as Virtual Reality (VR) to create engaging learning experiences is a broad research field known as immersive learning (Makransky and Petersen, 2021; Mystakidis and Lympouridis, 2023). Prior HCI research has investigated the design of immersive learning systems across various domains, ranging from digital fabrication (Radu et al., 2021b) to language learning (Shao et al., 2020). Different affordances and formats of immersive learning have been explored, including virtual laboratories (Radu et al., 2021b), simulations (Zhu et al., 2025), and virtual classrooms (Gao et al., 2021). While immersive learning environments have shown promise for better learning outcomes (Baceviciute et al., 2020; Makransky and Mayer, 2022), providing effective learner support in these environments remains a challenge (Petersen et al., 2021; Shen et al., 2025). A key setting where this challenge has been recognized is the immersive virtual classroom in VR (Gao et al., 2021; Han et al., 2022). Immersive virtual classrooms have emerged as a prominent example of immersive learning, both widely researched in HCI (Gao et al., 2021; Liu et al., 2025a; Liu et al., 2024) and increasingly adopted in practice (Moynihan, 2021). Prior work has examined design features such as embodied agents (Liu et al., 2024), large language model (LLM) integration (Tang et al., 2025), and annotation tools (Tsai et al., 2025) to optimize the learning experience. However, most of these efforts focus on in-lecture support, whereas post-lecture support for immersive virtual classrooms has received less attention.

Post-lecture support is well established as a critical component of learning success (Palmer et al., 2019). Educational research emphasizes that learning is a continuous and ubiquitous process extending beyond the classroom (Hwang and Tsai, 2011; Kuh, 1996). Techniques such as tutoring (D'mello and Graesser, 2012), quizzes (Raes et al., 2020), content reviews (Szafir and Mutlu, 2013), and other follow-up activities have been applied to improve learning outcomes. They have been shown to be particularly effective when tailored to individual learners' prior experiences and behaviors in lectures (Plooy et al., 2024; Ross et al., 2018; Szafir and Mutlu, 2013). For example, biosignal-based attention metrics have been used to drive adaptive content reviews for desktop-based learning systems (Szafir and Mutlu, 2013). We extend this approach to immersive virtual classrooms. In immersive learning, however, much of the individual learning behavior, preferences, and performance exhibited during the lecture is often lost, making it difficult to provide meaningful personalized support afterward. We identify this as a research gap: while personalized post-lecture support based on in-lecture metrics has shown potential to enhance motivation (Ross et al., 2018), engagement (Nuci et al., 2021), and learning outcomes (Ross et al., 2018), its role in immersive learning contexts remains underexplored, lacking a framework to guide the transfer of in-situ attention monitoring to ex-situ support.

To address this gap, we ground our framework and system design in two key streams of prior research. First, research has revealed that focused attention during lectures is a crucial cognitive factor for learning success (Keller et al., 2020). In HCI research, eye-tracking technologies have been widely used to support attention across diverse contexts (Duchowski, 2007; Roda and Thomas, 2006), including education (D'Mello et al., 2012; Hutt et al., 2017) and in VR systems (Han et al., 2022; Plopski et al., 2022). Yet, little research has explored how attention data can inform post-lecture support in immersive learning. Second, we identify quizzes as a pedagogically effective post-lecture support strategy. Quizzes are widely used to reinforce learning, provide self-assessment, and improve retention through the testing effect (Roediger III and Karpicke, 2006). They have also been shown to increase engagement (Raes et al., 2020). Moreover, adaptive and personalized quizzes have been explored as promising post-lecture scaffolds (Ross et al., 2018). However, prior work has not investigated using attention data from VR lectures as a personalization strategy for post-lecture quizzes.

Building on these rationales, we introduce a framework for attention-based post-lecture personalization in the context of immersive learning, instantiated with *AttentiveLearn*, a learning ecosystem that generates personalized quiz questions based on learners' attention distribution during VR lectures. The ecosystem integrates VR lectures with an existing mobile learning assistant application deployed at our local university. *AttentiveLearn* operationalizes the framework through three components: (a) a gaze-aware VR classroom that collects eye-tracking data during lectures, (b) a data processing pipeline that computes attention metrics and transfers them to an external server, and (c) a mobile learning assistant application that delivers post-lecture support.

To evaluate *AttentiveLearn*, we conducted a four-week between-subjects field study with 36 students. Each week, participants attended an on-site VR lecture on Bayesian data analysis. After the lecture, half of the participants used a variant of the mobile assistant with quizzes personalized using eye-tracking data collected during the lecture, while the other half used a version with randomly generated, non-attentive quiz questions. We collected survey data on motivation and engagement each week, as well as results from intermediate and final exams to evaluate learning outcomes. In addition, we analyzed interaction and chat logs of the learning assistant and conducted 10 semi-structured interviews to provide qualitative insights. Our findings suggest that the personalized assistant improved reported engagement and motivation throughout the study. There was also anecdotal evidence of improved attention for low-attention learners from the first to the last week. Learning outcomes showed improvement in the intermediate tests, though no significant differences were observed in the final test. In this work, we make the following three key contributions:

- Conceptualization of a novel learning ecosystem with a generalizable framework that bridges in-situ immersive learning with ex-situ support through attention-aware post-lecture quizzes.
- Integration of the attention-aware personalization pipeline into an existing mobile learning assistant, demonstrating feasibility in real-world educational settings.
- Empirical insights and design implications from a four-week field study ($n = 36$) investigating how personalized post-lecture support affects motivation, engagement, and learning outcomes.

5.2 Related Work

5.2.1 Immersive Learning Environments

Immersive learning refers to the use of technologies, such as VR and Augmented Reality (AR), to create engaging, situated learning experiences that support embodied cognition and enhance learning outcomes (Makransky and Petersen, 2021; Mystakidis and Lympouridis, 2023). Empirical evaluations of immersive learning systems reveal a broad range of learning benefits. For instance, a VR field study conducted by Petersen et al. (2021) demonstrated improved enjoyment and knowledge acquisition in the context of museum guides. Moreover, Radu and Schneider (2019) identified students' increased self-efficacy after using an immersive physics learning application. Other positive impacts include improved comprehension (Schnitzer et al., 2025), motivation (Thanyadit et al., 2023), and achievement across contexts (Radianti et al., 2020). However, research also reports mixed or contradictory effects (Makransky et al., 2021;

Radianti et al., 2020), which indicate unaddressed user challenges in immersive learning environments. Therefore, recent systematic reviews of immersive learning systems highlight the importance of cognitive and learner-centered support as a critical determinant of learning and motivation (Makransky and Petersen, 2021). To address the need for learner-centered support, Palmas et al. (2019) investigated public-speaking training in VR and how skill development can be supported through direct feedback and presence-enhancing mechanisms (Palmas et al., 2021).

As one form of immersive learning environments, virtual classrooms implemented in VR have attracted increasing attention in HCI. Researchers have leveraged VR-specific affordances, such as embodied agents (Liu et al., 2024) and VR videos as virtual excursions (Cheng and Tsai, 2019), to enhance engagement and interactivity. To further enrich the classroom experience, prior work has introduced both teacher-centric tools (e.g., authoring tools (Shen et al., 2025), teaching augmentation (An et al., 2020)) and learner-centered support mechanisms (Liu et al., 2024). Our work aligns with the latter, focusing on learner-centered support in virtual classrooms. While some of these support techniques have analogs in desktop or video-based learning (Szafir and Mutlu, 2013; Thanyadit et al., 2023), VR introduces unique immersive dynamics, such as spatial attention shifts and embodied presence, making dedicated support for immersive learning necessary (Liu et al., 2024).

For immersive virtual classrooms, learner-centered designs have included embodied agents acting as virtual peers during lectures (Liu et al., 2024), as well as studies of classroom dynamics such as avatar representation, teacher-student proximity, and classroom layout (Blume et al., 2019; Gao et al., 2021). However, the majority of this work has emphasized in-lecture support, with relatively little attention to how support can extend beyond the immersive lecture itself. Among the in-lecture support studies, a few have explored adaptive and cognitive support, for example, through gaze-based tools for note-taking (Tsai et al., 2025) and attention-based warnings during lectures (Han et al., 2022), pointing toward promising directions for post-lecture support in immersive learning environments. Targeting the virtual classroom as our design space, our work takes a first step toward transferring established attention-support techniques into immersive learning, while acknowledging the need for investigating learner-centered support in more embodied and immersive learning scenarios in future research.

In summary, immersive virtual classrooms have shown promise, but learner-centered support often stops at in-lecture support, leaving open the question of how learning can be

effectively extended and supported after the session.

5.2.2 Attention Aware Systems and Gaze-adaptive Support

To provide cognitive support in learning contexts, attention-aware systems focus on one key cognitive process of attention (Roda and Thomas, 2006). Attention has been defined as the process of selecting relevant perceived information, allowing individuals to become "active seekers and processors" of knowledge (Chun and Wolfe, 2008). In learning, attention is essential for effective information processing (Mayer, 2014a) and sustaining situational interest (Makransky and Petersen, 2021). In the cognitive-affective model of immersive learning proposed by Makransky and Petersen (2021), attention is identified as a crucial factor influencing learning outcomes.

HCI research has developed attention-aware systems to scaffold attentional processes (Roda and Thomas, 2006). A closely related concept is attentive user interfaces, which focus on making the interface responsive to users' attentional states, rather than treating attention support as the core design goal of the system (Bulling, 2016; Vertegaal et al., 2006). Regardless of the nuanced differences between the two concepts, eye-tracking is the most widely used mechanism for these systems, offering a non-intrusive proxy for user attention (Duchowski, 2007). In existing research, attention-aware systems have included gaze-adaptive intelligent tutoring (D'Mello et al., 2012), note-taking (Khan, 2019), reading & writing support (Buscher et al., 2012; Langner et al., 2023), and adaptive feedback during learning tasks (Jarodzka and Brand-Gruwel, 2017; Liu et al., 2025b). Besides these gaze-adaptive systems, EEG-based cognitive support and attention monitoring have also been explored (Roda and Thomas, 2006; Zander et al., 2010).

Specifically for lecture consumption and in classroom settings, Hutt et al. (2019) used eye-tracking to detect mind wandering during recorded lectures, demonstrating the feasibility of in-lecture attention monitoring. Xu et al. (2025) developed a learning feedback system that aggregates the gaze data of peers to inform the area of interests (AOI) during lectures. Furthermore, Szafir and Mutlu (2013) developed *ARTFul*, a system that leveraged EEG to monitor attention during lectures and provide learners with suggestions for content review. Their results demonstrated that with personalized content review, learners achieve better learning and recall. Their work established that attention signals can effectively guide post-lecture support and identified several future research directions which we address. First, they acknowledged the limitations of passive content review and advocated for exploring “other embedded assessment techniques” based on attention. We respond to this by using personalized quizzes administered both directly after the lecture and before the subsequent lecture, aiming to provide more comprehensive and active support. Second, they envisioned that advances in intelligent content summarization could provide learners with a “truly customized educational experience”; we design our system towards this by leveraging LLMs for quiz generation and post-lecture Q&A chat. Lastly, they noted the limitations of EEG, including that it’s prone to be affected by extraneous signals. Within our target design space of immersive learning support, we employ eye-tracking, which is increasingly integrated into VR headsets and seen as a less obtrusive sensing method (Plopski et al., 2022). Eye-tracking has been established as a reliable and non-intrusive approach for attention-aware support (Chun and Wolfe, 2008; Duchowski, 2007; Roda and Thomas, 2006). More broadly, researchers have begun to examine the potential of using gaze data for personalized content generation when combined with LLMs. For instance, Abdrabou et al. (2025) discussed both opportunities and ethical concerns of gaze-informed LLM systems, though their work was not situated in VR or immersive learning. Therefore, compared to existing attention-based support for non-immersive learning using EEG (Szafir and Mutlu, 2013), the open question we address is how to (a) capture attention more unobtrusively in immersive environments, (b) transfer those signals across devices after the lecture session, and (c) translate attention into personalized quizzes rather than content replays.

In immersive learning environments, however, leveraging gaze data for cognitive support has only recently emerged as a research focus. Recent explorations include the framework by Abeysinghe et al. (2025), who proposed measuring attention with eye-tracking in immersive settings to adapt the presentation of learning materials. While promising, their work remained limited to technical feasibility and did not address learner-centered content adaptation or post-lecture support. Similarly, Han et al. (2022) and Tsai et al. (2025) designed gaze-adaptive support mechanisms within VR sessions, such as adaptive hints and note-taking aids. Furthermore, Liu et al. (2024) integrated LLM-driven embodied agents to support students during VR lectures. However, none of these works have addressed gaze-informed personalization beyond the session itself.

Thus, while gaze-adaptive support in attention-aware systems demonstrates strong potential, prior work largely focuses on in-situ adaptation, leaving the post-lecture support

underexplored, especially in immersive learning contexts.

5.2.3 Personalized Quizzes and Learning Assistants

Personalized support after lectures has long been explored in educational research, with quizzes recognized as one validated strategy for learning reinforcement. Quizzes have been shown to enhance metacognition and self-regulation (Dunlosky et al., 2013). Besides, the testing effect of post-lecture quizzes has been widely observed to improve knowledge retention (McDaniel et al., 2011), self-efficacy (Dunlosky et al., 2013), and engagement (Nuci et al., 2021).

Beyond static, pre-defined questions, adaptive and personalized quizzes have also been supported in digital learning platforms. For instance, Ross et al. (2018) designed adaptive quizzes that adjusted question difficulty based on prior responses, leading to improvements in motivation and engagement. Similarly, Klaveren et al. (2017) showed that adaptive question sequencing better supported diverse learning needs. More recent work by Contrino et al. (2024) integrated personalized quizzes into a smart learning platform that dynamically adjusted course structures. In HCI, adaptive learning and personalization have also been studied in intelligent tutoring systems (D'mello and Graesser, 2012), assessment tools (Gamage et al., 2019), personalized learning analytics (Demmans Epp et al., 2023), etc. However, personalization in these works typically relies on performance data, with fewer systems incorporating cognitive states such as attention. Moreover, little research has adapted these strategies specifically to virtual classrooms.

In parallel, learning assistants as mobile applications have been widely studied as flexible tools for delivering post-lecture support (Hwang and Tsai, 2011). They are capable of delivering a ubiquitous learning experience independent of human tutors after the lectures (Hwang and Tsai, 2011). Studies have integrated quizzes into mobile assistants (Nuci et al., 2021). With AI-driven personalization, recent work has explored LLM-based assistants with adaptive learning support (Wambsganss et al., 2021). For immersive learning, cross-device interaction research in HCI already underscores the opportunity of extending immersive learning through mobile devices (Brudy et al., 2019; Zhu et al., 2024). Yet, immersive learning research has not sufficiently investigated mobile assistants that extend VR lectures with ex-situ personalized support.

In short, while personalized quizzes and mobile learning assistants have been researched,

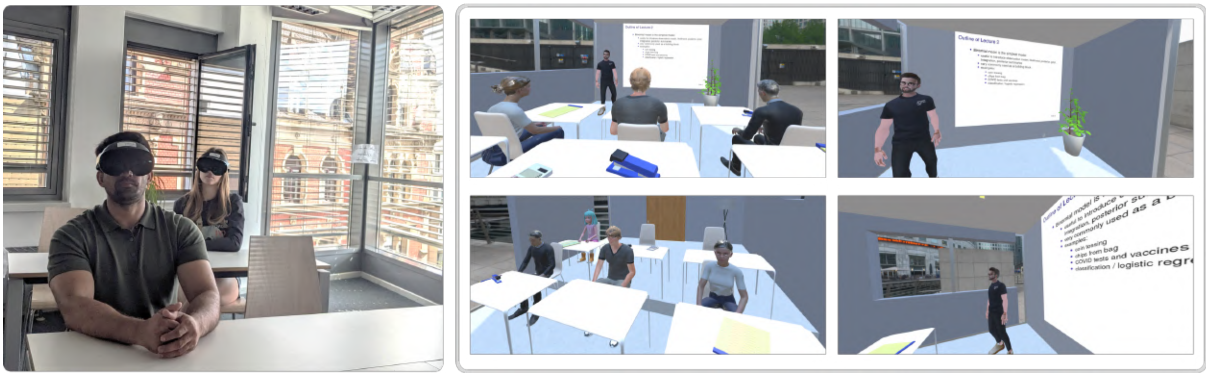

Figure 5.1: An immersive virtual classroom setup. Left: students using VR headsets in a real-world environment. Right: corresponding classroom scene showing multiple perspectives of the lecture space with avatars and slides.

their integration into immersive learning with attention-aware personalization has yet to be explored.

5.3 AttentiveLearn: Designing an Attention-Aware Learning Ecosystem

After identifying the research gap, we clarified our design goal and scope of *AttentiveLearn*: we aim to provide personalized post-lecture support that adapts to individual attention levels during immersive learning sessions. Rather than innovating in VR classroom design itself, our design serves as a reference implementation of a generalizable framework, focusing on bridging in-situ immersive learning with ex-situ personalized support through a mobile assistant. Specifically, we investigate how attention-aware personalization influences students' motivation, engagement, and learning outcomes.

5.3.1 An Immersive Virtual Classroom

In our framework, the virtual classroom serves as the *In-Situ Data Acquisition* layer. Immersive virtual classrooms represent an established problem space rather than a novel contribution of this work. We chose this setting for two reasons: (1) gaze-based attention metrics have been validated in classroom contexts (Han et al., 2022; Hutt et al., 2019), and (2) virtual classrooms remain a widely adopted format in both HCI research and practice (Gao et al., 2021; Moynihan, 2021), with open-source infrastructures further provide a practical foundation for the design space (Oehlberg, 2018).

For our implementation, we adopted the existing Unity-based virtual classroom from Liu et al. (2025a) with their consent. This system provides a standard classroom setup consisting of a slideshow and multiple embodied avatars. In their classroom setup, one avatar represents the lecturer, delivering a pre-recorded lecture with synchronized audio, lip movements, and slide transitions. Additional avatars simulate peer students, creating the impression of a live lecture. Furthermore, the system already integrates a gaze data collection pipeline and a basic gaze-duration–based attention metric (AOI coverage), which has been technically evaluated and aligns with established attention metrics for in-lecture interventions (Han et al., 2022) and teacher-oriented attention monitoring (Thanyadit et al., 2023). This provides a solid foundation for our work. Our modifications to this environment were minimal. We extended the system's attention tracking with the ability to define AOIs on each slide. This enables tracking of attention switches between AOIs besides the gaze duration. Based on the gaze duration and the number of attention switches during each lecture section (the sections can be predefined in the Unity application via timestamps), an Attention Distribution Index (ADI) can be later calculated for each section in the attention-aware personalization pipeline following Sharma et al. (2020).

Thus, the immersive virtual classroom functions as a validated problem space that allows us to investigate our main contribution: designing a framework for attention-aware personalization for post-lecture support.

5.3.2 Attention-Aware Personalization Pipeline

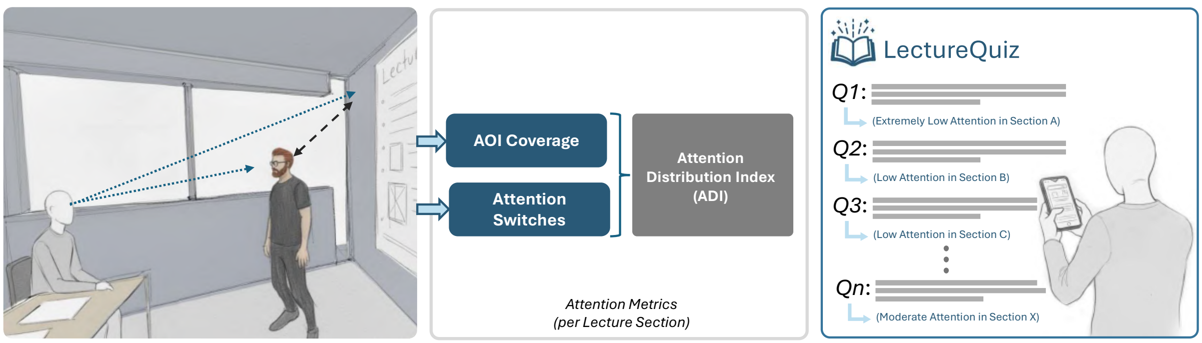

Figure 5.2: Attention-aware personalization pipeline. Eye-tracking data from VR lectures are processed into attention metrics, which are then used to generate personalized *LectureQuiz* questions targeting low-attention sections.

Data Processing We implemented a Python-based web server using Flask-SocketIO[1] that connects the VR classroom with a mobile assistant application. This pipeline instantiates the framework's *Data Translation* layer, converting raw signals into pedagogical insights. It executes three steps: (1) receiving raw eye-tracking data with timestamped gaze positions as CSV files from the VR application; (2) processing these into section-level attention metrics; and (3) forwarding metrics in a JSON file to the mobile assistant. The metrics included (a) AOI coverage percentage, (b) number of attention switches, and (c) the ADI metric. AOI coverage was calculated as in Liu et al. (2025a): the accumulated gaze duration on predefined learning-related objects (lecturer avatar and slides) divided by the total lecture length.

Attention switches were approximated from gaze shifts (Sharma et al., 2020). The ADI metric combines both measures to represent overall attention across lecture sections, following the methodology of Sharma et al. (2020) and Hutt et al. (2019). We intentionally relied on these established measures rather than proposing new ones, to ensure that the pipeline remains adaptable and modular, allowing the framework to accommodate alternative biosignals or attention definitions in future iterations.

LectureQuiz These metrics inform an attention-aware quiz module named *LectureQuiz*, which provides immediate feedback and assessment after each lecture, following the principle of test-enhanced learning (McDaniel et al., 2011). Personalization works as follows: *LectureQuiz* provides learners with LLM-generated questions focusing on sections with low AOI coverage and ADI (see Supplementary Materials for prompts). The number and difficulty of questions can be configured in the prompts based on the learning objectives. In our study, we investigate the integration of *LectureQuiz* in a mobile application, while also validating the perceived quality of *LectureQuiz* (see Section 5.5.2). This integration ensures that attention-aware personalization is not an isolated design feature but part of a comprehensive learning ecosystem.

5.3.3 Integrating Personalized Quizzes in a Mobile Assistant

LectureQuiz Integration To situate attention-aware personalization in students' ex-situ learning after lectures, we integrated *LectureQuiz* into an existing mobile assistant application deployed for self-regulated learning at our university, serving as the framework's *Ex-Situ Inter-*

[1] `https://flask-socketio.readthedocs.io`

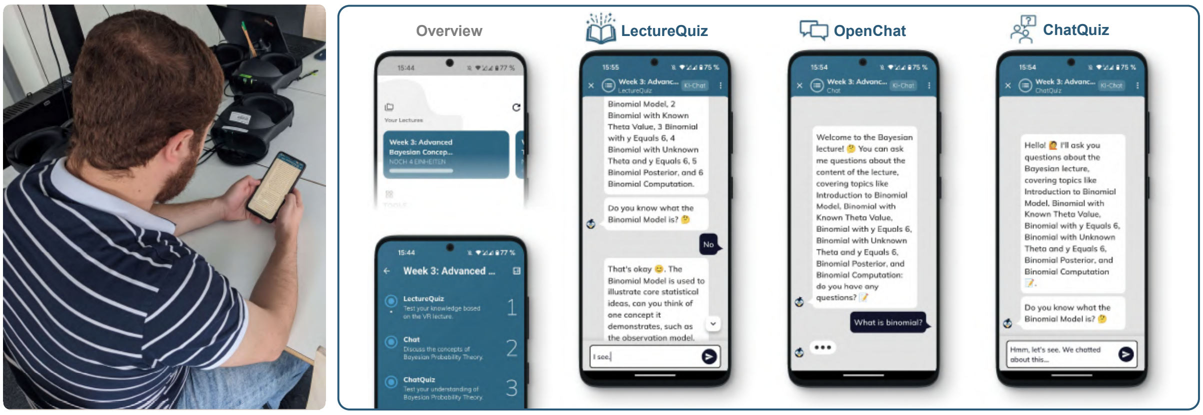

Figure 5.3: Integrating personalized support in the mobile assistant. Left: a student doing *LectureQuiz* after a VR lecture. Right: screenshots of the assistant, including personalized *LectureQuiz*, *OpenChat* for Q&A support, and *ChatQuiz* for additional practice.

vention interface. The assistant has been implemented as a native Android and iOS application using Dart and the Flutter framework[2]. This application provides a familiar platform for students, enabling us to examine how personalized support interacts with ongoing learning practices. As described in our attention-aware personalization pipeline, the mobile application receives the JSON file with attention metrics after VR lectures. These metrics are incorporated into the prompt provided to the LLM, which then generates *LectureQuiz* as a learning module within the application (see Figure 5.3).

OpenChat Beyond *LectureQuiz*, the assistant integrates other supporting features, including *OpenChat*, a conversational Q&A module allowing users to pose follow-up questions and receive answers after the lecture. For *OpenChat*, the mobile assistant leverages a knowledge base constructed from lecture slides, textbooks, and recorded lecture transcripts. When a student submits a question, a similarity search is performed on a PostgreSQL database using embeddings. The most relevant content is then appended to the LLM prompt as grounding information, following the retrieval-augmented generation approach.

ChatQuiz Furthermore, chat logs in *OpenChat* inform users' confusion levels per lecture section, following an established method of linguistic confusion detection (Atapattu et al., 2019). Later, the confusion levels can be used to generate an additional quiz module, *ChatQuiz*. While *ChatQuiz* explores further personalization possibilities, the main focus of our study remains on the attention-driven *LectureQuiz* and the effect of our learning ecosystem *AttentiveLearn* as a

[2]`https://flutter.dev/`

whole.

In our design, *AttentiveLearn* bridges immersive in-lecture experiences with ubiquitous post-lecture support. By embedding attention-aware personalization into a mobile assistant, we explore beyond using attention data as one-off feedback and provide a continuous, learner-centered ecosystem for real-world learning. To ensure transparency and reproducibility, we made the system components of *AttentiveLearn* open source and available as supplementary material.

5.4 Field Study

We evaluated *AttentiveLearn* in a between-subjects field study with 36 participants (12 female, 24 male). The study aimed to investigate the impact of attention-aware post-lecture support on students' motivation, engagement, and learning outcomes in an authentic setting. The study protocol was approved by the university IRB. Due to IRB constraints, the study could not be embedded in the official curriculum or offer credit points toward students' degrees, as this might have biased academic achievement. Instead, we recruited participants from a local university and organized a three-week lecture series on Bayesian data analysis and a final exam on the fourth week, offered as an optional non-credit course on the university's learning platform. The lectures were based on a real-world course by Vehtari (2024), which covers basic concepts of Bayesian data analysis including probability theory, single-parameter and hierarchical models, Bayesian inference, etc. and has been used as an open educational resource in universities.

5.4.1 Study Design

We employed a two-factor mixed design with one between-subjects factor *Group* (attentive vs. non-attentive) and one within-subjects factor *Week* (1–3). Participants were randomly assigned to one of the two groups. As the independent variable, the two groups differed only in the type of post-lecture support delivered through the mobile assistant: the attentive group received an attention-aware *LectureQuiz*, while the non-attentive group received a non-personalized version. For comparable expectations across two conditions, all participants were informed that the study involved "personalized learning support" and that eye-tracking data would be collected throughout the study. This provided a transparent and data privacy-compliant study design. However, the attention-based personalization mechanism for the attentive group was not revealed until the final interview. This prevented participants in the attentive group from knowing during the study that their attention data directly informed quiz generation. We examined three sets of dependent variables:

- **Engagement:** measured weekly with the short-form User Engagement Scale (UES-SF) (O'Brien et al., 2018).
- **Motivation:** measured weekly with the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, 1991).
- **Learning Outcomes:** assessed with three weekly *Mini-Exams* in the application and an on-site exam in week 4.

Additional behavioral data (eye-tracking in VR lectures, log data from *LectureQuiz*, etc.) were collected to support and validate findings. Finally, semi-structured interviews with 10 participants provided qualitative insights.

Incentivization Strategy Because the course did not contribute credit points, we implemented an incentive scheme to approximate a typical student workload. Participants earned €15 for completing each weekly on-site lecture and the subsequent *LectureQuiz*, plus €8 for completing the out-of-class activities (*ChatQuiz* and *Mini-Exam*) with the mobile assistant each week. Completion of the final exam yielded an additional €15. These rates were aligned with the standard hourly payment of student assistants at our university. In addition, participants could earn €2 bonuses for ranking in the top 25% of each *Mini-Exam* and the final exam. The incentive structure was intended to provide fair compensation across the four weeks, while avoiding undue pressure that might bias learning behavior.

5.4.2 Participant Information

We recruited 36 students (12 female, 24 male) from two local universities, aged 18-28 ($M = 23.61$, $SD = 3.26$). Seventeen were enrolled in a bachelor's program, 14 in a master's program, and five had recently completed their master's degree. Pre-study surveys confirmed low prior knowledge of Bayesian data analysis (15 reported none, 21 little; $M = 1.42$, $SD = 0.50$, on a

5-point scale). Familiarity with VR varied: five had never used VR, 29 reported rare use (less than twice per year), and two reported occasional use (a few times per month).

Group Comparability Participants were randomly assigned to the two groups (attentive vs. non-attentive). Balance checks on the initial sample ($n = 36$) showed no significant group differences in age ($t(34) = 0.710$, $p = 0.482$), prior knowledge ($U = 153.00$, $p = 0.753$), VR experience ($U = 153.50$, $p = 0.713$), or educational degree ($\chi^2(2) = 0.545$, $p = 0.762$). Assumptions of normality and variance homogeneity were confirmed where applicable.

Five participants dropped out after week 2 and three after week 3, leaving 28 participants (10 female, 18 male) who completed the full study. This drop-out rate was comparable to that of real lectures at our university. After dropout, group sizes remained comparable (13 attentive vs. 15 non-attentive), and the distribution of demographics and prior knowledge remained balanced.

5.4.3 Task and Procedure

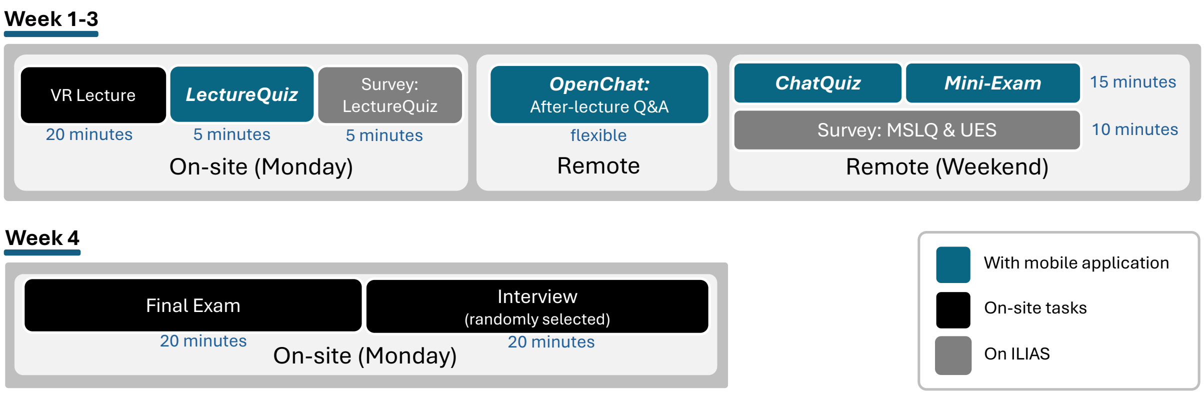

Figure 5.4: Study procedure over four weeks. Weeks 1–3 included a VR lecture, *LectureQuiz* and survey. Followed by out-of-class *OpenChat* for Q&A, *ChatQuiz*, weekly surveys, and a mini-exam. Week 4 concluded with a final exam and semi-structured interviews.

Participants took part in a four-week course that combined an immersive virtual classroom with post-lecture support via the mobile assistant.

Weeks 1–3 Three on-site VR lectures, each around 20 minutes long, were adapted from the original Bayesian data analysis course and offered each week (Vehtari, 2024). Following the structure of the source material, each lecture was divided into six sections. The section-division was not visible to participants and had no impact on their in-lecture experience, but served as analytic units for calculating section-level attention metrics in the *AttentiveLearn* backend.

Each Monday, participants attended an on-site lecture using a Meta Quest Pro headset, organized in groups of three to five to mirror the virtual classroom setting. This schedule resulted in a total of 10 on-site sessions every Monday. After the lecture, they completed a *LectureQuiz* in the mobile application with six questions (personalized for the attentive group based on in-lecture attention, random for the non-attentive group). Although the quizzes received by the non-attentive group were not personalized, these items were also pedagogically validated and reflect current practices in lectures: all quiz questions were generated based on pedagogically validated material including the original assessment of the open course (Vehtari, 2024), as well as the exercise questions in the course textbook (Gelman et al., 1995), and were reviewed retrospectively by a statistics expert and lecturer for content alignment and appropriate difficulty. They continued with a two-item survey on their perceived accuracy and helpfulness of *LectureQuiz*, measured on a 5-point Likert scale:

Q1. How accurately did *LectureQuiz* reflect the parts of the lecture where you paid less attention?

Q2. To what extent did *LectureQuiz* help you understand and review the lecture content?

After the weekly lecture on site, participants went home and carried out all subsequent activities independently in their own environments. During the week, they could freely use *OpenChat* at any time. At the end of the week, starting each Friday, a *ChatQuiz* and a *Mini-Exam* module (with 12 predefined questions validated by a statistics expert) were made available on the mobile assistant. Participants could complete these tasks flexibly over the weekend before the next Monday lecture. Weekly surveys on motivation (MSLQ) and engagement (UES-SF) were also administered on the online learning platform of the local university[3]. This weekly cycle was repeated for three weeks (see Figure 5.4).

Week 4 On Monday of week 4, participants completed a final on-site exam consisting of 12 multiple-choice questions. The items were drawn from a mix of original course assessments (Vehtari, 2024), textbook exercises (Gelman et al., 1995), and LLM-generated questions,

[3]with the built-in survey tool on the ILIAS platform

all of which were reviewed by a statistics expert, and none of these questions overlapped with the Mini-Exams. For the final exam, participants had 20 minutes and were not permitted to use any supporting materials. After the exam, and with consent, one participant per session was randomly selected for a semi-structured interview.

5.4.4 Data Collection and Analysis

Data Collection We collected multimodal data for mixed-methods analysis: (a) eye-tracking during lectures (AOI coverage percentage, attention switches, ADI), (b) log data from the mobile assistant including *LectureQuiz* scores, (c) on-site survey (two items on *LectureQuiz*) and weekly surveys (MSLQ, UES-SF), (d) weekly *Mini-Exams* and a final exam, and (e) post-study interviews.

For the weekly surveys, we followed established guidelines. We included the motivation scale of MSLQ with six subscales (intrinsic goal orientation, extrinsic goal orientation, task value, control beliefs, self-efficacy, test anxiety) on a 7-point Likert scale (Pintrich, 1991). The UES-SF measured four engagement dimensions (focused attention, perceived usability, aesthetic appeal, reward factor) on a 5-point Likert scale (O'Brien et al., 2018). We also analyzed *OpenChat* logs and *ChatQuiz* scores from the mobile assistant to capture patterns of self-directed study.

Finally, we collected qualitative data through 10 semi-structured interviews conducted after the final exam, with five participants randomly selected from each group. The interview protocol covered four areas: (a) study routines and overall impressions, (b) perceptions of post-lecture support, (c) motivation and engagement, and (d) suggestions for system improvement. The full interview guide and coding framework are provided as supplementary materials.

Analysis Methods Quantitative analyses combined frequentist and Bayesian approaches. Two-way mixed ANOVAs were applied for weekly surveys and exams, with *Group* (attentive vs. non-attentive) as a between-subjects factor and *Week* (1–3) as a within-subjects factor. Assumptions of normality, homogeneity of variances, and absence of outliers were confirmed using visual inspection of Q-Q plots and Levene's tests. Furthermore, because the MSLQ and UES-SF surveys contain multiple subscales, we accounted for the risk of family-wise error by applying a Holm-Bonferroni correction strategy across the scales. Final exam scores were analyzed using independent-samples t-tests, with assumptions of normality confirmed with a Shapiro-Wilk test ($W = 0.956, p = 0.287$) and homogeneity of variance confirmed using a Levene's test ($F(1, 26) = 1.344, p = 0.257$). Additionally, to assess the construct validity of the ADI, we performed a repeated measures correlation analysis between the per-section ADI and the subjective quiz ratings on accuracy and helpfulness.

Qualitative data from the interviews were analyzed using reflexive thematic analysis (TA). Following the six-phase approach (Braun and Clarke, 2012), the first author engaged in the process of familiarization with the data, initial coding, theme construction, reviewing, refining, and writing. Coding was conducted openly and inductively to remain grounded in participants' accounts through an interpretative and iterative process, with themes generated through analytical engagement with the data. As pointed out by Braun and Clarke (2019), reflexive TA does not depend on multiple coders. Therefore, the analysis was conducted by the first author. To ensure rigor, the first author revisited codes and themes multiple times to refine their scope after distancing and engagement with the dataset. A reflexive log was kept during the process to document analytic decisions and made available as supplementary material.

5.5 Results

We present our findings in six parts, combining quantitative and qualitative insights. First, we report on in-lecture attention patterns measured through eye-tracking in VR. Then, we analyze user feedback on the attention-aware *LectureQuiz*. Next, we examine user engagement and learning motivation. Furthermore, we present learning outcomes based on weekly mini-exams and the final exam. Finally, we describe how participants appropriated the support features, including *OpenChat* and *ChatQuiz*, and summarize additional suggestions from the interviews.

5.5.1 In-Lecture Attention

Overall Attention Distribution As shown in Figure 5.5, we visualize the eye-tracking data collected during the three-week VR lectures. The AOI coverage percentage (see Figure 5.5a) was calculated per minute by dividing the total gaze duration on the learning-related AOIs (lecturer avatar and lecture slides) by the 60-second duration. Our results indicate that both groups maintained similar attention levels throughout the three-week lectures.

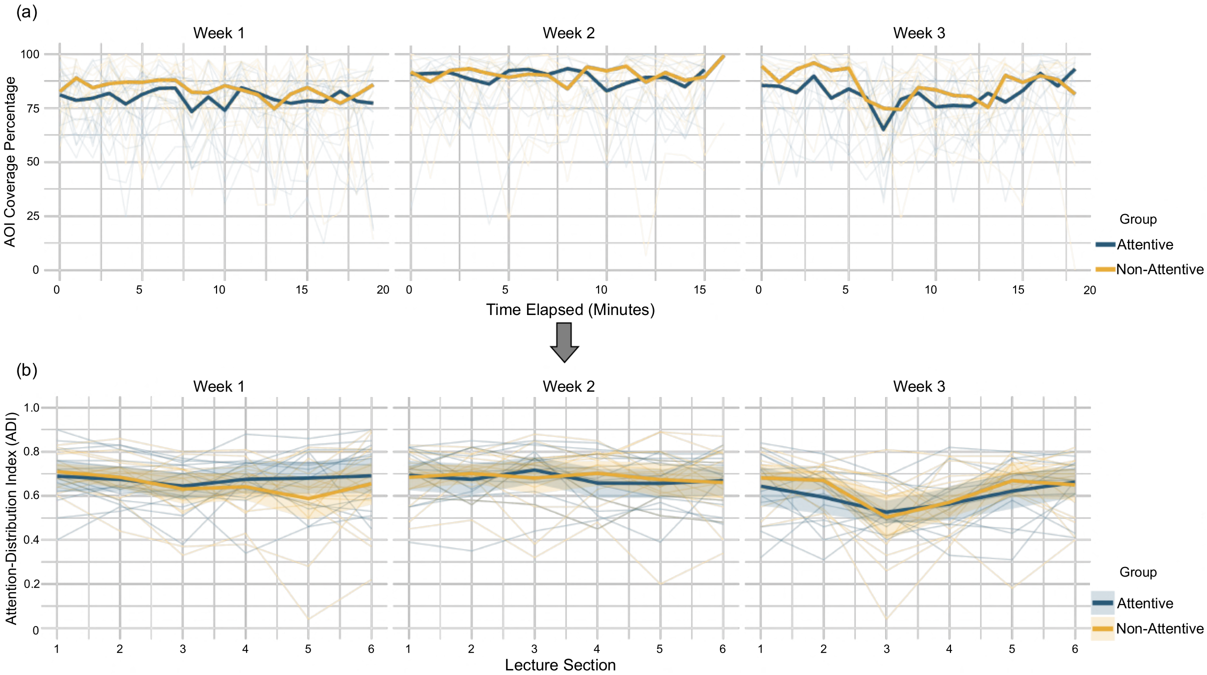

Figure 5.5: (a) In-lecture focused percentage over time (minutes) across three weeks for attentive and non-attentive groups; (b) Attention Distribution Index (ADI) across lecture sections for both groups.

We conducted a mixed ANOVA on the ADI metric (ranging 0–1; see Figure 5.5b), calculated at the lecture-section level based on AOI coverage and attention switches. The analysis revealed a significant within-subject effect of *Week*, $F(2, 52) = 5.015, p = .010, \omega^2 = 0.069$, indicating that ADI changed over time. This finding was supported by the Bayesian mixed ANOVA, which showed that the data were best explained by a model including only *Week* ($BF_{10} = 9.69$), with inclusion Bayes factors providing moderate evidence for *Week* ($BF_{\text{incl}} = 6.20$). Post-hoc pairwise comparisons (Holm-adjusted) localized the effect to a difference between week 2 and week 3 ($MD = 0.074, SE = 0.023; t(26) = 3.189, p_{\text{holm}} = .011; d = 0.728$), with the Bayesian comparison providing strong converging evidence ($BF_{10} = 13.08$). In contrast, both analyses provided little evidence for a main effect of *Group* ($F(1, 26) = 0.177, p = .677; BF_{\text{incl}} = 0.185$) or for a *Week* × *Group* interaction ($F(2, 52) = 0.523, p = .596; BF_{\text{incl}} = 0.160$).

Participants with Low-Attention Recognizing variation in attention levels among participants, we conducted a post-hoc and exploratory descriptive analysis focusing on those with low attention. We defined this subset as participants whose week-1 ADI was at or below the 25th percentile (0.619), including eight participants evenly split across the two groups. For this subset of participants, those in the attentive group showed a greater average ADI increase ($M = +0.086, SD = 0.092$) than their non-attentive counterparts ($M = +0.026, SD = 0.187$) from week 1 to week 3. An independent-samples t-test showed that this difference was not statistically significant, $t(6) = 0.576$, $p = 0.585$. A complementary Bayesian t-test yielded $BF_{10} = 0.578$, providing anecdotal evidence for the null hypothesis and suggesting that a larger sample size would be needed to draw stronger conclusions.

Qualitative Results Thematic analysis of interview data further contextualized these findings. From 10 interviews, we identified six codes that informed **Theme 1: Promoting Attention Management in Subsequent Lectures**.

Participants described immersive VR lectures as engaging, regardless of prior VR familiarity (the interviewed participants included novice users of VR: P1, P3, P15, P24, and more frequent users: P4, P7), echoing prior work on the benefits of immersive virtual classroom (Gao et al., 2021). However, participants also highlighted challenges of attention management. Several reported initial disorientation, noting they "didn't know what to look for in this new type of classroom" (P24) or felt "lost in another world" (P3), pointing to the need for learner-centered attention support. In addition, six participants mentioned that attention was harder to sustain in week 3 as "the lecture became more difficult" (P1)—a pattern consistent with both course design (Vehtari, 2024), as well as the observed ADI decrease (Figure 5.5b). This pattern validates existing findings that attention management remains a challenge in immersive virtual classrooms (Han et al., 2022).

While most participants agreed that *AttentiveLearn* did not alter the in-lecture experience directly (since personalization occurred post-lecture), three participants in the attentive group (P7, P15, P19) reported consciously trying to sustain greater focus in later weeks, motivated by anticipating the quizzes, even though the attention-based personalization mechanism was not explicitly apparent to them. As P19 explained: "because I know there is a quiz afterwards and I don't want to perform bad there, I try to focus more listening (to the lecture) to make full use of the quiz". P7 described a related effect, saying: "I'm aware that the quiz will help me afterwards, so I became more confident and comfortable during the lectures". None of the non-attentive participants reported similar impressions. Importantly, these three participants from the attentive group were also among those classified as low-attentive in week 1, and their ADI scores indeed improved over time ($M = +0.13$, $SD = 0.03$).

These results first validate existing research that attention management is a challenge in immersive learning environments. At the same time, our findings suggest that *AttentiveLearn* may have motivated low-attentive students to develop more effective attention management strategies in subsequent lectures.

5.5.2 Attention-Aware *LectureQuiz*

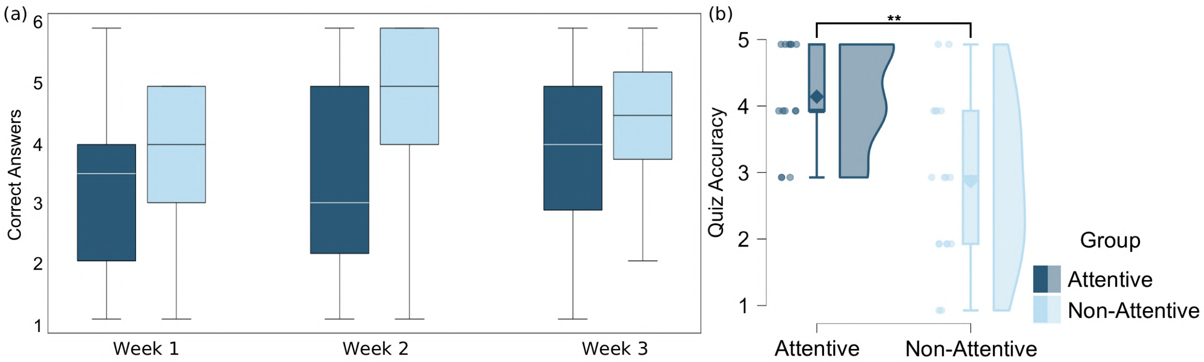

Figure 5.6: (a) Correct answers in *LectureQuiz* across three weeks by group. (b) Aggregated *LectureQuiz* accuracy ratings on a 5-point Likert scale; asterisks denote significant differences (** $p < 0.01$).

Quiz Scores As shown in Figure 5.6, participants completed the *LectureQuiz* with six questions immediately after each lecture, with responses evaluated in real time by an LLM to calculate quiz scores. Overall, the attentive group scored slightly lower across weeks ($M = 2.67$, $SD = 1.721$) compared to the non-attentive group ($M = 3.21$, $SD = 1.413$), but both frequentist and Bayesian analyses suggest that these differences were not statistically meaningful.

For the between-subject factor *Group*, the mixed ANOVA indicated no significant difference, $F(1, 26) = 2.64, p = 0.11$, and the Bayesian analysis likewise provided little evidence for including *Group* ($BF_{\text{incl}} = 0.396$). Similarly, for the within-subject factor *Week*, no effect was observed, $F(2, 52) = 1.59, p = 0.21$, with the Bayesian analysis again indicating little support ($BF_{\text{incl}} = 0.234$). Finally, the *Group* $\times$ *Week* interaction was not significant, $F(2, 52) = 0.55, p = 0.58$, and the Bayesian inclusion factor strongly favored exclusion ($BF_{\text{incl}} = 0.077$).

Survey on *LectureQuiz* To assess participants' perceptions, responses from the two-item survey on quiz accuracy and helpfulness were aggregated across the three weeks. Independent-samples t-tests were then conducted.

For **quiz accuracy**, participants in the attentive group ($M = 4.21, SD = 0.80$) rated their quizzes as more accurate in revealing attention gaps than the non-attentive group ($M = 2.94, SD = 1.25$). An independent-samples t-test confirmed that this difference was statistically significant, $t(29) = 3.292, p = 0.003, Hedges'g = 1.157$. Consistently, the Bayesian t-test supported the group difference, with $BF_{10} = 14.28$. Under the directional hypothesis (attentive > non-attentive), the one-sided test also supported higher perceived accuracy in the attentive group, $BF_{+0} = 28.45$ (posterior median $d = 1.003$). Aiming to validate the objective attention metric and subjective perception, we analyzed the correlation between ADI and perceived quiz accuracy per section. For the attentive group, a repeated measures correlation indicated a significant negative association ($r_{rm} = -0.72, p < 0.001$), confirming that lecture sections with lower detected attention were significantly more likely to be rated as having accurate quizzes. In contrast, no significant correlation was found for the non-attentive group ($r_{rm} = 0.02, p = 0.714$).

For **quiz helpfulness**, a Shapiro–Wilk test indicated non-normality ($p < 0.001$), so we used a Mann–Whitney U test, which revealed no significant difference between groups ($U = 124.000, p = 0.834$). A Bayesian Mann–Whitney U test converged, indicating that the data are about 2.8 times more likely under the null hypothesis ($BF_{10} = 0.357$). Regarding helpfulness, the repeated measures analysis revealed no significant correlation with ADI for either the attentive group or the non-attentive group.

Qualitative Results Thematic analysis provided further insights into participants' experiences with *LectureQuiz*. We generated **Theme 2: Effective Attention-Aware Personalization of *LectureQuiz*** from 12 codes related to quiz accuracy, timing, and learner perceptions.

Participants across both groups valued quizzes as a form of post-lecture support. For example, non-attentive group participants noted that quizzes had helped them "refresh knowledge in memory" (P21). However, personalization was distinctly emphasized by attentive group participants, who described the questions as covering content that was "new and unfamiliar" (P7). While this sometimes led to lower quiz scores, participants interpreted it positively, recognizing that these questions reflected missed content during lectures and thus helped surface attention gaps. This aligns with the higher perceived accuracy ratings in the attentive group.

Participants further indicated a preference for attention-aware quizzes compared to potential in-lecture interventions. Several participants stated that in-lecture attention support, such as warnings (Han et al., 2022), would have been "disruptive to an already complicated lecture"

(P15) or introduced an undesirable "pressure of being monitored" (P19). In contrast, post-lecture quizzes were described as "just in time" (P1), making attention gaps actionable without interrupting the lecture. Finally, many participants (four out of five in the attentive group; two out of five in the non-attentive group) reported that *LectureQuiz* created a positive carry-over effect with pressure, motivating them to "study harder through the rest of the week" (P15).

In conclusion, attention-aware quizzes were perceived as accurate, timely, and non-disruptive, enabling participants in the attentive group to identify attention gaps and sustain learning beyond the lecture.

5.5.3 User Engagement

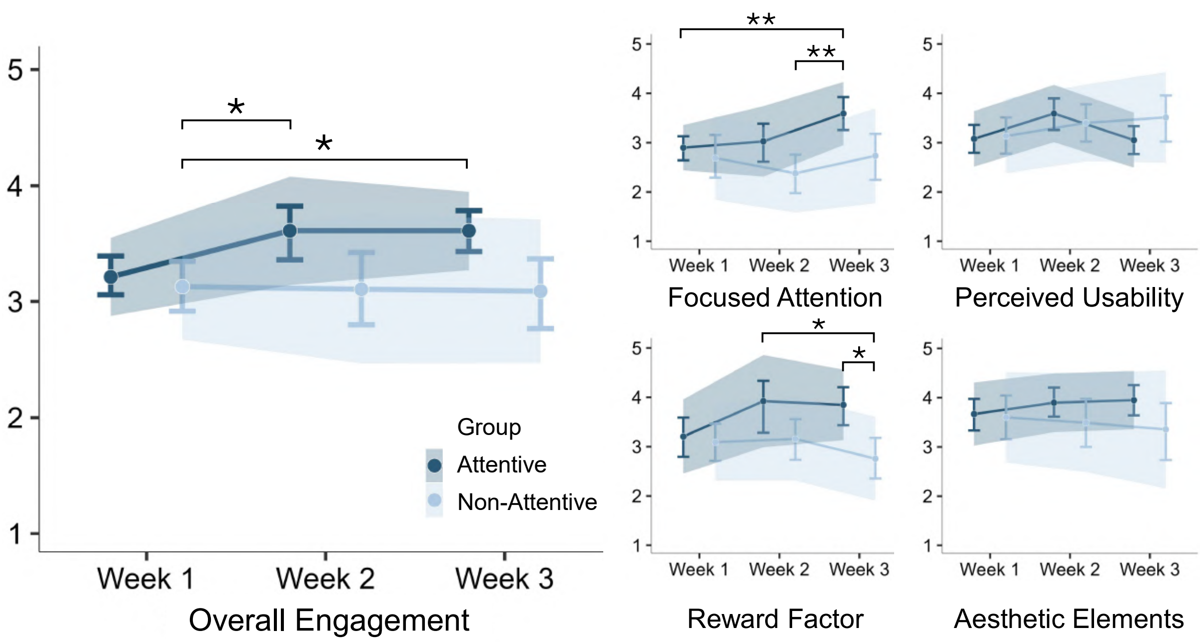

Figure 5.7: Engagement measured with the UES-SF scale. The plots show mean ratings for attentive and non-attentive groups across three weeks on four subscales. Shaded ribbons represent standard deviations, error bars denote 95% confidence intervals, and brackets indicate significant group differences (* $p < 0.05$, ** $p < 0.01$).

Overall Engagement To assess overall engagement, we averaged the results across the four UES-SF subscales, following O'Brien et al. (2018). Descriptive analysis suggested an increase in engagement from week 1 ($M = 3.21, SD = 0.34$) to week 2 ($M = 3.61, SD = 0.47$), which was sustained in week 3 ($M = 3.60, SD = 0.32$) for the attentive group. By contrast, the non-attentive group showed a slight, non-significant decrease over the same period, from week 1 ($M = 3.13, SD = 0.46$) to week 3 ($M = 3.08, SD = 0.62$).

Before running the mixed ANOVA, Mauchly's test indicated that the assumption of sphericity was violated, $\chi^2(2) = 11.454, p = 0.003$, so a Greenhouse–Geisser correction was applied. The analysis revealed a significant between-subject effect of *Group*, $F(1, 26) = 6.319, p = 0.018, \omega^2 = 0.09$, with participants in the attentive group showing higher overall engagement ($MD = 0.369, SD = 0.147$). This result was supported by the Bayesian ANOVA, which provided moderate evidence for including *Group* ($BF_{\text{incl}} = 3.064$) and strong evidence for higher engagement in the attentive group ($BF_{10} = 25.94$). In contrast, neither analysis indicated meaningful effects of *Week* ($F(1.462, 38.042) = 2.138, p = 0.144$; $BF_{\text{incl}} = 0.543$) or of the *Week* $\times$ *Group* interaction ($F(1.462, 38.042) = 2.909, p = 0.082$; $BF_{\text{incl}} = 1.061$).

Reward Factor and Focused Attention At the subscale level, analyses indicated varying degrees of evidence for between-subject effects on **focused attention** and **reward factor**. For focused attention, the attentive group scored higher overall, $F(1, 26) = 5.768, p = 0.024, \omega^2 = 0.081, p_{\text{Holm}} = 0.072$, with Bayesian model comparison providing converging evidence ($BF_{\text{incl}} = 3.479$) and a direct comparison showing strong evidence for the attentive group ($BF_{10} = 24.65$). In addition, the within-subject effect of *Week* was significant, $F(2, 52) = 5.817, p = 0.005, \omega^2 = 0.055$, supported by Bayesian model comparison ($BF_{\text{incl}} = 6.017$). Post-hoc tests showed that scores were higher in week 3 compared to week 1 ($t(26) = 3.090, p = 0.009, d = 0.483$; $BF_{10} = 3.361$) and compared to week 2 ($t(26) = 3.365, p = 0.007, d = 0.602$; $BF_{10} = 15.53$). For reward factor, the attentive group also scored higher, $F(1, 26) = 8.482, p = 0.007, \omega^2 = 0.122, p_{\text{Holm}} = 0.028$, with Bayesian model comparison providing moderate support for including *Group* ($BF_{\text{incl}} = 6.622$) and a direct comparison showing strong evidence for the attentive group ($BF_{10} = 50.02$).

The other two subscales showed comparable results on perceived usability (attentive: $M = 3.329, SD = 0.607$, non-attentive: $M = 3.348, SD = 0.817$) and aesthetic elements (attentive: $M = 3.838, SD = 0.606$, non-attentive: $M = 3.482, SD = 1.026$), with no significant differences between the attentive and non-attentive groups.

Qualitative Results Our thematic analysis produced **Theme 3: Engagement Through Self-Reflection and Sustained Focus Across Weeks**.

The codes grouped under this theme highlighted two mechanisms. First, personalized post-lecture support encouraged participants to actively reflect on their attention and learning gaps, fostering a sense of progress and reward. Second, repeated engagement with the assistant helped participants concentrate and sustain focused attention over time.

For instance, P15 described that realizing gaps after the *LectureQuiz* made problems "visible" and encouraged him to work on them during the week. Similarly, P4 noted how daily interactions with the assistant helped him prepare for the *Mini-Exam*, creating a sense of closure at the end of each week:

> After the lecture, I went home and thought about where I didn't understand. Then I would chat with it (*OpenChat*) and solve some questions day by day. Over the weekend, when I did the *Mini-Exam*, I felt very confident and it was satisfying to wrap up the week.

Beyond this sense of fulfillment, participants also reported that the assistant helped them sustain focus during post-lecture learning and review. Participants in the attentive group not only spent more time using the assistant (see Section 5.5.6) but also reported that they "know what to ask" (P1) and felt more focused during its use. These reflections show that participants not only perceived a reward factor but also felt more attentive with the learning assistant.

Four participants in the attentive group emphasized that this pattern became evident in week 3, when content difficulty and external exam pressure increased. For example, P19 explained that the challenges made him more "curious to test the capability of the learning assistant", which helped him sustain focused attention and engagement despite pressure. By contrast, participants in the non-attentive group reported lacking this reinforcement, describing week 3 as a point where they "just felt like doing the bare minimum because it (the task) was intimidating" (P21).

Therefore, post-lecture support delivered in *AttentiveLearn* fostered reflection, reward, and sustained focus, enabling participants to remain engaged even under increased pressure.

5.5.4 Learning Motivation

To assess changes in learning motivation, a mixed ANOVA was conducted across the six MSLQ dimensions. Following Pintrich (1991), we examined each dimension separately rather than calculating an overall score. Three dimensions showed significant between-group differences:

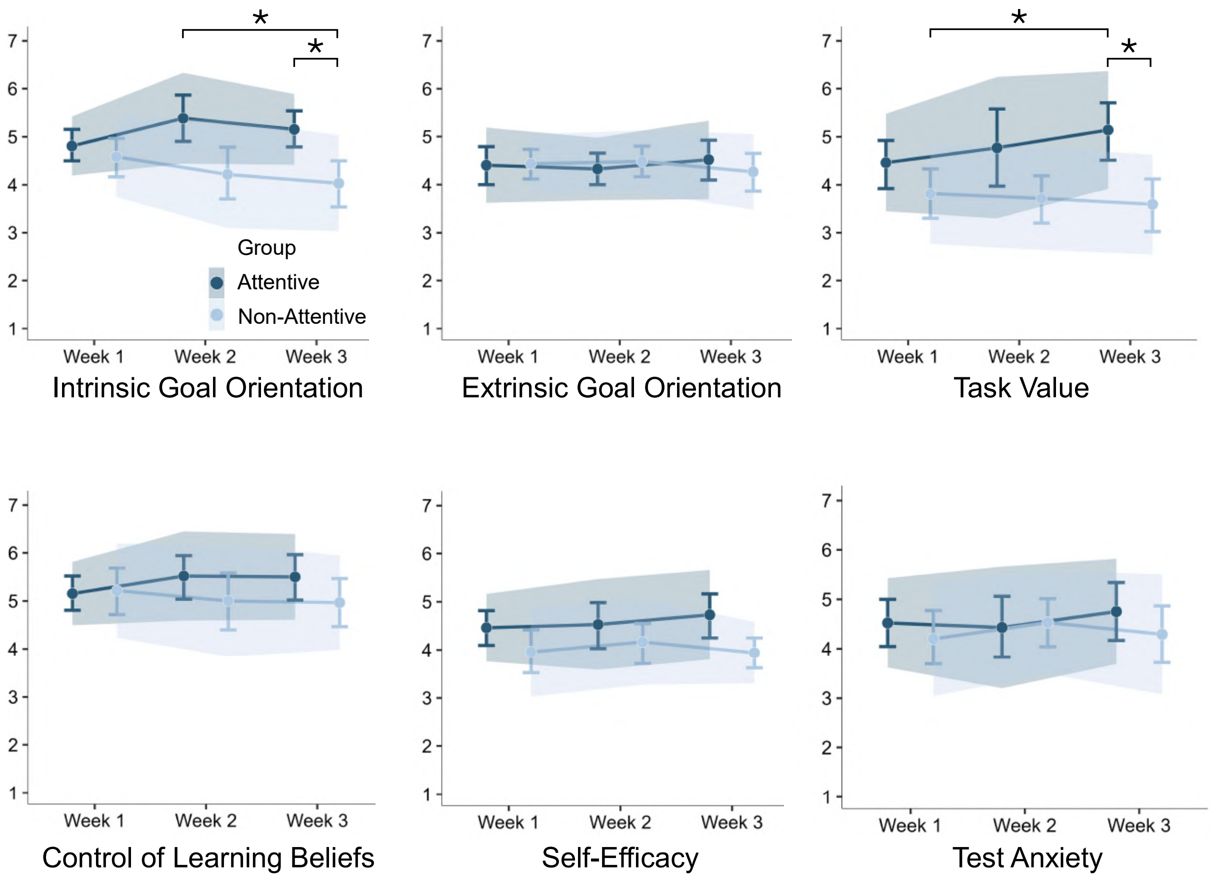

Figure 5.8: Mean scores on the six MSLQ subscales for attentive and non-attentive groups across three weeks. Shaded ribbons represent standard deviations, error bars denote 95% confidence intervals, and brackets indicate significant group differences (* $p < 0.05$).

intrinsic goal orientation, **self-efficacy**, **task value**. The assumption of sphericity was met for all analyses.

Intrinsic Goal Orientation For intrinsic goal orientation, both analyses indicated a strong between-subject effect of *Group*, $F(1, 26) = 12.610, p = 0.001, \omega^2 = 0.177, p_{\text{Holm}} = 0.006$, with the attentive group scoring higher than the non-attentive group. The Bayesian model comparison converged, providing strong evidence for including *Group* ($BF_{\text{incl}} = 13.91$), and a direct group comparison showed decisive evidence for higher scores in the attentive group ($BF_{10} = 330.3$). In contrast, neither approach indicated meaningful effects of *Week* ($F(2, 52) = 0.472, p = 0.626, \omega^2 = 0.000$; $BF_{\text{incl}} = 0.272$) or of the *Week* $\times$ *Group* interaction ($F(2, 52) = 3.114, p = 0.053, \omega^2 = 0.039$; $BF_{\text{incl}} = 0.887$).

Self-Efficacy For self-efficacy, both analyses revealed a between-subject effect of *Group*, $F(1, 26) = 5.035, p = 0.034, \omega^2 = 0.070, p_{\text{Holm}} = 0.136$, with the attentive group reporting higher self-efficacy ($M = 4.57, SD = 0.86$) than the non-attentive group ($M = 4.02, SD = 0.82$). Bayesian model comparison provided converging evidence, yielding anecdotal support for including *Group* ($BF_{\text{incl}} = 1.477$), and a direct group comparison indicated strong evidence for higher self-efficacy in the attentive group ($BF_{10} = 11.23$). In contrast, neither approach indicated meaningful effects of *Week* ($F(2, 52) = 0.371, p = 0.692, \omega^2 = 0.000$; $BF_{\text{incl}} = 0.120$) nor of the *Week* $\times$ *Group* interaction ($F(2, 52) = 0.752, p = 0.477, \omega^2 = 0.000$; $BF_{\text{incl}} = 0.096$).

Task Value For task value, there was a significant between-subject effect of *Group*, $F(1, 26) = 9.586, p = 0.005, \omega^2 = 0.137, p_{\text{Holm}} = 0.025$, with higher scores in the attentive group ($M = 4.79, SD = 1.25$) compared to the non-attentive group ($M = 3.70, SD = 1.05$). This was supported by Bayesian model comparison, which provided moderate evidence for including *Group* ($BF_{\text{incl}} = 6.848$), and by a direct Bayesian comparison showing decisive evidence for higher task value in the attentive group ($BF_{10} = 548.12$). Neither analysis indicated effects of *Week* ($F(2, 52) = 0.533, p = 0.590, \omega^2 = 0.000$; $BF_{\text{incl}} < 0.2$) or of the *Week* $\times$ *Group* interaction ($F(2, 52) = 2.074, p = 0.136, \omega^2 = 0.012$; $BF_{\text{incl}} < 0.2$).

Qualitative Results Thematic analysis further clarified how attention-aware support influenced students' motivation, resulting in **Theme 4: Motivation Through Goal-Directed Learning and Transferable Skills**.

Participants in the attentive group emphasized that *LectureQuiz* made their attention gaps explicit, which helped them define concrete goals for review. As P1 explained, identifying gaps provided "a clear goal of what to do after the lecture", enabling targeted Q&A with *OpenChat*. This active goal-setting resonates with the higher self-efficacy found in survey data. Meanwhile, non-attentive participants described being "not sure what to do" (P3) after lectures, sometimes engaging with the assistant on less relevant or already familiar content. Their gaps often only surfaced during the *Mini-Exams*, which some described as "frustrating" (P3).

Moreover, participants framed their motivation less around the specific lecture topic of Bayesian data analysis and more around the perceived value of refining their learning strategies. Seven of 10 interviewees stated the topic itself was not of primary interest. Yet, attentive group participants emphasized that the system still improved their motivation by helping them develop transferable skills. As P7 reflected, becoming aware of gaps and addressing them was "useful for real lectures and exams", even if the content itself was not central to their studies. This interpretation aligns with survey findings of higher task value in the attentive group: students perceived the system as helping them not only retain knowledge but also improve transferrable learning skills.

By revealing attention gaps and supporting targeted review, *AttentiveLearn* improved participants' self-efficacy and perceived task value, motivating them through clearer goals and transferable learning skills.

5.5.5 Learning Outcomes

To examine learning outcomes, we analyzed the results of the intermediate *Mini-Exams* in the mobile application and the final on-site exam.

Mini-Exams For the mini-exams, the mixed ANOVA revealed a significant within-subject effect of *Week*, $F(2, 52) = 5.989, p = 0.005, \omega^2 = 0.052$, with strong supporting evidence from the Bayesian analysis ($BF_{\text{incl}} = 11.63$). In addition, both approaches indicated a *Week* $\times$ *Group* interaction, $F(2, 52) = 3.413, p = 0.040, \omega^2 = 0.026; BF_{\text{incl}} = 3.81$, suggesting that

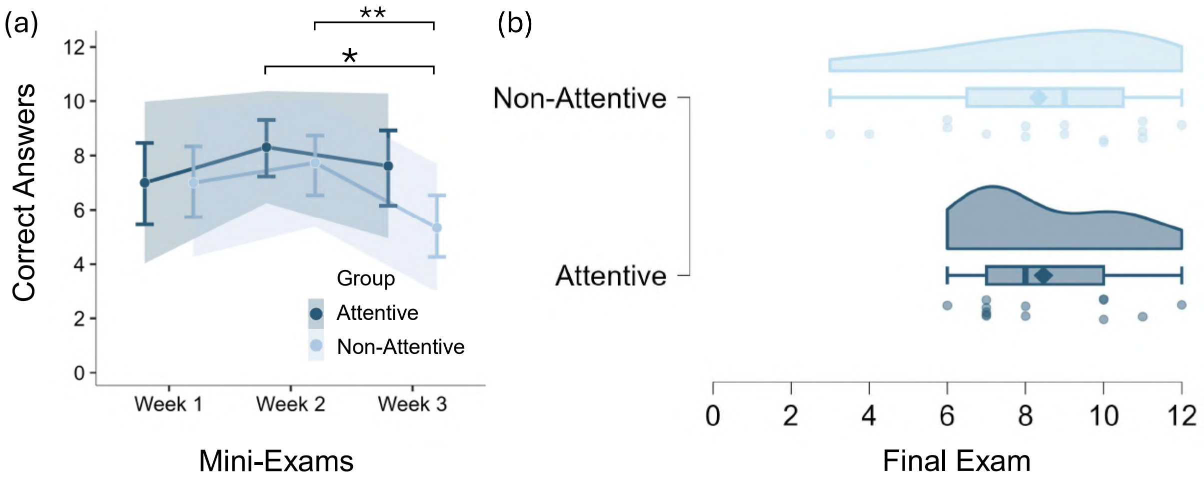

Figure 5.9: Performance comparison between groups. (a) Mini-exam scores across the three weeks of the study. Asterisks denote significant differences (* $p < 0.05$, ** $p < 0.01$). Shaded ribbons represent standard deviations; (b) Final exam scores in week 4.

performance changed differently across groups over time. Post-hoc comparisons confirmed that the non-attentive group's scores declined significantly from week 2 ($M = 7.73, SD = 2.34$) to week 3 ($M = 5.33, SD = 2.35$), $p = 0.004$, with Bayesian tests providing strong evidence for this decline ($BF_{10} = 32.54$). In contrast, the attentive group's scores remained stable (week 2: $M = 8.31, SD = 2.06$; week 3: $M = 7.62, SD = 2.66$). These results suggest that attention-aware personalization buffered against performance decline in later weeks.

Final Exam For the final exam, an independent-samples t-test found no significant difference between the attentive group ($M = 8.46, SD = 1.90$) and the non-attentive group ($M = 8.33, SD = 2.69$), $t(26) = 0.143, p = 0.887$. Bayesian analysis also supported no evident difference between the groups ($BF_{10} = 0.357$), indicating the data were approximately 2.8 times more likely under the null hypothesis.

These results show that participants benefited from attention-aware support in terms of sustaining performance during intermediate assessments, though the final on-site exam shows no significant differences.

5.5.6 Appropriation of Support Features

OpenChat* and *ChatQuiz We collected and analyzed the interaction logs of 31 participants (14 attentive, 17 non-attentive) who completed the three-week lectures and used the mobile assistant. Of these, three neither attended the final exam nor submitted the surveys, resulting in a final sample of 28 participants for the previous analyses.

On average, participants in the attentive group spent 150.83 minutes ($SD = 221.1$) per week using the mobile assistant, compared to 139.1 minutes ($SD = 340.04$) in the non-attentive group. This difference was not statistically significant. However, it may reflect higher engagement in the attentive group, consistent with the survey results.

For *OpenChat*, message activity varied significantly across weeks, $F(2, 58) = 8.91, p < .001$, but showed no group difference, $F(1, 29) = 0.32, p = .58$, and no interaction effect, $F(2, 58) = 0.26, p = .77$ (Figure 5.10b). For *ChatQuiz* scores (Figure 5.10a), mixed ANOVAs revealed no significant effects of group or week (*Group*: $F(1, 29) = 0.09, p = 0.77$; *Week*: $F(2, 58) = 0.54, p = 0.58$; *Interaction*: $F(2, 58) = 2.17, p = 0.13$).

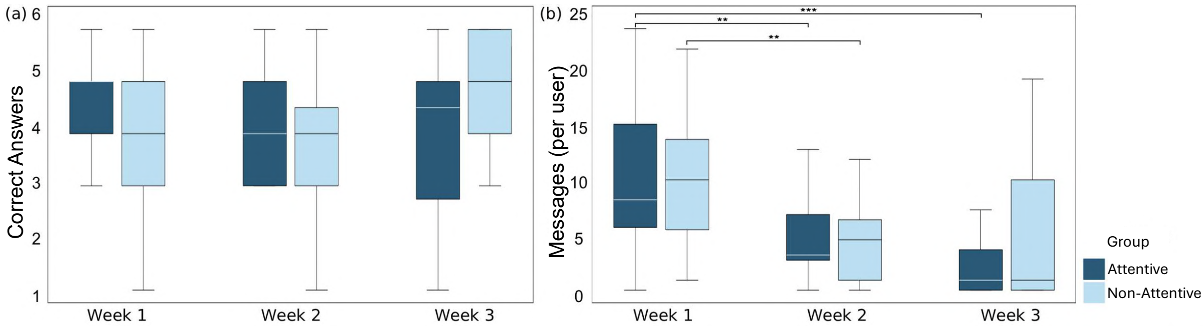

Figure 5.10: Boxplots of interaction data]: (a) Number of correctly answered questions in *ChatQuiz* module; (b) Number of messages sent in *OpenChat* module per week. Asterisks denote significant differences (** $p < 0.01$, *** $p < 0.001$).

Qualitative Results Thematic analysis revealed **Theme 5: Tool Appropriation for Diverse Learning Strategies**. Five codes within this theme highlighted how participants appropriated the available features in distinct ways, reflecting their personal study practices.

For *OpenChat*, some participants (e.g., P7, P15, P19) used the assistant in a highly targeted manner, entering with specific questions or hypotheses and seeking confirmation. Others—especially in the non-attentive group—reported using it more exploratively to probe for gaps and discover unfamiliar content (e.g., P3, P21). Similar appropriation patterns were observed in *ChatQuiz*. Some positioned it as an exam-preparation step, completing it immediately before the weekly *Mini-Exam*, while others used it diagnostically to probe their understanding before returning to *OpenChat* for clarification. As P24 explained:

> I completed the (Chat)Quiz first, went back (to *OpenChat*) and consolidated the knowledge. I would then take a break before the weekly exam in the app.

Participants also suggested additional features to extend *AttentiveLearn*, including memorization aids such as spaced repetition with flashcards (P1, P15, P21, P24), automated lecture summaries (P1, P13), and practice exam generation (P15). While beyond the scope of the current design, these suggestions reflect diverse learner strategies and highlight opportunities to expand personalization beyond attention.

These findings suggest that *AttentiveLearn* was flexibly appropriated to align with learning practices, while participants' suggestions indicate opportunities for extending *AttentiveLearn* with broader personalization.

5.6 Discussion

5.6.1 Interpretation of Results

Attention Improvement Our eye-tracking analyses confirmed that both groups began the study with comparable attention levels. The observed fluctuations in AOI coverage and ADI throughout the lecture validate existing research (Han et al., 2022), which reveals challenges of maintaining attention in immersive VR lectures.

Although the personalized *AttentiveLearn* did not produce a significant overall difference in attention trajectories between groups, exploratory analyses suggested potential benefits for low-attention participants. These students showed a tendency toward greater improvement over the three weeks when using *AttentiveLearn*. The interviews provide additional context for this pattern. Participants described two possible effects: greater confidence during lectures and a stronger sense of responsibility to sustain attention in order to perform well on the quiz. Taken together, these findings tentatively indicate: ***AttentiveLearn* not only facilitated post-lecture review, but may have also contributed to attention management during subsequent lectures.**

Motivation and Engagement Quantitative measures indicate that *AttentiveLearn* supported higher levels of engagement and motivation, particularly in the motivational dimensions of intrinsic goal orientation, task value, and self-efficacy, as well as the engagement factors of reward and focused attention. Other subscales did not differ across groups, which aligns with the fact that our design was not aimed to directly influence those dimensions. This demonstrates that the validated benefits of personalized quizzes, as shown in the work of McDaniel et al. (2012), can also be extended to an attention-aware immersive learning context. Because participants were not explicitly informed about the attention-based mechanism beforehand, any observed differences between groups should not be primarily attributed to participants' belief in the attentive group that they were being monitored more closely. This helps rule out expectation bias as an explanation for the observed motivational and engagement effects.

The combined interpretation of log data, survey results, and interviews suggests that revealing attention gaps and enabling participants to address them through post-lecture support made learning challenges more visible and actionable. This process enhanced participants' sense of task value, consistent with prior research showing that solving identifiable problems fosters motivation (Pintrich, 1991). In our study, this improvement in task value was achieved both through better retention of lecture content and the development of transferrable skills.

At the same time, participants who used the personalized *AttentiveLearn* reported greater self-efficacy and higher focused attention once they became aware of their attention gaps. In contrast, those in the non-attentive condition often described treating the lectures more casually, as the topic was not central to their study programs. This observation aligns with broader research on how attention-aware systems can enhance situational interest and attention management (Roda and Thomas, 2006; Vertegaal et al., 2006).

We also observed context-dependent dynamics. In week 3, coinciding with the exam phase at the universities and the increased difficulty of the VR lecture content, both groups reported higher learning pressure and lower overall MSLQ scores compared to the previous week. However, the attentive group maintained relatively higher motivation and engagement. This suggests that *AttentiveLearn* helped participants regulate their motivation more effectively under heightened external demands. **Overall, *AttentiveLearn* improved participants' motivation and engagement throughout the study, even when they were under pressure.**

Learning Outcomes The intermediate *Mini-Exams* showed that participants using *AttentiveLearn* achieved higher performance, with the performance gap widening across the three-week lectures. This effect was most pronounced in week 3, when task difficulty and external pressures were greatest. Supporting evidence comes from higher reported confidence in interviews, suggesting that personalized support improved comprehension and sustained learning progress. These findings extend prior work showing that effective quizzes enhance knowledge retention after lectures (McDaniel et al., 2011).

However, these benefits did not extend to the final exam. Two factors may explain this discrepancy. First, although students could revisit modules from previous weeks, *AttentiveLearn* didn't provide an overarching exam-preparation feature. Without it, learners may have faced challenges to consolidate knowledge across the full course. Second, as observed in other field studies (Alazmi and Alemtairy, 2024; Cheng and Tsai, 2019), final learning outcomes can be heavily influenced by students' broader curriculum and other academic demands. In our study, the exam phase at the local universities constrained participants' time for review in the final week. This challenge mirrors prior research showing that measuring long-term learning outcomes is difficult in immersive learning studies, particularly in real-world settings (Petersen et al., 2021). **Therefore, *AttentiveLearn* enhanced short-term comprehension and helped sustained progress, but its long-term impact was constrained and requires further investigation.**

5.6.2 Limitations

We recognize several limitations of our study. First, the sample size of 36 participants, combined with an unbalanced gender distribution, restricts the generalizability of findings. While this reflects the actual demographics of students enrolled in the two local universities, future work with a larger and more diverse sample can strengthen external validity. To mitigate the limitation, we employed Bayesian statistical analysis, a method increasingly used in immersive learning research with moderate sample sizes (Liu et al., 2025b).

Second, the lecture length in our study was around 20 minutes. Although this aligns with existing design recommendations (Horizon, 2025), real-world lectures often extend far longer. As current research lacks validated design guidelines for VR lecture duration, future work should investigate optimal session lengths and strategies for segmenting extended learning content for immersive learning. Another pedagogical limitation is the quiz design. While all questions used in the *Mini-Exams* and final exam were validated by a statistics expert and lecturer before the study, and all generated questions in the *LectureQuiz* and *ChatQuiz* modules were validated afterward for content alignment and pedagogical difficulty, we acknowledge that this still underrepresents the richer pedagogical intent present in real courses, where instructors hand-design non-personalized quizzes. Our comparison therefore isolates the effect of attention-aware personalization in an idealized experimental setting, rather than against a fully realistic classroom baseline.

Third, while our learning ecosystem integrated multiple support features (e.g., *OpenChat*, *ChatQuiz*), our evaluation focused specifically on attention-driven personalization of *Lecture-Quiz*. Between-subject comparisons of every design features were beyond the scope of this four-week field study. Future research should examine how different support mechanisms (e.g., attention-based and confusion-based quizzes, flashcards, tutoring, etc.) compare or complement one another when integrated into immersive learning pipelines. For attention-driven personalization, we also acknowledge important limitations. While visual attention is often used as a proxy for cognitive processing under the eye–mind link hypothesis (Just and Carpenter, 1980), this assumption does not always hold. As noted in prior work (Szafir and Mutlu, 2013), low attention levels indicated by biosignals (e.g., gaze, EEG) may reflect a learner's prior knowledge of the material rather than actual disengagement. To mitigate this effect, we measured participants' prior knowledge and they reported little familiarity with the lecture content, yet residual confounds cannot be ruled out. For this, Szafir and Mutlu (2013) recommend combining multiple assessment methods and biosignals to disambiguate such cases. Therefore, we validated the attention-driven post-lecture support using eye-tracking and complemented it with quizzes as an additional assessment method. However, our current system still does not prompt users with subjective confirmations of attention gaps during learning. Future work should integrate mechanisms that couple system-driven attention signals with user-reported focus and perceived difficulties, enabling more accurate and learner-centered personalization.

Finally, our study was conducted at our university with one lecture topic. Although this ensured ecological validity in an authentic setting, additional studies across varied domains, institutions, and cultural backgrounds are needed to establish robustness. Moreover, the use of non-credit lectures and compensated participation limits the generalizability of our findings to authentic university courses where motivation, stakes, and classroom dynamics differ.

5.6.3 Design Implications for Future Research

Extending Immersive Learning with Post-Lecture Support In this paper, we connected the virtual learning experience with a mobile learning assistant that delivers personalized post-lecture support. Our findings show that students in both groups continued engaging with the content after the VR session, revisiting material multiple times per week with the assistant. This indicates that even independent of gaze-adaptive personalization, extending immersive learning into post-lecture contexts increases engagement and fosters more continuous learning.

This finding aligns with ubiquitous and seamless learning frameworks (Dede, 2011; Milrad et al., 2013), which emphasize that learning should not be restricted to a single context or time, but instead span across settings and devices. Our findings also align with the design implications of existing research (Klaveren et al., 2017; Szafir and Mutlu, 2013), suggesting that revisiting content, especially when tailored to learners' attention, improves the general learning experience. Additionally, our work demonstrates how immersive learning can be complemented by mobile assistants to bridge in-situ and ex-situ learning. As Cheng and Tsai (2019) suggest, cross-device continuity is particularly valuable in creating an active and ubiquitous learning experience. Our findings reinforce this promise and point to design opportunities for cross-device immersive learning ecosystems.

Ex-situ Personalization for Immersive Experiences *AttentiveLearn* implements a pipeline that processed in-situ eye-tracking data to deliver ex-situ, attention-aware personalization. This approach extends earlier visions of cross-device interaction (e.g., (Brudy et al., 2019)) by not only bridging modalities but also adapting user experience across devices based on users' cognitive states. Current cross-reality research often highlights a tension between continuity and immersion, with post-experience supports potentially disrupting the sense of flow and immersion (Auda et al., 2023). Our work demonstrates that using in-situ attention data to inform ex-situ personalization can increase engagement and motivation after the immersive session, which may address this challenge in cross-reality interaction. We selected a virtual classroom as a starting point because it allowed established attention-support techniques, primarily studied in classroom or video-based learning, to be transferred into VR in a controlled and interpretable way.

This points to a broader research agenda on cross-reality personalization strategies. Beyond gaze, additional signals could also inform the adaptation of ex-situ support features in immersive learning and related domains. Personalization does not need to be limited to attention; other cognitive and affective aspects, such as empathy (Wambsganss et al., 2022), may also serve as valuable input for personalized support. Prior work has demonstrated that biosignal-driven personalization can enhance user experience in diverse contexts (Schultz and Maedche, 2023). Extending our pipeline to integrate such multimodal signals would enable richer personalization, offering new opportunities to investigate how ex-situ support, informed by in-situ data, fosters sustained engagement across cross-reality experiences. Future work should apply this approach to more embodied and interactive immersive environments, such as virtual laboratories and makerspaces (Radu et al., 2021b), skill-training scenarios with simulation (Zhu et al., 2025), to understand how ex-situ personalization can support learning beyond virtual classrooms.

Flexible Support Techniques for Different Learning Styles In our study, personalized quizzes were the primary support technique, effectively revealing knowledge gaps and guiding learners to address them. While participants recognized the value of quizzes, they also expressed interest in other support techniques, such as flashcards or automatically generated example exams. These suggestions reflect heterogeneous user needs and learning preferences, consistent with prior research on individual learning strategies (Makransky et al., 2021).

Future work can advance in two directions. First, quizzes themselves have broader potential beyond post-lecture review: prior studies show they can also enhance in-lecture engagement (Raes et al., 2020). Future immersive learning systems could therefore integrate quizzes dynamically within both in-situ and ex-situ contexts. Second, attention is only one cognitive dimension of learning (Makransky and Petersen, 2021). Other aspects such as memorization, comprehension, or decision-making may be supported through complementary techniques like spaced-repetition flashcards or gamification (Zhang et al., 2023). Combined with our ecosystem, such techniques could create a more comprehensive support system.

Improving Engagement and Motivation as Design Goals While many studies in immersive learning focus on learning outcomes (Baceviciute et al., 2020; Makransky et al., 2021; Radianti et al., 2020; Radu et al., 2021a; Szafir and Mutlu, 2013), our work shows the value of also targeting engagement and motivation. Indeed, learning outcomes alone may not capture the full challenges students face in immersive environments (Zhang et al., 2022). By designing for broader learner-centered goals such as motivation and engagement, which covers more aspects including task value, self-efficacy, and reward factors, systems such as *AttentiveLearn* can improve the overall quality of the learning experience, even when the subject matter itself is not directly relevant to students' curricula. For evaluating engagement and motivation throughout the learning process, we advocate conducting field studies that assess these factors over an extended period of learning. Existing research typically investigates single-session effects and focuses primarily on recall and learning outcomes (Blume et al., 2019; Szafir and Mutlu, 2013). In contrast, we conducted a four-week deployment within a real course context and additionally measured engagement and motivation, uncovering design implications for sustaining learning beyond the immersive session itself.

In our field study, participants using *AttentiveLearn* reported higher motivation and engagement even when they perceived the topic as peripheral to their degree programs. This suggests that designing for engagement and motivation may help sustain learner involvement across less intrinsically motivating content. Building on existing research into in-lecture support for engagement (Han et al., 2022), future work should investigate how in-situ and ex-situ supports can be integrated to maintain motivation and engagement throughout the entire learning journey. Such integration would broaden the scope of immersive learning design, shifting from outcome-focused designs to more holistic, learner-centered experiences.

5.7 Conclusion

In this paper, we presented *AttentiveLearn*, an attention-aware learning ecosystem that extends immersive learning with post-lecture support delivered through a mobile learning assistant. Leveraging attention metrics derived from eye-tracking, the system generates personalized quizzes to help learners address attention gaps and integrates these with complementary features in the mobile application. We conducted a four-week field study with university students to investigate the effects of *AttentiveLearn* on motivation, engagement, and learning outcomes. Our results show that *AttentiveLearn* enhanced participants' motivation and engagement—particularly through improvements in perceived self-efficacy and task value—as well as contributed to better intermediate learning outcomes. With these findings, we contribute empirical evidence to HCI research on learner-centered support for immersive learning. While prior work has predominantly focused on in-lecture interventions, our study demonstrates the value of extending immersive learning beyond the lecture through ex-situ personalization. Looking forward, we see opportunities for personalized support with other immersive and cross-reality ecosystems. We believe this work can contribute toward making immersive learning a more ubiquitous, personalized, and learner-centered experience.

6 Discussion

Ongoing research on immersive learning, which leverages technologies such as AR, VR, and MR, has contributed to transforming learning activities from relatively passive online content consumption to active situated experiences. As outlined in Chapter 1, the core premise of immersive learning lies in its capacity to foster *presence* and *situated cognition* (Mohammadhossein et al., 2022). While acknowledging the potential of immersive learning, my research was driven by the recognition of a critical challenge in immersive system design: the affordances that make these environments effective can also impose cognitive demands on learners (C1). Without learner-centered adaptive support, immersive systems risk causing cognitive overload rather than facilitating deep understanding (Kockord and Bodensiek, 2021).

Guided by the CAMIL framework (Makransky and Petersen, 2021), *attention* has been identified as a central moderator between the technological affordances of immersive technologies and successful learning outcomes. Yet, existing research and design knowledge on gaze-adaptive support for attention management in immersive learning are scarce (C2). Therefore, I propose the overarching goal of this dissertation as a research question: *How can gaze-adaptive support be designed to effectively assist learners in immersive learning?* To answer this question, I followed a research roadmap spanning the *Reality–Virtuality Continuum* (Milgram and Fumio, 1994), investigating immersive learning systems implemented with both MR and VR technologies. Furthermore, the studies examine multiple gaze-adaptive support techniques, including attention feedback, note-taking, and post-lecture quizzes. Finally, two learning formats have been investigated as contexts for gaze-adaptive support: self-directed exploratory learning and lecture consumption in virtual classrooms.

Study I addressed the lack of a comprehensive understanding of immersive learning systems, particularly the subset of head-mounted MR learning systems (Gap 1), by reviewing the state of the art of MR systems in higher education. This review revealed that while technical feasibility has promoted the adoption of MR learning systems, pedagogical design implications for adaptive support remains scarce. **Study II** addressed the need for in-situ support (Gap 2) by designing and evaluating *AF-Mix*, a system demonstrating how gaze-based feedback can facilitate attention management during learning sessions without removing the head-mounted MR device. Recognizing that learning is not an isolated event (C3, Gap 3), **Study III** and **Study IV** expanded the scope to *cross-device* ecosystems and other immersive technologies (i.e., immersive VR). Study III bridged the gap between immersive and external environments through cross-device note-taking, while Study IV extended support throughout the learning journey, utilizing eye-tracking data captured during immersion to personalize ex-situ review activities.

In the following sections, I discuss the theoretical and practical contributions of my dissertation, drawing on the classification of HCI research contributions proposed by Wobbrock and Kientz (2016). Section 6.1 reports the theoretical contributions that extend the understanding of gaze-adaptive support in immersive learning. Adapting the contribution types outlined by Wobbrock and Kientz (2016), I describe the following types of contributions. First, a **conceptual framework**, as one type of survey contribution, is derived from a systematic review of prior work on a research topic, with the goal of "exposing trends and gaps" (Wobbrock and Kientz, 2016, p. 42). Second, a **design space** provides an overview of the characteristics of an artifact design, revealing potential design possibilities. Third, empirical contributions take the form of **descriptive knowledge**, which facilitates understanding of HCI-related phenomena in immersive learning, or **prescriptive knowledge**, which provides guidelines for improving user experience and other desired outcomes by addressing identified user challenges.

Section 6.2 outlines the practical contributions. Here, I present the following contribution types. First, an **artifact contribution** highlights the "creation and realization of interactive artifacts" that "reveal new possibilities, enable new explorations, facilitate new insights, or compel us to consider new possible futures" (Wobbrock and Kientz, 2016, p. 40). Second, **methodological contributions** inform future research by describing how to perform research activities. Lastly, **design implications** can assist HCI researchers and practitioners in creating more learner-centered adaptive support.

6.1 Theoretical Contributions

This dissertation makes several theoretical contributions that advance the understanding of learner-centered, gaze-adaptive support in immersive learning systems. These contributions are primarily positioned within the research field of HCI, providing a conceptual framework, a design space, as well as prescriptive and descriptive knowledge. Table 6.1 provides an overview of these theoretical contributions. In the subsequent section, each contribution is described in detail, and I briefly discuss the results of the studies from which these contributions were derived.

In **Study I**, I investigated RQ1: *What is the state of the art of mixed reality learning systems in higher education?* To address this question, I conducted a systematic literature review following the PRISMA guidelines (Page et al., 2021) and analyzed 80 research articles to characterize the research landscape of head-mounted MR in educational contexts. The main theoretical contribution of this study is an integrated conceptual framework. This framework structures the fragmented domain of MR learning along five dimensions: (1) technological characteristics, including device types; (2) learning theories and paradigms; (3) fields of education; (4) design features; and (5) research design, including evaluated outcomes and methodologies. With this framework, the study revealed a critical research gap: while current research heavily emphasizes technological novelty and usability, pedagogical grounding in system design remains limited. To identify this gap, I categorized the learning paradigms and theories used in the reviewed literature — such as constructivism, behaviorism, and embodied learning—based on existing classifications in education research (Schunk, 2012). The analysis shows that few articles explicitly leveraged established learning theories and paradigms to guide system design. Moreover, adaptive support is largely absent as a key design feature. In particular, biosignal-adaptive technologies, including gaze-adaptive systems with eye-tracking, have the potential of effectively supporting the cognitive processes in immersive learning. Overall, this contribution provides a conceptual foundation for future research by shifting the focus from technology-driven to learner-centered design. It also highlights the opportunity to design biosignal-adaptive systems to support learners in immersive learning. Finally, the framework can be used as a design space for immersive learning systems by outlining design possibilities that can inform MR learning systems designs.

Building on the conceptual framework and design space, **Study II** addressed RQ2: *How to design attention feedback for immersive learning to improve students' self-reflection?* To answer this question, I conducted a participatory design workshop to inform the conceptualization and implementation of *AF-Mix*, a gaze-aware MR system for HMDs, and evaluated it in a laboratory experiment with 22 participants. The evaluation provides an empirical research contribution in the form of descriptive knowledge about the interplay between visual attention feedback and learners' self-reflection in immersive environments. The results show that learning analytics based on attention distribution can promote self-reflection and content coverage during the review process; however, this increase does not automatically translate into immediate improvements in knowledge retention or learning success. This finding adds nuance to the CAMIL framework by suggesting that, although attention is an important cognitive factor for learning, a gap remains between informing learners about attention-related knowledge gaps and achieving measurable improvements in learning outcomes. In addition, the study provides prescriptive knowledge in the form of design rationales. Specifically, the results indicate that attention feedback derived from eye-tracking data can promote learners' self-reflection during review. Nevertheless, individual learning and review strategies should be considered when presenting such feedback. Finally, although this study primarily focuses on gaze-adaptive support in situ (i.e., during the learning session), the results also indicate that ex-situ support is crucial for learning success. Accordingly, subsequent studies investigated additional gaze-adaptive support techniques and provided further empirical evidence.

In **Study III**, I investigated RQ3: *How to design gaze-adaptive note-taking for immersive learning to support effective cross-device and context-aware information transfer?* To address this question, I designed *GazeNote*, a system that uses eye-tracking to synchronize context between a head-mounted display and a handheld tablet. The theoretical contribution of this study lies in extending the understanding of cross-device interaction within the *Reality–Virtuality Continuum* (Milgram and Fumio, 1994). It provides evidence that gaze data can serve as a modality to bridge *in-situ* immersive experiences and *ex-situ* review tasks. Existing research grounded in the theoretical framework of seamless and ubiquitous learning often leverages digital tools such as mobile phones to achieve desired learning outcomes — i.e., learning as a journey embedded in everyday activities (Hwang and Tsai, 2011; Milrad et al., 2013; Sharples, 2015). This study provides prescriptive knowledge in the form of design rationales, showing that by leveraging note-taking techniques and gaze data, immersive learning systems can be integrated into broader cross-device ecosystems to support ubiquitous and seamless learning. Furthermore, descriptive knowledge derived from behavioral analyses in a within-subject experiment indicates that preserving contextual information is crucial for note-taking support, whereas gaze-aware heatmaps combined with cross-device notes provide only marginal benefits. Given the interdisciplinary nature of note-taking research (Kiewra et al., 1991), these findings can guide future work in HCI, educational psychology, and information science aimed at optimizing techniques for context-aware cognitive offloading and effective information retrieval.

In **Study IV**, I addressed RQ4: *How to design gaze-adaptive post-lecture support for immersive learning to improve engagement, motivation, and learning outcomes?* In contrast to the previous studies, I investigated this question through a four-week field study with 36 participants. I developed the learning ecosystem *AttentiveLearn*, which consists of a gaze-aware virtual classroom in VR, a personalization pipeline that computes attention metrics and informs gaze-adaptive quiz questions, and a mobile assistant that delivers the quiz modules. The system evaluation provided empirical evidence on the efficacy of gaze-informed attention metrics for longitudinal adaptive support. The results indicate that participants who received personalized support based on their specific attention gaps achieved significantly higher motivation, engagement, and learning gains across the lecture weeks than those who received non-personalized quiz questions. Notably, learners who were less attentive during the initial lecture week benefited more from the gaze-adaptive quizzes and showed improved attention levels in subsequent lectures. These findings provide descriptive knowledge that gaze-adaptive quizzes can increase motivation and engagement throughout the learning journey, while their effects on attention management are moderated by individual learner characteristics. Furthermore, the study offers prescriptive knowledge in the form of design rationales, demonstrating that gaze data collected in immersive VR systems can be used to model a learner's attention state and drive subsequent learning activities, including review and self-testing on a mobile device. In doing so, *AttentiveLearn* bridges the gap between the immersive learning experience and the ex-situ review process.

In summary, the four studies provide both descriptive and prescriptive knowledge that advances understanding of learner-centered, gaze-adaptive support in immersive learning. First, this work structures the previously fragmented MR learning landscape through a conceptual framework and design space. Second, the empirical evidence adds nuance to the CAMIL framework, indicating that gaze-adaptive support must go beyond simple attention guidance to facilitate the cognitive processing required to improve learning outcomes. Third, the research extends the scope of immersive support to cross-device ecosystems, showing that gaze can effectively bridge in-situ immersion and ex-situ productivity tasks to facilitate seamless learning. Fourth, in-the-wild evidence validates the efficacy of gaze-adaptive support by demonstrating that personalization based on attention metrics enhances learning outcomes and motivation, particularly for learners with lower initial attention.

Table 6.1: Overview of theoretical contributions

	Theoretical Contributions
Study I	**Conceptual Framework & Design Space:** An integrated framework that structures MR learning systems in higher education across five dimensions (technology, learning theories/paradigms, field of education, design features, and research design), while also outlining a design space for immersive learning system design.
Study II	**Descriptive Knowledge:** Empirical evidence that attention feedback can promote learners' self-reflection and content coverage during review, but does not necessarily yield immediate gains in retention or learning success. **Prescriptive Knowledge:** Design rationales for presenting gaze-based attention feedback to support self-reflection, highlighting the need to account for individual learning/review strategies when delivering attention feedback.
Study III	**Descriptive Knowledge:** Evidence that preserving contextual information is crucial for note-taking support, whereas gaze-aware visualizations (e.g., heatmaps) provide only marginal benefits when combined with cross-device notes. **Prescriptive Knowledge:** Design rationales for integrating gaze-adaptive note-taking into cross-device ecosystems to support seamless and ubiquitous learning and context-aware cognitive offloading.
Study IV	**Descriptive Knowledge:** In-the-wild evidence that gaze-adaptive quizzes based on attention gaps can increase motivation, engagement, and learning gains compared to non-personalized quizzes; benefits are stronger for learners with lower initial attention, and effects are moderated by individual learner characteristics. **Prescriptive Knowledge:** Design rationales for longitudinal, gaze-informed post-lecture support that uses attention metrics from immersive sessions to drive subsequent review and self-testing on a mobile device, thereby bridging the temporal gap between in-situ and ex-situ learning.

6.2 Practical Contributions

In addition to the theoretical contributions presented in the previous section, this dissertation provides actionable practical insights for designers, researchers, and educators. Across the studies, I developed novel systems and interaction techniques that not only demonstrate the technical feasibility of gaze-adaptive support, but also reveal design implications for practitioners. Building on this knowledge, researchers and educators can create learner-centered immersive learning experiences that improve learners' motivation, engagement, and learning outcomes. Table 6.2 provides an overview. In the following sections, I describe the specific practical contributions and design implications of each study.

In **Study I**, the systematic literature review reveals a set of practical design implications for implementing MR learning systems. By examining 80 existing studies, this work offers an integrated resource to support practitioners in making design decisions about immersive technologies. In particular, it highlights trade-offs between Optical See-Through (OST) and Video See-Through (VST) head-mounted displays, depending on the educational domain and learning tasks. Furthermore, the review identifies a practical gap in current evaluation methodologies for MR learning systems: an over-reliance on simple usability metrics (e.g., the SUS questionnaire) as the primary measured construct. By contrast, cognitive constructs such as engagement and attention — which can better capture learners' experiences during immersive learning sessions — are measured only to a limited extent. Similarly, study designs in the reviewed articles rarely assess actual learning outcomes. The studies that did measure learning outcomes often rely on basic instruments such as knowledge tests with a few items that hadn't been validated before. This contrasts with established educational practice, which advocates comprehensive assessment using validated instrument across the entire learning journey, from the learning session itself to the review process (McDaniel et al., 2012; Raes et al., 2020). Therefore, the study contributes methodological recommendations toward a more comprehensive evaluation approach that assesses immersive learning from a learner-centered perspective and accounts for different cognitive aspects. Finally, biosignal-adaptive technologies could be leveraged both as design features for providing learning support and as assessment instruments, enabling a better understanding of cognitive processes such as attention management in immersive learning.

Study II provides a novel artifact contribution through *AF-Mix*, a gaze-aware learning system implemented on the Microsoft HoloLens 2. This artifact demonstrates the technical feasibility of accessing and processing raw eye-tracking data on a standalone head-mounted MR device to generate attention feedback as a design intervention. The system serves as a reference implementation for developers, showcasing a closed-loop architecture in which gaze data are captured, analyzed, and transformed into attention feedback within the Unity engine, in line with, and preceding, more elaborate frameworks such as one proposed by Abeysinghe et al. (2025). Although the artifact functions as a standalone learning system, the attention-feedback component was designed to be transferable to other immersive learning systems as a modular extension. This design principle of modularity informs the dissertation: the artifacts in Stud-

Table 6.2: Overview of practical contributions

	Practical Contributions
Study I	**Design Implications:** Practical guidance for design decisions regarding the technology, as well as varying interaction levels suited for different learning tasks and contexts. **Methodological Contribution:** Recommendations for a more comprehensive, learner-centered evaluation approach that incorporates cognitive constructs (e.g., engagement and attention), validated instruments, and assessment across the learning journey (in and ex-situ). Biosignals are positioned as both support mechanisms and potential assessment instruments.
Study II	**Artifact Contribution:** *AF-Mix*, a gaze-aware MR learning system on HoloLens 2 that demonstrates the technical feasibility of accessing and processing raw eye-tracking data on-device and delivering closed-loop attention feedback in real time within Unity. **Methodological Contribution:** An exemplary application of participatory design for immersive system design. **Design Implications:** Transforming raw gaze data into attention metrics and feedback that supports learners in retrieving missed information without introducing additional cognitive overhead; a principle of modular design that enables gaze-adaptive components to be integrated within other immersive systems.
Study III	**Artifact Contribution:** *GazeNote*, a modular cross-device system architecture that connects an MR learning system with a secondary mobile device (smartphone or tablet) to support context-synchronized note-taking and review. **Design Implications:** cross-device design recommendations for capturing contextual information during immersive learning and supporting ex-situ review with note-taking (e.g., leveraging familiar note-taking GUIs, gaze-aware context linkage, and PDF export for retrieval).
Study IV	**Artifact Contribution:** *AttentiveLearn*, a learning ecosystem that integrates a gaze-aware VR lecture environment with a companion mobile learning assistant, enabling longitudinal post-lecture support. **Methodological Contribution:** A workflow showing how biosignals or personal data collected on one device can drive personalized experiences on another device, supporting hybrid learning setups. This methodology can potentially support both learner-centered and instructor-centered goals. **Design Implications:** A practical pipeline for generating attention-driven, gaze-adaptive quizzes and review materials via LLM prompting based on attention metrics, thereby bridging synchronous immersive lectures with asynchronous mobile review and self-testing.

ies III and IV are similarly designed to be flexibly integrated as components within existing immersive learning ecosystems, demonstrating the generalizability of gaze-adaptive support.

Additionally, the study contributes methodological insights from the participatory design process, in which the attention-feedback visualization was co-designed with 15 participants. Participatory design for immersive systems has historically been limited, as established design methods are previously often applied to 2D interface design (GaliOlga et al., 2024). Accordingly, this study serves as a methodological example of applying participatory design specifically to immersive system design. Finally, the implemented system demonstrates a pipeline for transforming raw gaze data into attention metrics that help learners retrieve missed information without introducing additional cognitive overhead, providing a blueprint for attentional learning analytics in immersive learning support.

In **Study III**, addressing the user challenge of retrieving contextual information for later review after an immersive learning session, I developed *GazeNote*, which is another artifact contribution. *GazeNote* is a cross-device system architecture that enables interaction between a head-mounted display and an external mobile device (smartphone or tablet). From a software engineering perspective, the system contributes a modular extension for Unity-based immersive learning systems developed for MR headsets. This architecture addresses the challenge of capturing contextual information during the learning process and the research gap (Gap 3) concerning the lack of designs that bridge immersive learning and the ex-situ review process.

Specifically, the system connects the MR headset with a secondary mobile device, allowing users to leverage the mobile device's familiar interface (a web-based note-taking tool) for note-taking while maintaining a sense of immersion. This provides HCI practitioners with a replicable cross-reality design pattern and practical design implications, demonstrating how to support complex learning workflows — such as information capture and retrieval for seamless learning — through cross-device interaction techniques (e.g., gaze-aware attention heatmaps and PDF generation) without requiring learners to leave the immersive simulation during note-taking.

Finally, **Study IV** contributes *AttentiveLearn* as an additional *artifact contribution*. *AttentiveLearn* is a learning ecosystem that integrates gaze-aware VR lectures with an LLM-based mobile learning assistant. The system demonstrates a practical end-to-end pipeline for generating gaze-adaptive, attention-driven quizzes: attention metrics inferred from in-lecture eye-tracking data are processed and then used to prompt an LLM to generate personalized quiz questions and review materials. The quiz and reviews modules are then delivered via a companion mobile assistant application.

Beyond demonstrating technical feasibility, the artifact and its four-week field evaluation yield design implications for bridging immersive learning experiences with the ex-situ review process. The ecosystem also provides educators with a methodological approach for integrating immersive learning formats into hybrid learning activities, which have strong potential for real-world adoption (Liu et al., 2023). More importantly, it illustrates how biosignals and personal data collected on one device (e.g., gaze data collected with VR headsets) can enable personalized learning support on another device that lacks the sensing capabilities required for such personalization (e.g., smartphones without eye trackers).

This methodological contribution benefits both learners and educators. For learners, attention-informed personalization supports self-reflection and more focused review. For educators, it offers a foundation for attention-aware learning analytics and learning management across lecture and post-lecture phases. In HCI, such dual perspectives — supporting both learner-centered and instructor-centered goals — have been explored previously (e.g., by Thanyadit et al. (2023)). My work advances this field of research by extending it with cross-device data transfer and interaction. In conclusion, it provides researchers with a validated workflow for longitudinal gaze-adaptive support that links immersive learning in the virtual classrooms with asynchronous mobile-based review activities.

In summary, the practical contributions of this dissertation provide a toolkit of systems and pipelines for learner-centered, gaze-adaptive support. They include methodological contributions — such as the proposed evaluation strategies in Study I and the exemplary participatory design approach in Study II — as well as artifacts that incorporate architectures for attention feedback (Study II), cross-device interaction (Study III), and post-lecture quizzes (Study IV). Together, these artifacts and the derived design implications reduce technical and design barriers for future researchers and practitioners seeking to implement intelligent immersive learning environments and gaze-adaptive learning support.

6.3 Limitations and Future Research

The studies in this dissertation aim to rigorously address the challenges and research gaps identified in prior work on immersive learning. Nevertheless, several limitations and threats to validity remain. This section discusses them and outlines directions for future research.

First, the sample size, statistical power, and longitudinal scope were limited in the conducted studies. Studies II – IV used mixed-method evaluations. While the study designs and evaluation plans were aligned with their respective research questions and underwent peer review, they involved relatively small samples ranging from 18 to 36 participants. In HCI, such sample sizes are common and are not necessarily considered a critical limitation (Caine, 2016). This pattern of limited sample sizes was also reflected in Study I, which analyzed sample sizes reported in the reviewed research articles on MR learning systems. However, from the perspective of educational research, larger samples are often needed to obtain more conclusive evidence about learning outcomes and how the intervention influenced students' learning behavior. Across the dissertation, there is consistent evidence that gaze-adaptive support is perceived as useful and can facilitate key learning processes and tasks, including self-reflection (Study II), contextual information capture and transfer (Study III), and post-lecture review (Study IV). In addition, positive effects on user experience, motivation, and engagement were also consistently observed. However, evidence for improvements in learning outcomes is more limited. In Study IV, intermediate assessments over the first three weeks showed a positive trend in learning gains; nevertheless, the final exam did not yield a statistically significant difference between the attentive and the non-attentive conditions. Future work should therefore evaluate the proposed gaze-adaptive support techniques with larger samples to improve statistical power and to better estimate the magnitude and robustness of learning effects. Related to this, educational research suggests that changes in learning behavior and measurable learning gains often emerge over extended periods (Raes and Depaepe, 2020). Although Study IV includes an in-the-wild evaluation, practical constraints including ethics and data-privacy considerations prevented deploying the system within an actual course embedded in students' curricula over a longer timeframe (e.g., throughout an entire semester). The other studies were not longitudinal field studies. Future research should thus conduct longer-term studies in authentic educational settings to assess behavioral change, knowledge retention, and learning outcomes.

Second, the scope of learning content, participant population, and immersive technology remain limited. To strengthen external validity, the studies covered multiple immersive technologies and learning formats. Nonetheless, the scope remains limited in several ways. First, the learning content in Studies II and III focused on classic HCI concepts (human information processing), whereas Study IV used Bayesian analysis. While these topics are interdisciplinary and relevant across study programs, it remains unclear to what extent the proposed gaze-adaptive support techniques transfer to other domains, disciplines, and learning tasks (e.g., language learning, procedural training, or collaborative problem solving). Future work should validate these gaze-adaptive support techniques across a broader set of contents and instructional approaches. Furthermore, participants were recruited from the KD2Lab pool. This sampling strategy may not fully represent the diversity of learners. For example, target user groups also include students from different institutions and age groups. It also remains unclear whether students with varying levels of prior knowledge would use the systems similarly or appropriate the tools to suit their specific needs. Future research should replicate the studies with more diverse learner populations and in different institutional contexts. Regarding the immersive technology, while the dissertation covers common head-mounted MR and VR setups, other technologies used for immersive learning were outside the scope, including handheld AR on smartphones or tablets and projection-based MR environments. This limits the generalizability of the design implications. Future research could systematically explore additional configurations using the cross-device taxonomy of Brudy et al. (2019) and the Reality-Virtuality Continuum (Milgram and Fumio, 1994) to validate whether the proposed gaze-adaptive support generalizes across other devices and degrees of immersion.

Third, from a methodological perspective, this dissertation follows a learner-centered approach and incorporates participatory design workshops. However, participatory design is a broad paradigm with many complementary methods. In this work, I primarily relied on the 6-3-5 brainwriting technique (VanGundy, 1984), which does not fully explore the broader space of participatory methods or clarify when particular techniques are most appropriate across stages of immersive learning system design. Future research should investigate additional participatory methods (e.g., bodystorming, enactments in VR, and longitudinal co-design) to better understand how to conduct effective participatory research for immersive learning. This can provide prescriptive knowledge on the integration of participatory methods into iterative development cycles of immersive systems. Another methodological limitation is related to the learner-centered perspective throughout the dissertation. while the dissertation emphasizes gaze-adaptive support for learners, several of the proposed gaze-adaptive techniques could also be leveraged from an instructor-centered perspective, for example, to monitor learner attention, diagnose learning difficulties, and provide interventions. Prior work has explored educator-oriented support in immersive learning, such as authoring tools and learning analytics (Hubenschmid et al., 2022; Shen et al., 2025), but gaze-adaptive systems are only beginning to be researched in this context (Thanyadit et al., 2023). Future research can further investigate gaze-adaptive support for educators. This can contribute to a more comprehensive understanding of gaze-adaptive support that benefits both learners and educators.

7 Conclusion

This dissertation investigates how to design gaze-adaptive support for immersive learning systems, with the goal of improving learners' attention management and learning outcomes. It contributes design knowledge and novel artifacts for developing learner-centered immersive learning systems. Particularly, it extends the scope of gaze-adaptive learning support beyond the immersive session itself to the ex-situ learning process, aligning with perspectives from ubiquitous and seamless learning frameworks.

Immersive technologies, including VR and MR, have been leveraged to create innovative digital learning experiences known as immersive learning. They can benefit learners by fostering a strong sense of presence and enabling active exploration in situated learning environments. However, the interactive elements and technological novelty that brings benefits to learners can also increase cognitive load and make attention allocation challenging. These challenges have not been addressed as current research on immersive learning prioritizes technical capability and content delivery over learner-centered pedagogical support. As a result, learners may struggle to self-regulate their attention without guidance and do not necessarily achieve better learning outcomes when using immersive systems. Against this background, this dissertation identifies an opportunity to leverage eye-tracking and gaze data, now available on many VR and MR headsets, to actively support learners during immersive learning.

I addressed three challenges. First, there is a need to understand the current design landscape and user challenges of immersive learning. One significant user challenge is attention management, as educational research highlights focused attention as a critical cognitive prerequisite for effective learning. Meanwhile, in-situ attention and cognitive support in immersive learning is crucial for improving learning outcomes. Second, while gaze-adaptive support has been studied in non-immersive contexts, transferring this design knowledge to immersive learning environments remains challenging. Third, immersive learning sessions are often isolated: learning support typically ends when learners remove the headset. To address this disconnect between the immersive experience and the ex-situ learning process, this dissertation introduces cross-device interaction techniques that extend support beyond the headset.

To tackle these challenges, my dissertation consists of four studies. Study I, a systematic literature review of 80 research articles, contributes a conceptual framework and design space for MR learning. It identifies limited pedagogical grounding in existing systems, a lack of adaptive learning support, and a tendency to evaluate primarily technical aspects rather than cognitive aspects of immersive learning. Study II introduces *AF-Mix*, a gaze-aware MR system that provides attention feedback during learning. The results show that visualizing attention helps learners reflect on their attention allocation and retrieve missed information during review. Study III presents *GazeNote*, a cross-device note-taking architecture that links immersive experiences with a mobile device. It demonstrates that cross-device interaction can help learners capture contextual information without breaking immersion, supporting the transition to ex-situ review. Study IV introduces *AttentiveLearn*, a learning ecosystem that uses eye-tracking data from VR lectures to generate personalized quizzes on a companion mobile assistant. Evaluated in a four-week field study, *AttentiveLearn* increased learners' motivation and engagement in a realistic setting and showed promising evidence toward improved learning outcomes.

Together, these studies contribute to HCI research by providing a foundation for future cross-device learning ecosystems that connect immersive experiences with everyday mobile devices. They also offer concrete artifacts and design knowledge for implementing gaze-adaptive support, supporting future research toward more learner-centered immersive learning experiences. In conclusion, the conceptual framework, design knowledge, methodological insights, and artifacts presented in this dissertation demonstrate how support can be integrated across devices and across phases of the learning journey. Taking a learner-centered perspective, this dissertation aims to advance immersive learning research with gaze-adaptive support.

Appendix

A. Appendix for Study 1

Table 7.1: Types of HMDs for MR learning systems

Technology	Explanation
Video See-through	The HMD blends the video stream of the surroundings captured by the device with the virtual elements, the blended content is then displayed on a screen in front of users' eyes.
Optical See-through	The HMD has a transparent display that reflects the generated virtual elements in front of users' eyes. As a result, users can perceive the physical surroundings with the virtual overlay through the transparent display.

Table 7.2: Learning paradigms and theories

Learning Paradigms	Explanation
Behaviorism	Behaviorism suggests that learning happens when behaviors are rewarded or punished. It emphasizes conditioning and reinforcement, focusing on the external manifestations of learning, not internal mental processes.
Cognitivism	Cognitivism suggests that learning is driven by internal mental processes like thinking, memory, and problem-solving, highlighting the importance of understanding the information processing and cognitive process of learning.
Constructivism	Constructivism proposes that learning is an active process where individuals build their understanding of the world through experiences and reflect on those experiences, emphasizing the role of personal meaning-making and active interaction in learning.

Table 7.3: Design features of MR learning systems

Design Features	Explanation
Passive observation	The system mainly allows users to observe virtual environments or educational content without active participation or interaction.
Basic Interaction	Users can interact with virtual elements including UI elements and 3D objects in basic ways, including moving, rotating, and scaling the virtual objects.
High interactivity	Users can interact with the virtual objects in a more engaging way, including creating new objects by assembling given objects
Social Interaction	Collaborative learning experiences that allow users to engage with peers or instructors in shared virtual spaces, encouraging communication and knowledge exchange during the learning experience
Contextual Instruction	Guidance and instructions during the use of the MR system to help students better understand and retain information
Feedback	The system gives feedback information on users' input and performance to help users reflect on their own learning strategies
Embodied movement	The system enables users to use their own body movements to interact with the learning content and make progress in learning.
Rewards and achievement	Gamification elements that motivate users by presenting virtual rewards and acknowledgment of their learning progress

Table 7.4: Research methods

Research Methods	Explanation
Qualitative Research	Research that evaluates subjective experiences and interactions. It emphasizes collecting and analyzing non-numerical data, such as interviews, observations, and textual analysis, to uncover underlying meanings, patterns, and insights (Edmonds and Kennedy, 2016).
Quantitative Research	Research that involves collecting and analyzing numerical data to deliver quantified results and trends. It emphasizes objective measurement, statistical analysis, and controlled experimentation to draw conclusions.
Mixed-methods	Mixed-methods research combines elements of both qualitative and quantitative research methods to complement each other and allows researchers to conclude more comprehensively (Creswell, 2009).

Table 7.5: Summary of identified literature reviews on XR learning systems

Authors	Keywords	Methodology	Databases	# Reviewed articles	Education Level
Won et al. (2023)	immersive VR, education, HMD, participant	not mentioned	Scopus, ProQuest, WoS, Google Scholar	219	All
Radianti et al. (2020)	VR, education, higher education	Webster and Watson (2002), Kitchenham and Charters (2007)	IEEE Xplore, ProQuest, Scopus, WoS	38	Higher Education
Ibáñez and Delgado-Kloos (2018)	AR, education, STEM	not mentioned	ACM DL, ERIC, IEEE Xplore, WoS, ScienceDirect, Scopus, Springer	28	All
Jensen and Konradsen (2018)	VR, HMD, education, training	not mentioned	Scopus, WoS, EBSCOHost, PubMed, IEEE Xplore, ERIC, PsycINFO, IBSS	21	All
Mystakidis et al. (2022)	AR, STEM, higher education	Petersen et al. (2015)	Scopus, Springer, WoS, ERIC, EBSCO, IEEE Xplore	45	Higher Education
Hidayat and Wardat (2023)	AR, STEM, education	Page et al. (2021)	ERIC, ScienceDirect, Scopus	42	All
Yu et al. (2022)	AR, game-based learning	Kitchenham and Charters (2007)	Google Scholar, ISI Web of Science, ProQuest, ProQuest Dissertation, PubMed, Engineering Village, IEEE Xplore	46	All

Table 7.6: Data collection methods

Data Collection	Explanation
Interview	A qualitative data collection method where researchers engage in conversation with participants to gather information about their experiences.
Survey	Collecting data by asking participants questions about their opinions, behaviors, or demographic characteristics.
Knowledge Test	An assessment of participants' understanding and learning progress of a given topic with multiple questions.
Interaction Log	Log data of how participants interact with the presented MR learning system such as the time of each interaction, the user's actions or inputs, system responses, errors encountered, and any other relevant information related to the user-system interaction.
Biosignal	Physiological signals or measurements generated by the human body, collected using biosignal sensors during the use of the MR system, including heart rate data, eye-tracking data, etc.

Table 7.7: Inclusion and exclusion criteria of the literature review

Inclusion Criteria	Exclusion Criteria
Peer-reviewed and published in journals or conferences	Non-peer-reviewed
With an empirical user evaluation	Pure conceptual papers
Full Papers	Short papers or work-in-progress papers
Published in 2013 - 2025	Not applied in higher education context

Table 7.8: Participant Demographics for Design Workshop

Participant ID	Major	Gender	Degree	Prior XR Experience	XR Experience Details (if Yes)
P1	Industrial Engineering	Male	Bachelor	Yes	VR Games
P2	Information Systems	Male	Bachelor	No	
P3	Computer Science	Female	Bachelor	Yes	Developed AR Mobile Applications
P4	Industrial Engineering	Male	Bachelor	No	
P5	Information Systems	Female	Bachelor	No	
P6	Information Systems	Male	Bachelor	Yes	Assistant in VR Lab

Table 7.9: Participant Demographics for System Evaluation

Participant ID	Major	Gender	Degree	Prior XR Experience	XR Experience Details (if Yes)
P1	Industrial Engineering	Male	Bachelor	Yes	VR games
P2	Mechanical Engineering	Male	Master	No	
P3	Biochemistry	Female	Bachelor	No	
P4	Computer Science	Male	Master	Yes	VR app development
P5	Electrical Engineering	Male	Bachelor	No	
P6	Information Systems	Female	Bachelor	No	
P7	Industrial Engineering	Female	Master	Yes	VR training simulation
P8	Mechanical Engineering	Male	Bachelor	No	
P9	Chemistry	Male	Bachelor	No	
P10	Information Systems	Female	Master	Yes	Museum VR exhibit
P11	Industrial Engineering	Male	Bachelor	No	
P12	Biochemistry	Female	Bachelor	Yes	HoloLens medical training app (HoloPatient)
P13	Computer Science	Female	Master	No	
P14	Electrical Engineering	Male	Bachelor	No	
P15	Biology	Male	Bachelor	Yes	AR mobile games
P16	Mechanical Engineering	Female	Master	No	
P17	Chemistry	Female	Bachelor	Yes	VR experience at a science center
P18	Information Systems	Male	Master	No	

C. Appendix for Study 3

C.1. Participant Information

C.2. Evaluation Protocol and Interview Guideline

Pre-Questionnaire (5 minutes, LimeSurvey)

1. How important is note-taking in your regular learning routines?
2. How good is your memory and ability to pay attention in learning?
3. Have you used XR technology before?
   (a) Which specific type of XR technology (AR, VR, MR, etc.)?
   (b) Have you used an XR learning system?

Note-Taking Phase (around 10 minutes per scenario with HoloLens, no time limit)

- **Scenario 1/2/3 (Counterbalanced order, paired with a random learning module)**
  - NASA-TLX questionnaire (0-100 scale, 5 points step, according to the official instruction)
  - Short UEQ questionnaire (7-point-Likert)
- **Scenario 1/2/3 (Counterbalanced order, paired with a random learning module)**
  - NASA-TLX questionnaire
  - Short UEQ questionnaire
- **Scenario 1/2/3 (Counterbalanced order, paired with a random learning module)**
  - NASA-TLX questionnaire
  - Short UEQ questionnaire
- *Note for Experimenter: Ensure proper system setup and calibration before each scenario (IP address configuration, eye-tracking calibration). Remind participants to save notes.*

Interruption Task and Rest Period (15-30 minutes)

- Lego Serious Play task (record task completion time).
- Rest period to control for time between tasks.

Note Review Phase (5-10 minutes, no time limit, on iPad)

- Review notes taken in all three scenarios on the iPad.
- Subjective rating of notes.

Semi-Structured Interview (15-20 minutes)

1. **Note-taking Experience**
   (a) What are your impressions of each note-taking method you tried?
      i. Why do you think one method was better than others? Which one was better?
   (b) What did you like and dislike when using each note-taking approach?
2. **Effectiveness of Notes**
   (a) Which method was most productive and effective in your opinion, after completing the quiz?
   (b) (Optional follow-up) Considering a one-week retention period, which method do you think would be most effective for long-term recall?
3. **Gaze Support (Scenario 3)**
   (a) Have you had experience with eye-tracking before?
   (b) The heatmap visualized your attention distribution. What are your thoughts on this feature?
   (c) How helpful was the heatmap for your note-taking and review process?
      i. Explain why in relation to (1) note-taking and (2) reviewing notes.
      ii. If not helpful, could you elaborate on why it was not helpful compared to other methods?
      iii. If helpful, could you describe how you would prefer to receive such support in the future?
4. **General Learning and Note-Taking**
   (a) How do you typically use note-taking techniques in your regular learning (e.g., exams, lectures)?
      i. What is your primary purpose for taking notes?
      ii. For this purpose, what requirements do you have for note-taking tools?
   (b) Compared to your regular note-taking methods, how did you find the methods provided in this study? Did they meet your requirements?

(c) Are there unique aspects of learning in MR that require you to adjust your note-taking techniques?

i. If so, what are those aspects?

ii. If you prefer not to adjust your techniques, how would you like the system to adapt to your needs?

(d) How would you suggest improving the system? What features would enhance your note-taking strategy within MR learning?

C.3. Descriptive Information of the System Evaluation

Table 7.10: Descriptive Statistics of NASA-TLX

Note-taking	Dimension	N	Mean	SD	SE	CV
S1	Mental	18	48.611	20.991	4.948	0.432
	Physical	18	35.556	27.434	6.466	0.772
	Temporal	18	36.389	22.932	5.405	0.630
	Performance	18	34.167	22.179	5.228	0.649
	Effort	18	49.167	22.961	5.412	0.467
	Frustration	18	33.889	23.424	5.521	0.691
S2	Mental	18	46.111	23.424	5.521	0.508
	Physical	18	27.778	19.037	4.487	0.685
	Temporal	18	30.278	19.664	4.635	0.649
	Performance	18	39.167	15.554	3.666	0.397
	Effort	18	41.389	20.422	4.814	0.493
	Frustration	18	31.389	23.314	5.495	0.743
S3	Mental	18	57.222	24.925	5.875	0.436
	Physical	18	33.611	24.242	5.714	0.721
	Temporal	18	36.667	26.066	6.144	0.711
	Performance	18	52.833	15.990	3.769	0.303
	Effort	18	50.556	26.728	6.300	0.529
	Frustration	18	38.333	24.193	5.702	0.631

Table 7.11: Descriptive Statistics of Behavioral Data

	Time (s)			**Words**			**Gaze Shifts**			**Nr. Notes**		
	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3
Mean	362.28	239.89	305.22	66.00	33.89	38.22	13.33	4.94	4.89	4.61	3.94	4.00
SD	114.47	86.61	73.01	34.23	24.60	36.63	9.15	2.26	2.06	2.03	1.98	1.88
SE	26.98	20.41	17.21	8.07	5.80	8.63	2.16	0.53	0.48	0.48	0.47	0.44
CV	0.316	0.361	0.239	0.519	0.726	0.958	0.686	0.457	0.420	0.441	0.503	0.470
Minimum	181.00	103.00	187.00	5.00	0.00	0.00	5.00	2.00	2.00	2.00	0.00	2.00
Maximum	603.00	483.00	455.00	116.00	91.00	104.00	35.00	9.00	8.00	9.00	8.00	8.00

Table 7.12: Descriptive Statistics of UEQ Scales Across Scenarios

	S1				S2				S3			
UEQ Scale	M	SD	SE	CV	M	SD	SE	CV	M	SD	SE	CV
Obstructive ∣ Supportive	4.89	2.00	0.47	0.408	5.56	1.72	0.41	0.310	4.94	1.86	0.44	0.377
Complicated ∣ Easy	4.56	1.69	0.40	0.371	5.33	1.72	0.40	0.322	4.61	1.98	0.47	0.428
Confusing ∣ Clear	5.11	1.57	0.37	0.307	5.39	1.69	0.46	0.375	4.78	2.02	0.48	0.422
Boring ∣ Exciting	5.56	1.15	0.27	0.207	5.50	1.25	0.40	0.313	4.50	2.20	0.52	0.490
Not Interesting ∣ Interesting	5.94	1.11	0.26	0.187	5.56	1.29	0.29	0.227	4.83	1.72	0.41	0.357
Conventional ∣ Inventive	5.94	1.35	0.32	0.227	5.67	1.57	0.31	0.233	5.50	1.62	0.38	0.294
Usual ∣ Leading Edge	5.78	1.31	0.31	0.226	5.67	1.50	0.37	0.277	5.83	1.65	0.39	0.284
Inefficient ∣ Efficient	5.17	1.43	0.34	0.276	5.22	1.96	0.35	0.264	5.72	1.71	0.40	0.299

D. Appendix for Study 4

D.1. Survey on *LectureQuiz*

Table 7.13: Post-lecture survey on *LectureQuiz* (5-point Likert scale).

Q1	How accurately did *LectureQuiz* reflect the parts of the lecture where your attention was lower?
Q2	To what extent did *LectureQuiz* help you understand and review the lecture content?

D.2. Prompts Used in the Mobile Assistant

This appendix documents the prompt templates used in the mobile assistant. The assistant name is anonymized as [Name of Assistant].

Initial Message

Purpose: Greet the learner after a VR lecture and invite questions.

Instruction to the assistant: You are [Name of Assistant], an adaptive tutor assisting a student who recently completed a VR lecture. You assist the student in understanding the content of the lecture.

Generate a welcome message to the user in which you offer to the user to ask questions about the lecture content and ask if they have any. Explain that you are a tutor that has knowledge about these topics:

- Topic A
- Topic B
- Topic C
- ...

Example phrases:

- “Welcome! You can ask me questions about today’s lecture. We covered topics such as … Do you have any questions?”
- “Hello! You can ask me anything about the lecture. Do you have a question?”

Summarizer for *OpenChat*

Purpose: Keep a short running summary of the dialogue for context.

Instruction to the assistant: Too keep the conversation focused and the prompt short, you have to summarize the conversation so far.

For the summary focus on the topics that are covered, questions that were asked and the knowledge of the user. Summarize whether or which question the user has asked.

For your context info, this is a summary of the conversation so far. The summary you generate will be appended to the conversation and will be used for the next steps. {{summary}}

Examples:

- “The user had no question.”
- “The user asked about Topic X.”
- “The user showed prior knowledge of Topic Y.”
- “The user was confused about Topic Z.”

Q&A Instruction for *OpenChat*

Purpose: Provide answers grounded in lecture materials and prior conversation.

Instruction to the assistant:

You are [Name of Assistant], an adaptive tutor assisting a student who recently completed a VR lecture, and you simulate a human tutor for the student. You assist the student in understanding the content of the lecture.

Answer the user's question based on the content of the lecture and the conversation so far.

This is basic knowledge about the content of the lecture:

- Topic A with content description
- Topic B with content description
- Topic C with content description
- ...

Additionally, you have the following knowledge based on the conversation's context: {{chat_context}}

This is the source of the context information: {{chat_context_source}}

Only answer if the question relates to the content of the lecture. Use only this knowledge to answer the question. When you answer, you must include links to the sources of the content you used. Always invite the learner to ask another question.

You have no knowledge about the relevance of topics for the exam. Do **not** make any statements about the relevance of topics for the exam.

For your context information, this is a summary of the conversation so far: {{summary}}

Additionally, you also have the latest two messages directly, which are already part of the summary: {{message_history}}

Confusion Update for *OpenChat*

Purpose: Maintain per-topic confusion scores for each lecture topic in the interval $[0, 1]$ with a float field, updated consistently throughout the *OpenChat*.

Instruction to the assistant:

You are a precise analysis and decision system. Your name is [Name of Assistant] and you are a tutor at a university.

You have to decide based on the conversation whether the user is having confusion about a topic of the lecture or not.

If the user asked a question about a topic, it can be a sign of confusion if they do not understand something, or even the opposite: the user has already understood it and just wants to explore more details. Consider the linguistic analysis approach of the existing research in this process of confusion update. (The paper of Atapattu et al. (2020) is attached here in the prompt)

For each topic given, you have to update the confusion level, which is a float between 0 and 1.

- 0 means no confusion,
- 1 means high confusion.

Currently, the confusion levels are as follows:

- Topic A: {{confusion_topic_A}}
- Topic B: {{confusion_topic_B}}
- Topic C: {{confusion_topic_C}}

When updating, also consider the old value. Changes should always be small (e.g., 0.1 or 0.2) to avoid overreacting to a single question.

For your context information, this is a summary of the conversation which was already used for the calculation of the current confusion levels. This summary does **not** include the latest two messages, which are the **main** messages you have to consider for your decision: {{summary}}

The next two messages are the latest two messages of the conversation, which you have to consider for your decision. If there is no message relevant for the confusion level, just keep the levels as they are: {{message_history}}

Bibliography

Abdrabou, Y., S. Özdel, V. Maquiling, E. Bozkir, and E. Kasneci (2025). "From Gaze to Data: Privacy and Societal Challenges of Using Eye-tracking Data to Inform GenAI Models". *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*. ETRA '25. New York, NY, USA.

Abeysinghe, Y., K. Cauchi, V. Ashok, and S. Jayarathna (2025). "Framework for Measuring Visual Attention in Gaze-Driven VR Learning Environments Using Meta Quest Pro". *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*. Tokyo Japan.

Akçayır, M. and G. Akçayır (2017a). "Advantages and Challenges Associated with Augmented Reality for Education: A Systematic Review of the Literature". *Educational Research Review*.

— (2017b). "Advantages and challenges associated with augmented reality for education: A systematic review of the literature". *Educational Research Review*.

Alazmi, H. S. and G. M. Alemtairy (2024). "The effects of immersive virtual reality field trips upon student academic achievement, cognitive load, and multimodal presence in a social studies educational context". *Education and Information Technologies*.

Albeedan, M., H. Kolivand, R. Hammady, and T. Saba (2024). "Seamless crime scene reconstruction in mixed reality for investigation training: a design and evaluation study". *IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES*.

Alenezi, M. (2023). "Digital Learning and Digital Institution in Higher Education". *Education Sciences*.

Ali, A. A., G. A. Dafoulas, and J. C. Augusto (2019). "Collaborative Educational Environments Incorporating Mixed Reality Technologies: A Systematic Mapping Study". *IEEE Transactions on Learning Technologies*.

Alzahrani, N. M. (2020). "Augmented reality: A systematic review of its benefits and challenges in E-learning contexts". *Applied Sciences*.

An, P., K. Holstein, B. d'Anjou, B. Eggen, and S. Bakker (2020). "The TA Framework: Designing Real-time Teaching Augmentation for K-12 Classrooms". *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. New York, NY, USA.

Arici, F., P. Yildirim, Ş. Caliklar, and R. M. Yilmaz (2019). "Research Trends in the Use of Augmented Reality in Science Education: Content and Bibliometric Mapping Analysis." *Computers & Education*.

Asaumi, K., M. Oki, W. Ohashi, H. Sato, and K. Tanaka (2025). "Effect of augmented reality-based endotracheal suctioning skill training of undergraduate nursing students: An open-label randomized controlled trial". *CLINICAL SIMULATION IN NURSING*.

Atapattu, T., K. Falkner, M. Thilakaratne, L. Sivaneasharajah, and R. Jayashanka (2019). *An Identification of Learners' Confusion through Language and Discourse Analysis*.

— (2020). "What Do Linguistic Expressions Tell Us about Learners' Confusion? A Domain-Independent Analysis in MOOCs". *IEEE Transactions on Learning Technologies*.

Auda, J., U. Gruenefeld, S. Faltaous, S. Mayer, and S. Schneegass (2023). "A Scoping Survey on Cross-reality Systems". *ACM Comput. Surv.*

Azizo, A. S. binti, F. bin Mohamed, C. V. Siang, M. I. M. bin Isham, M. K. bin Mokhtar, and I. Wirasari (2022). "A prototype application of StickAR to enhance note-taking activity by using augmented reality technology". *Intelligent technologies for interactive entertainment*. Ed. by Z. Lv and H. Song. Cham.

Baceviciute, S., A. Mottelson, T. Terkildsen, and G. Makransky (2020). "Investigating Representation of Text and Audio in Educational VR using Learning Outcomes and EEG". *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. New York, NY, USA.

Banky, G. and A. Blicblau (2017). "Doing it with a glass onion: Investigating affordances provided by a head-mounted augmented reality immersive device for the real-time online supervision of experimental learning". *2017 IEEE 3rd International Conference on Engineering Technologies and Social Sciences, ICETSS 2017*.

Banquiero, M., G. Valdeolivas, D. Ramon, and M. C. Juan (2024). "A color Passthrough mixed reality application for learning piano". *VIRTUAL REALITY*.

Barteit, S., L. Lanfermann, T. Bärnighausen, F. Neuhann, and C. Beiersmann (2021). "Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review". *JMIR Serious Games*.

Benight, C. C. and A. Bandura (2004). "Social cognitive theory of posttraumatic recovery: the role of perceived self-efficacy". *Behaviour Research and Therapy*.

Bhatia, N. and O. K. Matar (2022). "Learning and teaching fluid dynamics using augmented and mixed reality". *2022 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct)*.

Bianchi, I., A. L. Zanatta, and R. Rieder (2020). "Augmented Reality in Medical Teaching-Learning Process Content: A Systematic Review". *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*. Porto de Galinhas, Brazil.

Blume, F., R. Göllner, K. Moeller, T. Dresler, A.-C. Ehlis, and C. Gawrilow (2019). "Do students learn better when seated close to the teacher? A virtual classroom study considering individual levels of inattention and hyperactivity-impulsivity". *Learning and Instruction*.

Bognar, B., V. Gajger, and V. Ivic (2015). *Constructivist E-Learning in Higher Education*. Tech. rep.

Boy, G. A. (1997). "The group elicitation method for participatory design and usability testing". *Interactions*.

Braun, V. and V. Clarke (2006a). "Using thematic analysis in psychology". *Qualitative Research in Psychology*.

— (2006b). "Using thematic analysis in psychology". *Qualitative Research in Psychology*.

— (2012). "Thematic analysis". *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. APA handbooks in psychology®. Washington, DC, US.

— (2019). "Reflecting on reflexive thematic analysis". *Qualitative Research in Sport, Exercise and Health*.

— (2021). "One Size Fits All? What Counts as Quality Practice in (Reflexive) Thematic Analysis?" *Qualitative Research in Psychology*.

Brown, J. S., A. Collins, and P. Duguid (1989). "Situated Cognition and the Culture of Learning". *Educational Researcher*.

Brudy, F., C. Holz, R. Rädle, C.-J. Wu, S. Houben, C. N. Klokmose, and N. Marquardt (2019). "Cross-Device Taxonomy: Survey, Opportunities and Challenges of Interactions Spanning Across Multiple Devices". *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA.

Bruner, J. S. (1960). *The process of education*. The process of education. Oxford, England.

Brysbaert, M. (2019). "How Many Words Do We Read per Minute? A Review and Meta-Analysis of Reading Rate". *Journal of Memory and Language*.

Buchner, J., K. Buntins, and M. Kerres (2022). “The impact of augmented reality on cognitive load and performance: A systematic review”. *Journal of Computer Assisted Learning*.

Bulling, A. (2016). “Pervasive Attentive User Interfaces”. *Computer*.

Buscher, G., A. Dengel, R. Biedert, and L. V. Elst (2012). “Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond”. *ACM Transactions on Interactive Intelligent Systems*.

Butler, D. L. and P. H. Winne (1995). “Feedback and Self-Regulated Learning: A Theoretical Synthesis”. *Review of Educational Research*.

Caine, K. (2016). “Local Standards for Sample Size at CHI”. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. San Jose California USA.

Cairns, P. and A. L. Cox, eds. (2008). *Research Methods for Human-Computer Interaction*. Cambridge.

Campo, A., A. Michalko, B. Van Kerrebroeck, B. Stajic, M. Pokric, and M. Leman (2023). “The assessment of presence and performance in an AR environment for motor imitation learning: A case-study on violinists”. *COMPUTERS IN HUMAN BEHAVIOR*.

Cao, Y., Y. Tang, and Y. Xie (2018). “A Novel Augmented Reality Guidance System for Future Informatization Experimental Teaching”. *Proceedings of 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2018*.

Card, S. K. (1981). “The Model Human Processor: A Model for Making Engineering Calculations of Human Performance”. *Proceedings of the Human Factors Society Annual Meeting*.

Chandran, K., P. Dimitriou, L. Wand, and M. McGinity (2024). “Co-located mixed reality classrooms: Four case studies from higher education”.

Charfi, S., E. Dubois, and D. L. Scapin (2009). “Usability recommendations in the design of mixed interactive systems”. *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*. EICS ’09. New York, NY, USA.

Chen, P.-H., H.-W. Ho, H.-C. Chen, K.-W. Tam, J.-C. Liu, and L.-F. Lin (2024). “Virtual reality experiential learning improved undergraduate students’ knowledge and evaluation skills relating to assistive technology for older adults and individuals with disabilities”. *BMC Medical Education*.

Cheng, K.-H. and C.-C. Tsai (2019). “A case study of immersive virtual field trips in an elementary classroom: Students’ learning experience and teacher-student interaction behaviors”. *Computers & Education*.

Cheng, S., Z. Sun, L. Sun, K. Yee, and A. K. Dey (2015). “Gaze-Based Annotations for Reading Comprehension”. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. New York, NY, USA.

Chernikova, O., N. Heitzmann, M. Stadler, D. Holzberger, T. Seidel, and F. Fischer (2020). “Simulation-Based Learning in Higher Education: A Meta-Analysis”. *Review of Educational Research*.

Chris Dede (2009). “Immersive interfaces for engagement and learning”. *Science (New York, N.Y.)*

Christopoulos, A., M. Styliou, N. Ntalas, and C. Stylios (2024). “The Impact of Immersive Virtual Reality on Knowledge Acquisition and Adolescent Perceptions in Cultural Education”. *Information*.

Chun, M. and J. Wolfe (2008). “Visual Attention”. *Blackwell’s Handbook of Perception*.

Connolly, M., G. Iohom, N. O’Brien, J. Volz, A. O’Muircheartaigh, P. Serchan, A. Biculescu, K. G. Gadre, C. Soare, L. Griseto, and G. Shorten (2024). “Delivering clinical tutorials to medical students using the Microsoft HoloLens 2: A mixed-methods evaluation”. *BMC MEDICAL EDUCATION*.

Contrino, M. F., M. Reyes-Millán, P. Vázquez-Villegas, and J. Membrillo-Hernández (2024). “Using an adaptive learning tool to improve student performance and satisfaction in online and face-to-face education for a more personalized approach”. *Smart Learning Environments*.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches, 3rd ed.* Research design: Qualitative, quantitative, and mixed methods approaches, 3rd ed. Thousand Oaks, CA, US.

D’mello, S. and A. Graesser (2012). “AutoTutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back”. *ACM Transactions on Interactive Intelligent Systems*.

D’Mello, S., A. Olney, C. Williams, and P. Hays (2012). “Gaze Tutor: A Gaze-Reactive Intelligent Tutoring System”. *International Journal of Human-Computer Studies*.

Dafoulas, G., A. Tsiakara, C. Maia, and D. Neilson (2019). “Evaluating the use of augmented reality in learning portfolios for different team roles”. *2019 10th International Conference on Information and Communication Systems, ICICS 2019*.

Dalton, J. and A. Ruhl (2021). *The Effectiveness of Virtual Reality Soft Skills Training in the Enterprise: A Study*. Tech. rep. PwC.

De Guzman, J. A., K. Thilakarathna, and A. Seneviratne (2019). "Security and Privacy Approaches in Mixed Reality: A Literature Survey". *ACM Computing Surveys*.

Dede, C. (2011). "Emerging Technologies, Ubiquitous Learning, and Educational Transformation". Ed. by C. D. Kloos, D. Gillet, R. M. Crespo García, F. Wild, and M. Wolpers. Berlin, Heidelberg.

Demmans Epp, C., B. K. Daniel, and K. Muldner (2023). "Editorial: Learning analytics for supporting individualization: data-informed adaptation of learning". *Frontiers in Education*.

Deng, Y., Y. Zhang, and R. Zhao (2025). "Investigating learning behaviors in desktop-based simulated and vr headset-based immersive 3D learning environments: a cross-media comparative study". *Education and Information Technologies*.

Dengel, A. and J. Mägdefrau (2018). "Immersive Learning Explored: Subjective and Objective Factors Influencing Learning Outcomes in Immersive Educational Virtual Environments". *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*.

Denoue, L., P. Chiu, and T. Fuse (2003). "Shared freeform input for note taking across devices". *CHI '03 extended abstracts on human factors in computing systems*. CHI EA '03. New York, NY, USA.

Di Gioia, F. R., E. Brasier, E. Pietriga, and C. Appert (2022). "Investigating the use of AR glasses for content annotation on mobile devices". *Proc. ACM Hum.-Comput. Interact.*

Diller, F., N. Henkel, G. Scheuermann, and A. Wiebel (2025). "SkillAR:omnipresent in-situ feedback for motor skill training using AR". *VIRTUAL REALITY*.

Doorn, J. van, D. van den Bergh, U. Böhm, F. Dablander, K. Derks, T. Draws, A. Etz, N. J. Evans, Q. F. Gronau, J. M. Haaf, M. Hinne, Š. Kucharský, A. Ly, M. Marsman, D. Matzke, A. R. K. N. Gupta, A. Sarafoglou, A. Stefan, J. G. Voelkel, and E.-J. Wagenmakers (2020). "The JASP guidelines for conducting and reporting a Bayesian analysis". *Psychonomic Bulletin & Review*.

— (2021). "The JASP guidelines for conducting and reporting a Bayesian analysis". *Psychonomic Bulletin & Review*.

Doswell, J. T. and A. Skinner (2014). "Augmenting Human Cognition with Adaptive Augmented Reality". *Foundations of Augmented Cognition. Advancing Human Performance* *and Decision-Making through Adaptive Systems*. Ed. by D. D. Schmorrow and C. M. Fidopiastis. Cham.

Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice*.

Dunlosky, J., K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham (2013). "Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology". *Psychological Science in the Public Interest*.

Edmonds, W. A. and T. D. Kennedy (2016). *An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods*.

Elor, A., S. Whittaker, S. Kurniawan, and S. Michael (2022). "BioLumin: An Immersive Mixed Reality Experience for Interactive Microscopic Visualization and Biomedical Research Annotation". *ACM Transactions on Computing for Healthcare*.

Essmiller, K., T. I. Asino, A. Ibukun, F. Alvarado-Albertorio, S. Chaivisit, T. Do, and Y. Kim (2020). "Exploring Mixed Reality Based on Self-Efficacy and Motivation of Users". *Research in Learning Technology*.

Evans, C., E. Cools, and Z. Charlesworth (2010). "Learning in higher education – how cognitive and learning styles matter". *Teaching in Higher Education*.

Faridan, M., B. Kumari, and R. Suzuki (2023). "ChameleonControl: Teleoperating Real Human Surrogates through Mixed Reality Gestural Guidance for Remote Hands-on Classrooms". *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. New York, NY, USA.

Fernandes, F. A., C. S. C. Rodrigues, E. N. Teixeira, and C. M. L. Werner (2023). "Immersive Learning Frameworks: A Systematic Literature Review". *IEEE Transactions on Learning Technologies*.

Fikar, P., F. Güldenpfennig, and R. Ganhör (2018). "The Use(fulness) of Therapeutic Toys: Practice-derived Design Lenses for Toy Design". *Proceedings of the 2018 Designing Interactive Systems Conference*. DIS '18. New York, NY, USA.

GaliOlga, TüterGözde, E. Beste, DelgadoÓscar, AthertonGray, CrossLiam, and ParesNarcis (2024). "Participatory Design of an Embodied Mixed Reality Experience Aimed to Assessing Individual and Collaborative Behaviors of Children". *Proceedings of the ACM on Human-Computer Interaction*.

Gamage, S. H. P. W., J. R. Ayres, M. B. Behrend, and E. J. Smith (2019). "Optimising Moodle quizzes for online assessments". *International Journal of STEM Education*.

Gao, H., E. Bozkir, L. Hasenbein, J.-U. Hahn, R. Göllner, and E. Kasneci (2021). "Digital Transformations of Classrooms in Virtual Reality". *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. New York, NY, USA.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995). *Bayesian Data Analysis*. New York.

George, O., J. Foster, Z. Xia, and C. Jacobs (2023). "Augmented Reality in Medical Education: A Mixed Methods Feasibility Study." *Curēus*.

Gieser, C., J. Schmitt, E. Loewenstein, C. Weber, V. Braun, and R. Brueck (2024). "SkillsLab plus -a new way to teach practical medical skills in an augmented reality application with haptic feedback". *IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES*.

Gießer, C., M. Pour, J. Schmitt, and R. Brück (2023). "Evaluation of MetaHumans as Digital Tutors in Augmented Reality in Medical Training". *Current Directions in Biomedical Engineering*.

GigXR (2025). *HoloPatient by GIGXR: Enhancing Patient Care Training*.

Giraudeau, P., A. Olry, J. S. Roo, S. Fleck, D. Bertolo, R. Vivian, and M. Hachet (2019). "CARDS: A Mixed-Reality System for Collaborative Learning at School". *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. New York, NY, USA.

Goldie, J. G. S. (2016). "Connectivism: A knowledge learning theory for the digital age?" *Medical Teacher*.

Golovchinsky, G., M. N. Price, and B. N. Schilit (1999). "From reading to retrieval: freeform ink annotations as queries". *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*.

Grad, P., A. M. Przeklasa-Bierowiec, K. P. Malinowski, J. Witowski, K. Proniewska, and G. Taton (2023). "Application of HoloLens-based augmented reality and three-dimensional printed anatomical tooth reference models in dental education". *ANATOMICAL SCIENCES EDUCATION*.

Group, H. (2025). *HoloStudy Demo*.

Hailpern, J., R. Vernica, M. Bullock, U. Chatow, J. Fan, G. Koutrika, J. Liu, L. Liu, S. Simske, and S. Wu (2015). "To Print or Not to Print: Hybrid Learning with METIS Learning Platform". *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. EICS '15. New York, NY, USA.

Hamilton, D., J. McKechnie, E. Edgerton, and C. Wilson (2021). “Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design”. *Journal of Computers in Education*.

Han, Y., Y. Miao, J. Lu, M. Guo, and Y. Xiao (2022). “Exploring Intervention Strategies for Distracted Students in VR Classrooms”. *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI EA ’22. New York, NY, USA.

Hanid, M. F. A., M. N. H. M. Said, and N. Yahaya (2020). “Learning Strategies Using Augmented Reality Technology in Education: Meta-Analysis”. *Universal Journal of Educational Research*.

Hart, S. G. (2006). “Nasa-Task Load Index (NASA-TLX); 20 Years Later”. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Hennerley, J., M. Dickinson, and M. Jiang (2017). “Augmented Reality Enhanced Human Robot Interaction for Social Robots as E-Learning Companions”. *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*. HCI ’17. Swindon, GBR.

Heslin, P. A. (2009). “Better than brainstorming? Potential contextual boundary conditions to brainwriting for idea generation in organizations”. *Journal of Occupational and Organizational Psychology*.

Hess, O., J. Qian, J. Bruce, E. Wang, S. Rodriguez, N. Haber, and T. J. Caruso (2022). “Communication Skills Training Using Remote Augmented Reality Medical Simulation: A Feasibility and Acceptability Qualitative Study.” *Medical science educator*.

Heyko, D. and D. R. Flatla (2021). “Identifying the Factors That Influence DHH Employee Success Under Hearing Supervisors”. *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. DIS ’21. New York, NY, USA.

Hidayat, R. and Y. Wardat (2023). “A systematic review of Augmented Reality in Science, Technology, Engineering and Mathematics education”. *Education and Information Technologies*.

Hidi, S. and K. A. Renninger (2006). “The Four-Phase Model of Interest Development”. *Educational Psychologist*.

Horizon, M. (2025). *Building Content to Optimize User Time: 20-40 Minute “Goldilocks” Session Length on Meta Quest*.

Hu, P.-C. and P.-C. Kuo (2017). “Adaptive learning system for E-learning based on EEG brain signals”. *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*.

Huang, G., X. Qian, T. Wang, F. Patel, M. Sreeram, Y. Cao, K. Ramani, and A. J. Quinn (2021). “AdapTutAR: An Adaptive Tutoring System for Machine Tasks in Augmented Reality”. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI ’21. New York, NY, USA.

Huang, Y.-H., H.-Y. Chang, W.-l. Yang, Y.-K. Chiu, T.-C. Yu, P.-H. Tsai, and M. Ouhyoung (2018). “CatAR: A Novel Stereoscopic Augmented Reality Cataract Surgery Training System with Dexterous Instruments Tracking Technology”. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA.

Huang, T.-C. (2019). “Seeing creativity in an augmented experiential learning environment”. *UNIVERSAL ACCESS IN THE INFORMATION SOCIETY*.

Huang, T.-C., C.-C. Chen, and Y.-W. Chou (2016). “Animating Eco-Education: To See, Feel, and Discover in an Augmented Reality-Based Experiential Learning Environment”. *Computers & Education*.

Hubenschmid, S., J. Wieland, D. I. Fink, A. Batch, J. Zagermann, N. Elmqvist, and H. Reiterer (2022). “ReLive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies”. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI ’22. New York, NY, USA.

Hutt, S., K. Krasich, C. Mills, N. Bosch, S. White, J. R. Brockmole, and S. K. D’Mello (2019). “Automated gaze-based mind wandering detection during computerized learning in classrooms”. *User Modeling and User-Adapted Interaction*.

Hutt, S., K. Krasich, J. R. Brockmole, and S. K. D’Mello (2021). “Breaking out of the Lab: Mitigating Mind Wandering with Gaze-Based Attention-Aware Technology in Classrooms”. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI ’21. New York, NY, USA.

Hutt, S., C. Mills, N. Bosch, K. Krasich, J. Brockmole, and S. D’Mello (2017). “"Out of the Fr-Eye-ing Pan": Towards Gaze-Based Models of Attention during Learning with Technology in the Classroom”. *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. Bratislava Slovakia.

Hwang, G.-J. and C.-C. Tsai (2011). “Research trends in mobile and ubiquitous learning: a review of publications in selected journals from 2001 to 2010”. *British Journal of Educational Technology*.

Ibáñez, M.-B. and C. Delgado-Kloos (2018). “Augmented reality for STEM learning: A systematic review”. *Computers & Education*.

Ibáñez, M.-B., Á. Di-Serio, D. Villarán-Molina, and C. Delgado-Kloos (2016). “Support for Augmented Reality Simulation Systems: The Effects of Scaffolding on Learning Outcomes and Behavior Patterns”. *IEEE Transactions on Learning Technologies*.

Iqbal, M. Z., E. Mangina, and A. G. Campbell (2019). “Exploring the Use of Augmented Reality in a Kinesthetic Learning Application Integrated with an Intelligent Virtual Embodied Agent”. *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*.

Jacob, R. J. K. (1995). “Eye tracking in advanced interface design”. *Virtual Environments and Advanced Interface Design*. USA. Chap. 6.

Janeras, M., J. Roca, J. A. Gili, O. Pedraza, G. Magnusson, M. A. Núñez-Andrés, and K. Franklin (2022). “Using Mixed Reality for the Visualization and Dissemination of Complex 3D Models in Geosciences—Application to the Montserrat Massif (Spain)”. *Geosciences*.

Jansen, R. S., D. Lakens, and W. A. IJsselsteijn (2017). “An integrative review of the cognitive costs and benefits of note-taking”. *Educational Research Review*.

Jarodzka, H. and S. Brand-Gruwel (2017). “Tracking the reading eye: towards a model of real-world reading”. *Journal of Computer Assisted Learning*.

Jensen, L. and F. Konradsen (2018). “A review of the use of virtual reality head-mounted displays in education and training”. *Education and Information Technologies*.

John, B., J. Kurian, R. Fitzgerald, and D. Goh (2022). “Students’ Learning Experience in a Mixed Reality Environment: Drivers and Barriers”. *Communications of the Association for Information Systems*.

Johnson, G. and S. Davies (2014). “Self-regulated learning in digital environments: theory, research, praxis”. *British Journal of Research*.

Julia, K., V. R. Peter, and K. Marco (2021). “Educational scalability in MOOCs: Analysing instructional designs to find best practices”. *Computers & Education*.

Just, M. A. and P. A. Carpenter (1980). “A theory of reading: From eye fixations to comprehension”. *Psychological Review*.

Kapp, S., M. Thees, F. Beil, T. Weatherby, J.-P. Burde, T. Wilhelm, and J. Kuhn (2020). "The effects of augmented reality: A comparative study in an undergraduate physics laboratory course". *CSEDU 2020 - Proceedings of the 12th International Conference on Computer Supported Education*.

Keller, A. S., I. Davidesco, and K. D. Tanner (2020). "Attention Matters: How Orchestrating Attention May Relate to Classroom Learning". *CBE—Life Sciences Education*.

Kennel, K. (2022). "Using Eye Tracking Data for Enhancing Adaptive Learning Systems". *2022 Symposium on Eye Tracking Research and Applications*. ETRA '22. New York, NY, USA.

Khan, A. A. (2019). "Gaze assisted voice note taking system". *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*. London United Kingdom.

Khan, A. A., S. Nawaz, J. Newn, R. M. Kelly, J. M. Lodge, J. Bailey, and E. Velloso (2022). "To type or to speak? The effect of input modality on text understanding during note-taking". *Proceedings of the 2022 CHI conference on human factors in computing systems*. CHI '22. New York, NY, USA.

Khan, A. A., J. Newn, R. M. Kelly, N. Srivastava, J. Bailey, and E. Velloso (2021). "GAVIN: Gaze-Assisted Voice-Based Implicit Note-taking". *ACM Trans. Comput.-Hum. Interact.*

Khan, M., F. Trujano, A. Choudhury, and P. Maes (2018). "Mathland: Playful Mathematical Learning in Mixed Reality". *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA '18. Montreal QC Canada.

Khan, T., K. Johnston, and J. Ophoff (2019). "The Impact of an Augmented Reality Application on Learning Motivation of Students". *Advances in Human-Computer Interaction*.

Kiewra, K. A., N. F. DuBois, D. Christian, A. McShane, M. Meyerhoffer, and D. Roskelley (1991). "Note-taking functions and techniques." *Journal of Educational Psychology*.

Kiewra, K. A. and B. M. Frank (1988). "Encoding and external-storage effects of personal lecture notes, skeletal notes, and detailed notes for field-independent and field-dependent learners". *The Journal of Educational Research*.

Kim, J., N. Heo, M. Seo, and H. Shin (2025). "Development and evaluation of an integrative-fidelity nursing simulation education program using extended-reality smart glasses: A quasi-experimental study". *CLINICAL SIMULATION IN NURSING*.

Kim, S. K., H. Yoon, S. Choonsung, C. J. Myung, and Y. Lee (2020). “Design and Implementation of a Smart Glass Application for XR Assisted Training of Core Nursing Skills”. *Journal of Multimedia Information System*.

Kim, T., J. Planey, and R. Lindgren (2022). “The Role of Augmented Reality in Multi-Device Small Group Learning Ecosystems”. *Proceedings of International Conference of the Learning Sciences, ICLS*.

Kirschner, P. A., J. Sweller, and R. E. Clark (2006). “Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching”. *Educational Psychologist*.

Kitchenham, B. and S. Charters (2007). “Guidelines for performing Systematic Literature Reviews in Software Engineering”.

Klaveren, C., S. Vonk, and I. Cornelisz (2017). “The Effect of Adaptive versus Static Practicing on Student Learning - Evidence from a Randomized Field Experiment”. *Economics of Education Review*.

Knierim, P., F. Kiss, M. Rauh, and A. Schmidt (2020). “Tangibility Is Overrated: Comparing Learning Experiences of Physical Setups and Their Virtual Equivalent in Augmented Reality”. *19th International Conference on Mobile and Ubiquitous Multimedia*. MUM 2020. New York, NY, USA.

Knierim, P., T. Kosch, M. Hoppe, and A. Schmidt (2018). “Challenges and Opportunities of Mixed Reality Systems in Education”. *Mensch und Computer 2018 - Workshopband*. Bonn.

Kockord, R. and O. Bodensiek (2021). “Cognitive Load During First Contact With Mixed Reality Learning Environments”. *Mensch Und Computer 2021*. Ingolstadt Germany.

Kristiansen, P. and R. Rasmussen (2014). *Building a better business using the Lego serious play method*.

Kuh, G. D. (1996). “Guiding Principles for Creating Seamless Learning Environments for Undergraduates”. *Journal of College Student Development*.

Kuhail, M. A., A. ElSayary, S. Farooq, and A. Alghamdi (2022). “Exploring Immersive Learning Experiences: A Survey”. *Informatics*.

Kurzhals, K., F. Göbel, K. Angerbauer, M. Sedlmair, and M. Raubal (2020). “A View on the Viewer: Gaze-Adaptive Captions for Videos”. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20. New York, NY, USA.

Lai, M.-L., M.-J. Tsai, F.-Y. Yang, C.-Y. Hsu, T.-C. Liu, S. W.-Y. Lee, M.-H. Lee, G.-L. Chiou, J.-C. Liang, and C.-C. Tsai (2013). “A Review of Using Eye-Tracking Technology in Exploring Learning from 2000 to 2012”. *Educational Research Review*.

Langner, M., P. Toreini, and A. Maedche (2020). “AttentionBoard: A quantified-self dashboard for enhancing attention management with eye-tracking”. *Information systems and neuroscience*. Ed. by F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, and T. Fischer. Cham.

— (2023). “Leveraging Eye Tracking Technology for a Situation-Aware Writing Assistant”. *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*. ETRA ’23. New York, NY, USA.

Lateef, F. (2010). “Simulation-based learning: Just like the real thing”. *Journal of Emergencies, Trauma, and Shock*.

Laumann, D., P. Schlummer, A. Abazi, R. Borkamp, J. Laustroeer, W. Pernice, C. Schuck, R. Schulz-Schaeffer, and S. Heusler (2024). “Analyzing the effective use of augmented reality glasses in university physics laboratory courses for the example topic of optical polarization”. *JOURNAL OF SCIENCE EDUCATION AND TECHNOLOGY*.

Li, K. and B.-M. Wong (2021). “A literature review of augmented reality, virtual reality, and mixed reality in language learning”. *International Journal of Mobile Learning and Organisation*.

Li, Z., M. Annett, K. Hinckley, K. Singh, and D. Wigdor (2019). “HoloDoc: Enabling Mixed Reality Workspaces that Harness Physical and Digital Content”. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI ’19. New York, NY, USA.

Lin, T., R. Singh, Y. Yang, C. Nobre, J. Beyer, M. A. Smith, and H. Pfister (2021). “Towards an Understanding of Situated AR Visualization for Basketball Free-Throw Training”. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI ’21. New York, NY, USA.

Lindlbauer, D., A. M. Feit, and O. Hilliges (2019). “Context-Aware Online Adaptation of Mixed Reality Interfaces”. *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. New Orleans LA USA.

Liu, S., M. Feick, L. Bierhoff, and A. Maedche (2026). “AttentiveLearn: Personalized Post-Lecture Support for Gaze-Aware Immersive Learning”. *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Barcelona, Spain.

Liu, S., A. Mädche, and M. Feick (2025a). "GazeClass: Towards Gaze-Adaptive Cross-Device Learning Support for Virtual Classrooms". *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI EA '25. New York, NY, USA.

Liu, S., T. Schulz, P. Toreini, and A. Maedche (2023). "An Immersive Learning Factory for AI & Data Literacy: An Exploratory Study in the Wild". *Proceedings of Mensch und Computer 2023*. MuC '23. New York, NY, USA.

Liu, S., P. Toreini, and A. Maedche (2022). "Designing Gaze-Aware Attention Feedback for Learning in Mixed Reality". *Proceedings of Mensch Und Computer 2022*. MuC '22. New York, NY, USA.

— (2025b). "AF-Mix: A gaze-aware learning system with attention feedback in mixed reality". *International Journal of Human-Computer Studies*.

— (2025c). "Mixed Reality Learning Systems with Head-Mounted Displays in Higher Education: A Systematic Review". *Technology, Knowledge and Learning*.

— (n.d.). "GazeNote: Designing Note-taking Support for Immersive Learning with Gaze-Aware and Cross-Device Interaction" ().

Liu, Z., Z. Zhu, L. Zhu, E. Jiang, X. Hu, K. A. Peppler, and K. Ramani (2024). "ClassMeta: Designing Interactive Virtual Classmate to Promote VR Classroom Participation". *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24. New York, NY, USA.

Lungu, A. J., W. Swinkels, L. Claesen, P. Tu, J. Egger, and X. Chen (2021). "A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery". *Expert Review of Medical Devices*.

Maas, M. J. and J. M. Hughes (2020). "Virtual, Augmented and Mixed Reality in K–12 Education: A Review of the Literature". *Technology, Pedagogy and Education*.

Maaß, K. and M. Artigue (2013). "Implementation of inquiry-based learning in day-to-day teaching: a synthesis". *ZDM*.

Magyar, N., X. Xu, and M. Maher (2020). "Creating and Evaluating a Goal Setting Prototype for MOOCs". *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. New York, NY, USA.

Makransky, G., N. K. Andreasen, S. Baceviciute, and R. E. Mayer (2021). "Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality". *Journal of Educational Psychology*.

Makransky, G. and R. E. Mayer (2022). "Benefits of Taking a Virtual Field Trip in Immersive Virtual Reality: Evidence for the Immersion Principle in Multimedia Learning". *Educational Psychology Review*.

Makransky, G. 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*.

Makransky, G., T. S. Terkildsen, and R. E. Mayer (2019). "Adding immersive virtual reality to a science lab simulation causes more presence but less learning". *Learning and Instruction*.

Manfredi, G., U. Erra, and G. Gilio (2023). "A Mixed Reality Approach for Innovative Pair Programming Education with a Conversational AI Virtual Avatar". *ACM International Conference Proceeding Series*.

Mayer, R. E. (1996). "Learners as information processors: Legacies and limitations of educational psychology's second.." *Educational Psychologist*.

— (2004). "Should There Be a Three-Strikes Rule Against Pure Discovery Learning?" *American Psychologist*.

— (2014a). "Cognitive theory of multimedia learning". *The Cambridge handbook of multimedia learning, 2nd ed.* Cambridge handbooks in psychology. New York, NY, US.

— (2014b). *The Cambridge Handbook of Multimedia Learning*.

Mayer, R. E., G. Makransky, and J. Parong (2023). "The Promise and Pitfalls of Learning in Immersive Virtual Reality". *International Journal of Human–Computer Interaction*.

McDaniel, M. A., P. K. Agarwal, B. J. Huelser, K. B. McDermott, and H. L. Roediger (2011). "Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement." *Journal of Educational Psychology*.

McDaniel, M. A., K. M. Wildman, and J. L. Anderson (2012). "Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study". *Journal of Applied Research in Memory and Cognition*.

Melo, G., R. Ravi, L. Jauer, and J. H. Schleifenbaum (2024). "Investigating the use of augmented reality head-mounted displays to teach design for additive manufacturing". *Springer Tracts in Additive Manufacturing*.

Microsoft (2023). *What is mixed reality?*

Milgram, P. and K. Fumio (1994). “A Taxonomy of Mixed Reality Visual Displays”. *IEICE Transactions on Information and Systems*. 77th ser.

Milrad, M., L.-H. Wong, M. Sharples, G.-J. Hwang, C.-K. Looi, and H. Ogata (2013). “Seamless Learning: An International Perspective on Next-Generation Technology-Enhanced Learning”. *Handbook of Mobile Learning*. Ed. by Z. L. Berge and L. Muilenburg. 0th ed.

Mitsuhara, H. and M. Shishibori (2015). “Hybrid Textbook: Fusion of Digital Textbooks and Traditional Learning Tools”. *Int. J. Intell. Syst. Technol. Appl.*

Mitsuhara, H., Y. Yano, and T. Moriyama (2010). “Paper-top interface for supporting note-taking and its preliminary experiment”. *2010 IEEE international conference on systems, man and cybernetics*.

Mohammadhossein, N., A. Richter, and S. Lukosch (2022). “Benefits of Using Augmented Reality in Learning Settings: A Systematic Literature Review”. *ICIS 2022 Proceedings*.

Moos, D. C. (2009). “Note-taking while learning hypermedia: Cognitive and motivational considerations”. *Computers in Human Behavior*. Including the Special Issue: Design Patterns for Augmenting E-Learning Experiences.

Moro, C., C. Phelps, P. Redmond, and Z. Stromberga (2021). “HoloLens and Mobile Augmented Reality in Medical and Health Science Education: A Randomised Controlled Trial.” *British Journal of Educational Technology*.

Moro, C., Z. Štromberga, A. Raikos, and A. Stirling (2017). “The effectiveness of virtual and augmented reality in health sciences and medical anatomy”. *Anatomical Sciences Education*.

Moser, T., M. Hohlagschwandtner, G. Kormann-Hainzl, S. Pölzlbauer, and J. Wolfartsberger (2019). “Mixed Reality Applications in Industry: Challenges and Research Areas”. *Software Quality: The Complexity and Challenges of Software Engineering and Software Quality in the Cloud*. Ed. by D. Winkler, S. Biffl, and J. Bergsmann. Cham.

Moynihan, B. (2021). *Immersive Connections - Teaching, Learning, and Creating in VR at Ithaca College*.

Muller, M. J. and S. Kuhn (1993). “Participatory Design”. *Communications of the ACM*.

Mystakidis, S. (2022). “Metaverse”. *Encyclopedia*.

Mystakidis, S., A. Christopoulos, and N. Pellas (2022). “A systematic mapping review of augmented reality applications to support STEM learning in higher education”. *Education and Information Technologies*.

Mystakidis, S. and V. Lympouridis (2023). “Immersive Learning”. *Encyclopedia*.

Nakazawa, A., M. Iwamoto, R. Kurazume, M. Nunoi, M. Kobayashi, and M. Honda (2023). “Augmented reality-based affective training for improving care communication skill and empathy”. *PLoS One*.

Nebeling, M., M. Speicher, X. Wang, S. Rajaram, B. D. Hall, Z. Xie, A. R. E. Raistrick, M. Aebersold, E. G. Happ, J. Wang, Y. Sun, L. Zhang, L. E. Ramsier, and R. Kulkarni (2020). “MRAT: The Mixed Reality Analytics Toolkit”. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20. New York, NY, USA.

Nguyen, C. and F. Liu (2016). “Gaze-based Notetaking for Learning from Lecture Videos”. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI ’16. New York, NY, USA.

Nguyen, T. V., S. Kamma, V. Adari, T. Lesthaeghe, T. Boehnlein, and V. Kramb (2021). “Mixed reality system for nondestructive evaluation training”. *VIRTUAL REALITY*.

Nuci, K. P., R. Tahir, A. I. Wang, and A. S. Imran (2021). “Game-Based Digital Quiz as a Tool for Improving Students’ Engagement and Learning in Online Lectures”. *IEEE Access*.

O’Keefe, C., L. H. Xu, and D. Clarke (2006). “Kikan-Shido: Between Desks Instruction”. *Making Connections*.

O’Brien, H. L., P. Cairns, and M. 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*.

O’Connor, K. (2022). “Constructivism, curriculum and the knowledge question: tensions and challenges for higher education”. *Studies in Higher Education*.

Oehlberg, J. (2018). *VR Classroom*.

Oh, S., H.-J. So, and M. Gaydos (2018). “Hybrid Augmented Reality for Participatory Learning: The Hidden Efficacy of Multi-User Game-Based Simulation”. *IEEE Transactions on Learning Technologies*.

Page, M. J., J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher (2021). “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”. *BMJ*.

Palmas, F., J. Cichor, D. A. Plecher, and G. Klinker (2019). “Acceptance and Effectiveness of a Virtual Reality Public Speaking Training”. *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*.

Palmas, F., R. Reinelt, J. E. Cichor, D. A. Plecher, and G. Klinker (2021). “Virtual Reality Public Speaking Training: Experimental Evaluation of Direct Feedback Technology Acceptance”. *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*.

Palmer, S., Y. Chu, and A. M. Persky (2019). “Comparison of Rewatching Class Recordings versus Retrieval Practice as Post-Lecture Learning Strategies”. *American Journal of Pharmaceutical Education*.

Pamparău, C., R.-D. Vatavu, A. R. Costea, R. Jurchiş, and A. Opre (2021). “MR4ISL: A Mixed Reality System for Psychological Experiments Focused on Social Learning and Social Interactions”. *Companion of the 2021 ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. EICS ’21. New York, NY, USA.

Panadero, E. (2017). “A Review of Self-regulated Learning: Six Models and Four Directions for Research”. *Frontiers in Psychology*.

Papadopoulos, T., K. Evangelidis, T. H. Kaskalis, G. Evangelidis, and S. Sylaiou (2021). “Interactions in Augmented and Mixed Reality: An Overview”. *Applied Sciences*.

Peden, R. G., R. Mercer, and A. J. Tatham (2016). “The use of head-mounted display eyeglasses for teaching surgical skills: A prospective randomised study”. *INTERNATIONAL JOURNAL OF SURGERY*.

Peng, N., L. Song, L. Wu, H. Wang, and C. Xue (2022). “Cueing effects of colour on attention management in multiple-view visualisations: evidence from eye-tracking by using a dual-task paradigm”. *Behaviour & Information Technology*.

Perez, R. and G. Torres-Delgado (2023). “Evaluation-Self-Regulated Learning in Higher Education in a Digital Environment Under Vulnerable Conditions and Confinement Due to Covid-19: Self-Regulated Learning in a Digital Environment under Vulnerable Conditions”. *Proceedings of the 6th International Conference on Digital Technology in Education*. ICDTE ’22. New York, NY, USA.

Peters, O. (2000). “Digital Learning Environments: New Possibilities and Opportunities”. *International Review of Research in Open and Distance Learning*.

Petersen, G. B., A. Mottelson, and G. Makransky (2021). “Pedagogical Agents in Educational VR: An in the Wild Study”. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI ’21. New York, NY, USA.

Petersen, K., S. Vakkalanka, and L. Kuzniarz (2015). “Guidelines for conducting systematic mapping studies in software engineering: An update”. *Information and Software Technology*.

Pinkl, J., J. Villegas, and M. Cohen (2024). “Multimodal Drumming Education Tool in Mixed Reality”. *Multimodal Technologies and Interaction*.

Pintrich, P. R. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Tech. rep.

Piolat, A., T. Olive, and R. T. Kellogg (2005). “Cognitive effort during note taking”. *Applied Cognitive Psychology*.

Plooy, E. du, D. Casteleijn, and D. Franzsen (2024). “Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement”. *Heliyon*.

Plopski, A., T. Hirzle, N. Norouzi, L. Qian, G. Bruder, and T. Langlotz (2022). “The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-Worn Extended Reality”. *ACM Comput. Surv.*

Porta, M. (2008). “Implementing Eye-Based User-Aware e-Learning”. *CHI ’08 Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’08. New York, NY, USA.

Quintana, C., B. J. Reiser, E. A. Davis, J. Krajcik, E. Fretz, R. G. Duncan, E. Kyza, D. Edelson, and E. Soloway (2004). “A Scaffolding Design Framework for Software to Support Science Inquiry”. *The Journal of the Learning Sciences*. JSTOR: 1466941.

Radianti, J., T. A. Majchrzak, J. Fromm, and I. Wohlgenannt (2020). “A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda”. *Computers & Education*.

Radu, I., V. Hv, and B. Schneider (2021a). “Unequal Impacts of Augmented Reality on Learning and Collaboration During Robot Programming with Peers”. *Proc. ACM Hum.-Comput. Interact.*

Radu, I., T. Joy, and B. Schneider (2021b). “Virtual Makerspaces: Merging AR/VR/MR to Enable Remote Collaborations in Physical Maker Activities”. *Extended Abstracts of the*

2021 CHI Conference on Human Factors in Computing Systems. CHI EA '21. New York, NY, USA.

Radu, I. and B. Schneider (2019). "What Can We Learn from Augmented Reality (AR)? Benefits and Drawbacks of AR for Inquiry-Based Learning of Physics". *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA.

Raes, A. and F. Depaepe (2020). "A longitudinal study to understand students' acceptance of technological reform. When experiences exceed expectations". *Education and Information Technologies*.

Raes, A., P. Vanneste, M. Pieters, I. Windey, W. Van Den Noortgate, and F. Depaepe (2020). "Learning and instruction in the hybrid virtual classroom: An investigation of students' engagement and the effect of quizzes". *Computers & Education*.

Rajaram, S. and M. Nebeling (2022). "Paper Trail: An Immersive Authoring System for Augmented Reality Instructional Experiences". *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22. New York, NY, USA.

Rappa, N. A., S. Ledger, T. Teo, K. Wai Wong, B. Power, and B. Hilliard (2022). "The use of eye tracking technology to explore learning and performance within virtual reality and mixed reality settings: a scoping review". *Interactive Learning Environments*.

Rayner, K. (1998). "Eye movements in reading and information processing: 20 years of research". *Psychological Bulletin*.

Rayner, K., T. J. Slattery, and N. N. Bélanger (2010). "Eye movements, the perceptual span, and reading speed". *Psychonomic Bulletin & Review*.

Ren, Y., Y. Li, and E. Lank (2014). "InkAnchor: Enhancing Informal Ink-Based Note Taking on Touchscreen Mobile Phones". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. New York, NY, USA.

Richards, S. (2023). "Student Engagement Using HoloLens Mixed-Reality Technology in Human Anatomy Laboratories for Osteopathic Medical Students: An Instructional Model." *Medical science educator*.

Roda, C. and J. Thomas (2006). "Attention aware systems: Theories, applications, and research agenda". *Computers in Human Behavior*. Attention aware systems.

Roediger III, H. L. and J. D. Karpicke (2006). "Test-enhanced learning: Taking memory tests improves long-term retention". *Psychological Science*.

Rojas-Munõz, E., M. Cabrera, C. Lin, N. Sánchez-Tamayo, D. Andersen, V. Popescu, K. Anderson, B. Zarzaur, B. Mullis, and J. Wachs (2020). "Telementoring in Leg Fasciotomies via Mixed-Reality: Clinical Evaluation of the STAR Platform". *Military Medicine*.

Rokhsaritalemi, S., A. Sadeghi-Niaraki, and S.-M. Choi (2020). "A Review on Mixed Reality: Current Trends, Challenges and Prospects". *Applied Sciences*.

Rolland, J. P. and H. Fuchs (2000). "Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization". *Presence*.

Romat, H., N. Marquardt, K. Hinckley, and N. Henry Riche (2022). "Style Blink: Exploring Digital Inking of Structured Information via Handcrafted Styling as a First-Class Object". *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22. New York, NY, USA.

Roo, J. S. and M. Hachet (2017). "One Reality: Augmenting How the Physical World Is Experienced by Combining Multiple Mixed Reality Modalities". *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. UIST '17. Québec City QC Canada.

Ross, B., A.-M. Chase, D. Robbie, G. Oates, and Y. Absalom (2018). "Adaptive quizzes to increase motivation, engagement and learning outcomes in a first year accounting unit". *International Journal of Educational Technology in Higher Education*.

Rouder, J. N., P. L. Speckman, D. Sun, R. D. Morey, and G. Iverson (2009). "Bayesian t tests for accepting and rejecting the null hypothesis". *Psychonomic Bulletin & Review*.

Rukangu, A. M., J. R. Morelock, K. J. Johnsen, and D. O. Moyaki (2025). "Virtual and physical robots in engineering education: a study on motivation and learning with augmented reality". *Computers in Education Journal*.

Ryan, G. V., S. Callaghan, A. Rafferty, M. F. Higgins, E. Mangina, and F. McAuliffe (2022). "Learning Outcomes of Immersive Technologies in Health Care Student Education: Systematic Review of the Literature". *Journal of Medical Internet Research*.

Ryan, R. M., V. Mims, and R. Koestner (1983). "Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory". *Journal of Personality and Social Psychology*.

Schnitzer, B. L., S. Gampe, A. Bieberstein, I. Hoffmann, D. Görlich, and O. Korn (2025). "Language Learning in Virtual Reality: Enhancement of Long-Term Vocabulary Recognition and Understanding through Full-Body Avatars". *Proceedings of the Extended Abstracts* *of the CHI Conference on Human Factors in Computing Systems*. CHI EA ’25. New York, NY, USA.

Schoeb, D. S., J. Schwarz, S. Hein, D. Schlager, P. F. Pohlmann, A. Frankenschmidt, C. Gratzke, and A. Miernik (2020). “Mixed reality for teaching catheter placement to medical students: A randomized single-blinded, prospective trial”. *BMC MEDICAL EDUCATION*.

Schrepp, M., A. Hinderks, and J. Thomaschewski (2017). “Construction of a Benchmark for the User Experience Questionnaire (UEQ)”. *International Journal of Interactive Multimedia and Artificial Intelligence*.

Schultz, T. and A. Maedche (2023). “Biosignals meet Adaptive Systems”. *SN Applied Sciences*.

Schulz, T., S. Liu, A. Mädche, and C. Weinhardt (2024). “Enhancing AI Education for Business Students through Extended Reality: An Exploratory Study”. *Proceedings of Mensch und Computer 2024*. MuC ’24. New York, NY, USA.

Schunk, D. H. (2012). *Learning Theories: An Educational Perspective*. 6th ed. Boston.

Schunk, D. H. and J. A. Greene, eds. (2017). *Handbook of Self-Regulation of Learning and Performance*. 2nd ed.

Seaman, J., M. Brown, and J. Quay (2017). “The Evolution of Experiential Learning Theory: Tracing Lines of Research in the JEE”. *Journal of Experiential Education*.

Seele, S., S. Misztal, H. Buhler, R. Herpers, and J. Schild (2017). “Here’s Looking At You Anyway! How Important is Realistic Gaze Behavior in Co-located Social Virtual Reality Games?” *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. CHI PLAY ’17. New York, NY, USA.

Serrano Vergel, R., P. Morillo Tena, S. Casas Yrurzum, and C. Cruz-Neira (2020). “A comparative evaluation of a virtual reality table and a HoloLens-Based augmented reality system for anatomy training”. *IEEE Transactions on Human-Machine Systems*.

Shadiev, R. and D. Li (2023). “A review study on eye-tracking technology usage in immersive virtual reality learning environments”. *Computers & Education*.

Shao, Q., A. Sniffen, J. Blanchet, M. E. Hillis, X. Shi, T. K. Haris, J. Liu, J. Lamberton, M. Malzkuhn, L. C. Quandt, J. Mahoney, D. J. M. Kraemer, X. Zhou, and D. Balkcom (2020). “Teaching American Sign Language in Mixed Reality”. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*

Sharma, K., H. S. Alavi, P. Jermann, and P. Dillenbourg (2016). “A Gaze-Based Learning Analytics Model: In-Video Visual Feedback to Improve Learner’s Attention in MOOCs”.

Proceedings of the Sixth International Conference on Learning Analytics & Knowledge. LAK '16. New York, NY, USA.

Sharma, K., M. Giannakos, and P. Dillenbourg (2020). "Eye-tracking and artificial intelligence to enhance motivation and learning". *Smart Learning Environments*.

Sharples, M. (2015). "Seamless Learning: Using Location-Aware Technology to Support Art Education". *Computer*.

Shen, S., C. T. Tan, H.-T. Chen, W. L. Raffe, and T. W. Leong (2025). "Educator Perceptions of XRAuthor: An Accessible Tool for Authoring Learning Content with Different Immersion Levels". *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. CHI '25. New York, NY, USA.

Shrager, L. and R. E. Mayer (1989). "Note-taking fosters generative learning strategies in novices". *Journal of Educational Psychology*.

Shuell, T. J. (1986). "Cognitive Conceptions of Learning". *Review of Educational Research*.

Skinner, B. F. (1961). *The Science of Learning and the Art of Teaching*. Cumulative record, Enlarged ed. East Norwalk, CT, US.

Speicher, M., B. D. Hall, and M. Nebeling (2019). "What Is Mixed Reality?" *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk.

Sprague, G. A. (1981). "Cognitive Psychology and Instructional Development: Adopting a Cognitive Perspective for Instructional Design Programs in Higher Education". *Educational Technology*.

Stemasov, E., J. Hohn, M. Cordts, A. Schikorr, E. Rukzio, and J. Gugenheimer (2023). "BrickStARt: Enabling In-situ Design and Tangible Exploration for Personal Fabrication using Mixed Reality". *Proceedings of the ACM on Human-Computer Interaction*.

Stolterman, E. and M. Wiberg (2010). "Concept-Driven Interaction Design Research". *Human–Computer Interaction*.

Stoner, R., M. Ahmad, S. B. Patel, A. Cowell, T. J. Hurkxkens, L. Bastrot, C. Caradu, D. G. Amiras, and J. Shalhoub (2024). "A comparison between augmented reality and traditional in-person teaching for vascular anastomotic surgical skills training". *JVS-Vascular Insights*.

Su, K.-W., H.-J. Chen, and C.-T. Chen (2025). "Development and evaluation of a mixed reality system for health education on neck stiffness and soreness". *INTERNATIONAL JOURNAL OF INDUSTRIAL ERGONOMICS*.

Szafir, D. and B. Mutlu (2013). “ARTFul: adaptive review technology for flipped learning”. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Paris France.

Tang, Y. M., K. M. Au, H. C. W. Lau, G. T. S. Ho, and C. H. Wu (2020). “Evaluating the effectiveness of learning design with mixed reality (MR) in higher education”. *VIRTUAL REALITY*.

Tang, Y., J. Situ, A. Y. Cui, M. Wu, and Y. Huang (2025). “LLM Integration in Extended Reality: A Comprehensive Review of Current Trends, Challenges, and Future Perspectives”. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. CHI ’25. New York, NY, USA.

Tauseef, M., A. Ullal, A. Watkins, Z. Ingram, C. Maxwell, J. Tate, L. Juckett, L. C. Mion, and N. Sarkar (2024). “From lab to a long-term care facility: Lessons learned from field deployment of augmented reality telepresence system as an interactive communication technology”. *2024 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-adjunct)*.

Thanyadit, S., M. Heintz, and E. L.-C. Law (2023). “Tutor In-sight: Guiding and Visualizing Students’ Attention with Mixed Reality Avatar Presentation Tools”. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI ’23. New York, NY, USA.

Tholander, J. and M. Normark (2020). “Crafting personal information - resistance, imperfection, and self-creation in bullet journaling”. *Proceedings of the 2020 CHI conference on human factors in computing systems*. CHI ’20. New York, NY, USA.

Thoravi Kumaravel, B., F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman (2019). “Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence”. *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. New Orleans LA USA.

Thorndike, E. L. (1932). *The fundamentals of learning*. The fundamentals of learning. New York, NY, US.

Toreini, P., M. Langner, A. Maedche, S. Morana, and T. Vogel (2022). “Designing Attentive Information Dashboards”. *Journal of the Association for Information Systems*.

Tsai, H.-R., S.-K. Chiu, and B. Wang (2025). “GazeNoter: Co-Piloted AR Note-Taking via Gaze Selection of LLM Suggestions to Match Users’ Intentions”. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. CHI ’25. New York, NY, USA.

Van Der Veer, G. C., S. R. Humayoun, V. M. Memmesheimer, and A. Ebert (2024). “Experience 2.0 and Beyond - Engineering Cross Devices and Multiple Realities”. *Companion Proceedings of the 16th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. EICS ’24 Companion. New York, NY, USA.

VanGundy, A. B. (1984). “BRAIN WRITING FOR NEW PRODUCT IDEAS: AN ALTERNATIVE TO BRAINSTORMING”. *Journal of Consumer Marketing*.

Vaze, A., A. Morris, and I. Clarke (2024). “CuriosityXR: Context-aware education experiences with mixed reality and conversation AI”. *2024 IEEE international conference on artificial intelligence and eXtended and virtual reality (AIxVR)*.

Vazquez, C. D., A. A. Nyati, A. Luh, M. Fu, T. Aikawa, and P. Maes (2017). “Serendipitous Language Learning in Mixed Reality”. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’17. Denver Colorado USA.

Vehtari, A. (2024). *Bayesian Data Analysis Course*.

Velichkovsky, B., A. Sprenger, and P. Unema (1997). “Towards Gaze-Mediated Interaction: Collecting Solutions of the “Midas Touch Problem””. *Human-Computer Interaction INTERACT ’97: IFIP TC13 International Conference on Human-Computer Interaction, 14th–18th July 1997, Sydney, Australia*. Ed. by S. Howard, J. Hammond, and G. Lindgaard. IFIP — The International Federation for Information Processing. Boston, MA.

Vergel, R. S., P. Morillo Tena, S. Casas Yrurzum, and C. Cruz-Neira (2020). “A Comparative Evaluation of a Virtual Reality Table and a HoloLens-Based Augmented Reality System for Anatomy Training”. *IEEE Transactions on Human-Machine Systems*.

Vertegaal, R., J. S. Shell, D. Chen, and A. Mamuji (2006). “Designing for augmented attention: Towards a framework for attentive user interfaces”. *Computers in Human Behavior*. Attention aware systems.

Viglialoro, R. M., S. Condino, G. Turini, M. Carbone, V. Ferrari, and M. Gesi (2021). “Augmented Reality, Mixed Reality, and Hybrid Approach in Healthcare Simulation: A Systematic Review”. *Applied Sciences*.

Vovk, A., F. Wild, W. Guest, and T. Kuula (2018). "Simulator Sickness in Augmented Reality Training Using the Microsoft HoloLens". *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. New York, NY, USA.

Wambsganss, T., T. Kueng, M. Soellner, and J. M. Leimeister (2021). "ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills". *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. New York, NY, USA.

Wambsganss, T., M. Soellner, K. R. Koedinger, and J. M. Leimeister (2022). "Adaptive Empathy Learning Support in Peer Review Scenarios". *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22. New York, NY, USA.

Wang, C., B. K. Daniel, M. Asil, P. Khwaounjoo, and Y. O. Cakmak (2020). "A Randomised Control Trial and Comparative Analysis of Multi-Dimensional Learning Tools in Anatomy". *Scientific Reports*.

Wang, T.-I. and P.-C. Chang (2017). "The Effect of 635 Brainstorming on the Creativity of Programming Design". *Emerging Technologies for Education*. Ed. by T.-T. Wu, R. Gennari, Y.-M. Huang, H. Xie, and Y. Cao. Lecture Notes in Computer Science. Cham.

Webster, J. and R. T. Watson (2002). "Analyzing the Past to Prepare for the Future: Writing a Literature Review". *MIS Quarterly*.

Weegar, M. and D. Pacis (2012). "A Comparison of Two Theories of Learning – Behaviorism and Construct ivism as applied to Face-to-Face and On line Learning".

Wilson, C. (2013). *Brainstorming and Beyond: A User-Centered Design Method*. 1st ed. Amsterdam.

Wish–Baratz, S., A. P. Gubatina, R. Enterline, and M. A. Griswold (2019). "A new supplement to gross anatomy dissection: HoloAnatomy". *Medical Education*.

Wobbrock, J. O. and J. A. Kientz (2016). "Research contributions in human-computer interaction". *Interactions*.

Wohlgenannt, I., J. Fromm, S. Stieglitz, J. Radianti, and T. A. Majchrzak (2019). "Virtual Reality in Higher Education: Preliminary Results from a Design-Science-Research Project". *Proceedings of the International Conference on Information Systems Development (ISD)*.

Won, M., D. A. K. Ungu, H. Matovu, D. F. Treagust, C.-C. Tsai, J. Park, M. Mocerino, and R. Tasker (2023). "Diverse approaches to learning with immersive Virtual Reality identified from a systematic review". *Computers & Education*.

Wu, H.-K., S. W.-Y. Lee, H.-Y. Chang, and J.-C. Liang (2013). "Current Status, Opportunities and Challenges of Augmented Reality in Education". *Computers & Education*.

Wunder, L., N. Gomez, J. Gonzalez, G. Mitzova-Vladinov, M. Cacchione, J. Mato, C. Foronda, and J. Groom (2020). "Fire in the operating room: Use of mixed reality simulation with nurse anesthesia students". *Informatics*.

Xu, S., D. Hu, R. Wang, and X. Zhang (2025). "PeerEdu: Bootstrapping online learning behaviors via asynchronous area of interest sharing from peer gaze". *Proceedings of the 2025 CHI conference on human factors in computing systems*. Chi '25. New York, NY, USA.

Yang, M.-T. and Y.-C. Chiu (2014). "Note-taking for 3D curricular contents using markerless augmented reality". *Interacting with Computers*.

Yang, M.-T. and W.-C. Liao (2014). "Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction". *IEEE Transactions on Learning Technologies*.

Yu, J., A. R. Denham, and E. Searight (2022). "A systematic review of augmented reality game-based Learning in STEM education". *Educational technology research and development*.

Yu, S.-J., J. C.-Y. Sun, and O. T.-C. Chen (2019). "Effect of AR-based online wearable guides on university students' situational interest and learning performance". *UNIVERSAL ACCESS IN THE INFORMATION SOCIETY*.

Yuan, Q., K. Shou, J. Xu, Z. Yao, and Z. Pan (2024). "Examining the impact of head-mounted augmented reality on learning engineering drawings: A case study for three-view drawing". *COMPUTER APPLICATIONS IN ENGINEERING EDUCATION*.

Yue, Y.-T., Y.-L. Yang, G. Ren, and W. Wang (2017). "SceneCtrl: Mixed Reality Enhancement via Efficient Scene Editing". *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. UIST '17. Québec City QC Canada.

Zander, T. O., C. Kothe, S. Jatzev, and M. Gaertner (2010). "Enhancing Human-Computer Interaction with Input from Active and Passive Brain-Computer Interfaces". *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*. Ed. by D. S. Tan and A. Nijholt. London.

Zarraonandia, T., P. Diaz, A. Montero, I. Aedo, and T. Onorati (2019). "Using a Google Glass-Based Classroom Feedback System to Improve Students to Teacher Communication". *IEEE access : practical innovations, open solutions*.

Zhang, X., Y. Chen, L. Hu, and Y. Wang (2022). “The metaverse in education: Definition, framework, features, potential applications, challenges, and future research topics”. *Frontiers in Psychology*.

Zhang, Y., A. Nowak, A. Romanowski, and M. Fjeld (2023). “Virtuality or Physicality? Supporting Memorization Through Augmented Reality Gamification”. *Companion Proceedings of the 2023 ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. EICS ’23 Companion. New York, NY, USA.

Zhao, Y., Y. Bozhkov, X. Chen, K. Fuchs, M. Buchfelder, L. Fester, D. S. de Oliveira, A. Del Vecchio, and T. Kinfe (2025). “First German experience using augmented reality for neuroanatomy education in undergraduate medical students: a feasibility and questionnaire-based study”. *ETR&D-EDUCATIONAL TECHNOLOGY RESEARCH AND DEVELOPMENT*.

Zhu, F., M. Sousa, L. Sidenmark, and T. Grossman (2024). “PhoneInVR: An Evaluation of Spatial Anchoring and Interaction Techniques for Smartphone Usage in Virtual Reality”. *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA.

Zhu, X. T., H. Cheerman, M. Cheng, S. R. Kiami, L. Chukoskie, and E. McGivney (2025). “Designing VR Simulation System for Clinical Communication Training with LLMs-Based Embodied Conversational Agents”. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI EA ’25. New York, NY, USA.

Zimmerman, B. J. (2002). “Becoming a Self-Regulated Learner: An Overview”. *Theory Into Practice*.

Zoghi, B., G. Bhutra, and A. Parameswaran (2018). “A novel use for wearable augmented reality in engineering education”. *Computers in Education Journal*.

References to Code Repositories and Data Sets

This dissertation follows open science principles. The code and data for the presented studies can be accessed on the following platforms:

Study I

The coding matrix and the data analysis using JASP of the reviewed research articles can be viewed at: `https://osf.io/duvqs/?view_only=6492fd17b9cb4af9a72c7d8681233438`

Study II

Artifact and Code: `https://gitlab.kit.edu/kit/win/h-lab/research/2195-liu-shi-af-mix`

Data Repository: `https://radar.kit.edu/radar/de/dataset/bt1rmgph2zxr29m2`

Study III

Artifact and Code: `https://gitlab.kit.edu/kit/win/h-lab/research/2212_liu_shi_gazenote`

Data Repository: `https://radar.kit.edu/radar/de/dataset/v2pxmbqhjvg8cdx9`

Study IV

Artifact and Code: `https://gitlab.kit.edu/kit/win/h-lab/research/2831_liu_shi_attentivelearn`

Data Repository: `https://radar.kit.edu/radar/de/dataset/q95h8fj6d460570j`

List of Publications

Accepted, Peer-Reviewed Publications

Liu, S., Toreini, P., & Maedche, A. (2022). Designing Gaze-Aware Attention Feedback for Learning in Mixed Reality. Proceedings of Mensch und Computer 2022, 503–508. https://doi.org/10.1145/3543758.3547565

Liu, S., Toreini, P., & Maedche, A. (2025). Mixed Reality Learning Systems with Head-Mounted Displays in Higher Education: A Systematic Review. Technology, Knowledge and Learning. https://doi.org/10.1007/s10758-025-09912-z

Liu, S., Toreini, P., & Maedche, A. (2025). AF-Mix: A gaze-aware learning system with attention feedback in mixed reality. International Journal of Human-Computer Studies, 198, 103467. https://doi.org/10.1016/j.ijhcs.2025.103467

Liu, S., Mädche, A., & Feick, M. (2025). GazeClass: Towards Gaze-Adaptive Cross-Device Learning Support for Virtual Classrooms. Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 1–12. https://doi.org/10.1145/3706599.3720232

Liu, S., Feick, M., Bierhoff, L., & Maedche, A. (2026). AttentiveLearn: Personalized Post-Lecture Support for Gaze-Aware Immersive Learning. Proceedings of the CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3772318.3790667

Publications Under Review

Liu, S., Toreini, P., & Maedche, A. (n.d.). GazeNote: Designing Note-taking Support for Immersive Learning with Gaze-Aware and Cross-Device Interaction. (Under Review at *Interacting with Computers*)

Eidesstattliche Versicherung

gemäß § 13 Abs. 2 Ziff. 3 der Promotionsordnung des Karlsruher Instituts für Technologie für die KIT-Fakultät für Wirtschaftswissenschaften

1. Bei der eingereichten Dissertation zu dem Thema *Designing Gaze-Adaptive Immersive Learning Support* handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Ich versichere an Eidesstatt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Karlsruhe, den 23.04.2026