

# AttentiveLearn: Personalized Post-Lecture Support for Gaze-Aware Immersive Learning

Shi Liu

human-centered systems lab (h-lab)  
Karlsruhe Institute of Technology (KIT)  
Karlsruhe, Germany  
shi.liu@kit.edu

Linus Bierhoff

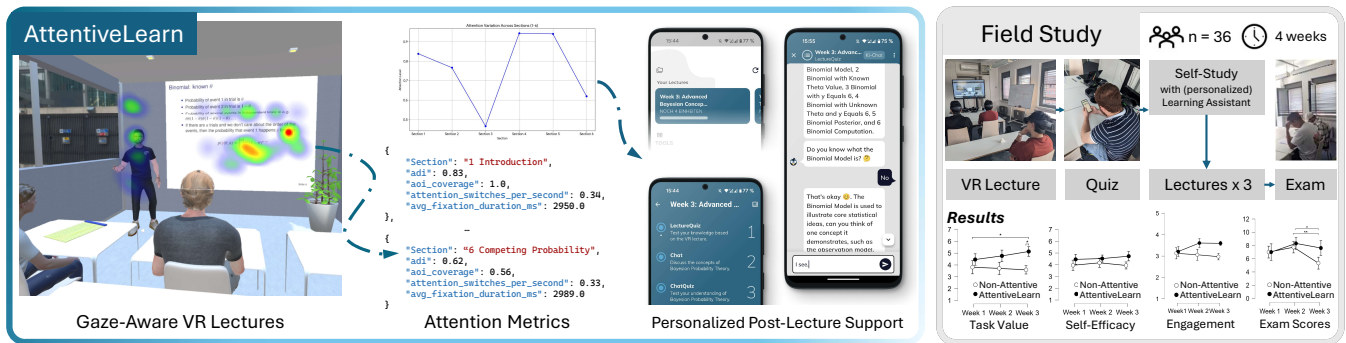
human-centered systems lab (h-lab)  
Karlsruhe Institute of Technology (KIT)  
Karlsruhe, Germany  
linus.bierhoff@student.kit.edu

Martin Feick

human-centered systems lab (h-lab)  
Karlsruhe Institute of Technology (KIT)  
Karlsruhe, Germany  
martin.feick@kit.edu

Alexander Maedche

human-centered systems lab (h-lab)  
Karlsruhe Institute of Technology (KIT)  
Karlsruhe, Germany  
alexander.maedche@kit.edu



**Figure 1: Overview of the *AttentiveLearn* system and study design. The system combines (a) gaze-aware VR lectures, (b) attention metrics, and (c) personalized post-lecture support via an LLM-based mobile assistant. The field study with 36 students over four weeks evaluated the system’s impact on engagement, motivation, and learning outcomes in a real-world lecture setting.**

## 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.

## CCS Concepts

• Human-centered computing → Mixed / augmented reality.

## Keywords

virtual reality, immersive learning, eye-tracking, adaptive support

### ACM Reference Format:

Shi Liu, Martin Feick, Linus Bierhoff, and Alexander Maedche. 2026. *AttentiveLearn: Personalized Post-Lecture Support for Gaze-Aware Immersive Learning*. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3772318.3790667>

## 1 Introduction

Using immersive technologies such as Virtual Reality (VR) to create engaging learning experiences is a broad research field known as immersive learning [44, 50]. Prior HCI research has investigated the design of immersive learning systems across various domains, ranging from digital fabrication [62] to language learning [70]. Different affordances and formats of immersive learning have been explored, including virtual laboratories [62], simulations [86], and



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI '26, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-2278-3/26/04  
<https://doi.org/10.1145/3772318.3790667>

virtual classrooms [25]. While immersive learning environments have shown promise for better learning outcomes [7, 43], providing effective learner support in these environments remains a challenge [57, 72]. A key setting where this challenge has been recognized is the immersive virtual classroom in VR [25, 27]. Immersive virtual classrooms have emerged as a prominent example of immersive learning, both widely researched in HCI [25, 39, 41] and increasingly adopted in practice [49]. Prior work has examined design features such as embodied agents [41], large language model (LLM) integration [74], and annotation tools [76] 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 [56]. Educational research emphasizes that learning is a continuous and ubiquitous process extending beyond the classroom [31, 37]. Techniques such as tutoring [19], quizzes [64], content reviews [73], 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 [21, 67, 73]. For example, biosignal-based attention metrics have been used to drive adaptive content reviews for desktop-based learning systems [73]. 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 [67], engagement [51], and learning outcomes [67], 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 [34]. In HCI research, eye-tracking technologies have been widely used to support attention across diverse contexts [22, 65], including education [20, 30] and in VR systems [27, 59]. 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 [66]. They have also been shown to increase engagement [64]. Moreover, adaptive and personalized quizzes have been explored as promising post-lecture scaffolds [67]. 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.

## 2 Related Work

### 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 [44, 50]. Empirical evaluations of immersive learning systems reveal a broad range of learning benefits. For instance, a VR field study conducted by Petersen et al. [57] demonstrated improved enjoyment and knowledge acquisition in the context of museum guides. Moreover, Radu and Schneider [63] identified students' increased self-efficacy after using an immersive physics learning application. Other positive impacts include improved comprehension [68], motivation [75], and achievement across contexts [60]. However, research also reports mixed or contradictory effects [42, 60], 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 [44]. To address the need for learner-centered support, Palmas et al. [54] investigated public-speaking training in VR and how skill development can be supported through direct feedback and presence-enhancing mechanisms [55].

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 [41] and VR videos as virtual excursions [14], to enhance engagement and interactivity. To further enrich the classroom experience, prior work has introduced both teacher-centric tools (e.g., authoring tools [72], teaching augmentation [4]) and learner-centered support mechanisms [41]. 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 [73, 75], VR introduces unique immersive dynamics, such as spatial attention shifts and embodied presence, making dedicated support for immersive learning necessary [41].

For immersive virtual classrooms, learner-centered designs have included embodied agents acting as virtual peers during lectures [41], as well as studies of classroom dynamics such as avatar representation, teacher-student proximity, and classroom layout [8, 25]. 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 [76] and attention-based warnings during lectures [27], 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.**

## 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 [65]. Attention has been defined as the process of selecting relevant perceived information, allowing individuals to become “active seekers and processors” of knowledge [15]. In learning, attention is essential for effective information processing [45] and sustaining situational interest [44]. In the cognitive-affective model of immersive learning proposed by Makransky and Petersen [44], attention is identified as a crucial factor influencing learning outcomes.

HCI research has developed attention-aware systems to scaffold attentional processes [65]. 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 [12, 78]. 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 [22]. In existing research, attention-aware systems have included gaze-adaptive intelligent tutoring [20], note-taking [35], reading & writing support [13, 38],

and adaptive feedback during learning tasks [32, 40]. Besides these gaze-adaptive systems, EEG-based cognitive support and attention monitoring have also been explored [65, 82].

Specifically for lecture consumption and in classroom settings, Hutt et al. [29] used eye-tracking to detect mind wandering during recorded lectures, demonstrating the feasibility of in-lecture attention monitoring. Xu et al. [81] developed a learning feedback system that aggregates the gaze data of peers to inform the area of interests (AOI) during lectures. Furthermore, Szaifir and Mutlu [73] 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 [59]. Eye-tracking has been established as a reliable and non-intrusive approach for attention-aware support [15, 22, 65]. 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. [1] 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 [73], 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. [2], 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. [27] and Tsai et al. [76] designed gaze-adaptive support mechanisms within VR sessions, such as adaptive hints and note-taking aids. Furthermore, Liu et al. [41] 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.

### 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 [23]. Besides, the testing effect of post-lecture quizzes has been widely observed to improve knowledge retention [46], self-efficacy [23], and engagement [51].

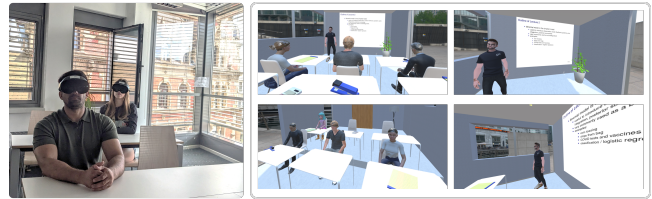
Beyond static, pre-defined questions, adaptive and personalized quizzes have also been supported in digital learning platforms. For instance, Ross et al. [67] designed adaptive quizzes that adjusted question difficulty based on prior responses, leading to improvements in motivation and engagement. Similarly, Klaveren et al. [36] showed that adaptive question sequencing better supported diverse learning needs. More recent work by Contrino et al. [16] 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 [19], assessment tools [24], personalized learning analytics [18], 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 [31]. They are capable of delivering a ubiquitous learning experience independent of human tutors after the lectures [31]. Studies have integrated quizzes into mobile assistants [51]. With AI-driven personalization, recent work has explored LLM-based assistants with adaptive learning support [79]. For immersive learning, cross-device interaction research in HCI already underscores the opportunity of extending immersive learning through mobile devices [11, 85]. 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, their integration into immersive learning with attention-aware personalization has yet to be explored.**

## 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



**Figure 2: 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.**

assistant. Specifically, we investigate how attention-aware personalization influences students' motivation, engagement, and learning outcomes.

### 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 [27, 29], and (2) virtual classrooms remain a widely adopted format in both HCI research and practice [25, 49], with open-source infrastructures further provide a practical foundation for the design space [52].

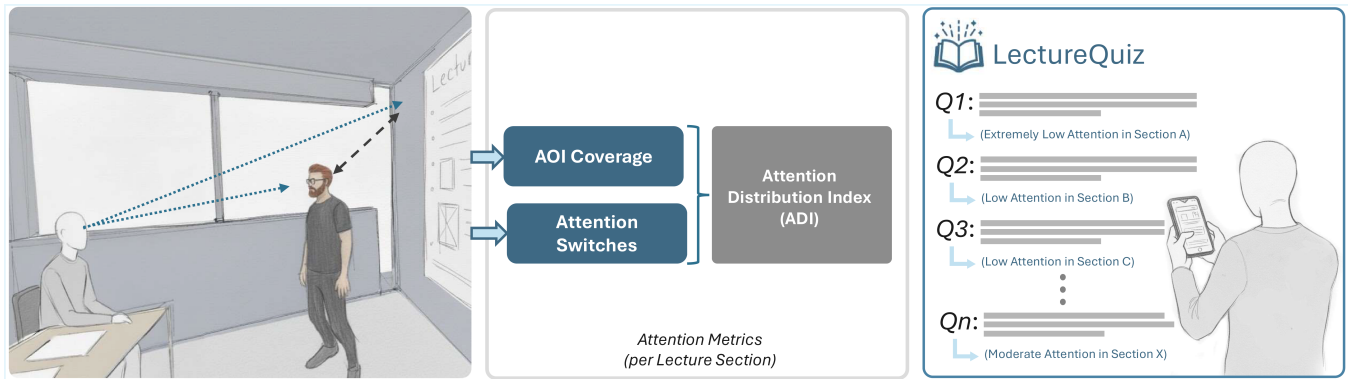
For our implementation, we adopted the existing Unity-based virtual classroom from Liu et al. [39] with their consent. This system provides a standard classroom setup consisting of a slideshow and multiple embodied avatars. In their classroom setup, one avatar represents the lecturer, delivering a pre-recorded 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 [27] and teacher-oriented attention monitoring [75]. 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. [71].

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.

### 3.2 Attention-Aware Personalization Pipeline

*Data Processing.* We implemented a Python-based web server using Flask-SocketIO<sup>1</sup> that connects the VR classroom with a mobile

<sup>1</sup><https://flask-socketio.readthedocs.io>



**Figure 3: 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.

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. [39]: 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 [71]. The ADI metric combines both measures to represent overall attention across lecture sections, following the methodology of Sharma et al. [71] and Hutt et al. [29]. 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 [46]. 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 5.2). This integration ensures that attention-aware personalization is not an isolated design feature but part of a comprehensive learning ecosystem.

### 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

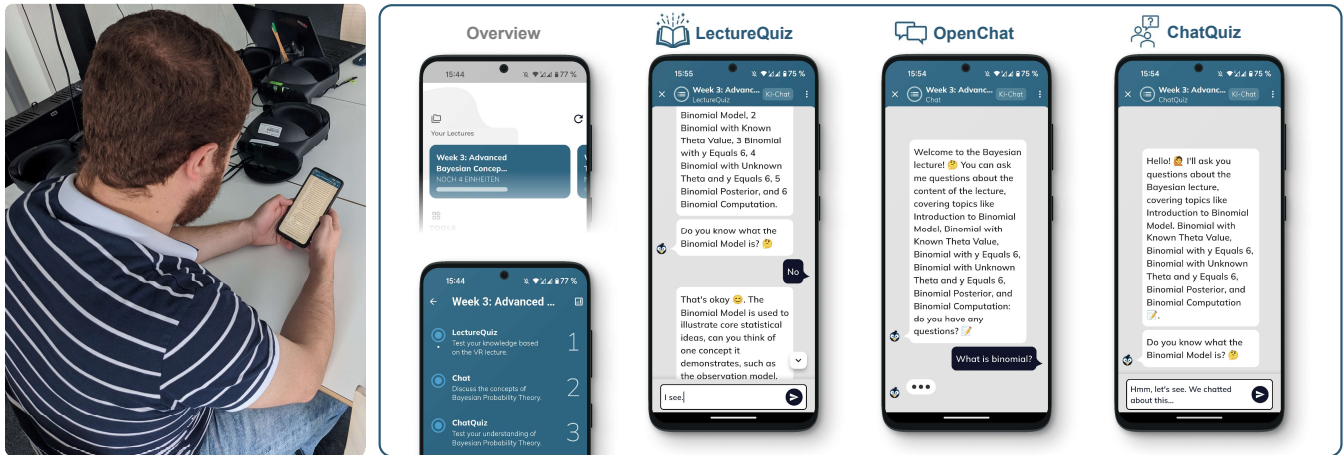
self-regulated learning at our university, serving as the framework’s *Ex-Situ Intervention* interface. The assistant has been implemented as a native Android and iOS application using Dart and the Flutter framework<sup>2</sup>. This application provides a familiar platform for students, enabling us 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 4).

*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 [5]. 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 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.

<sup>2</sup><https://flutter.dev/>



**Figure 4: 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.**

## 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 [77], 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.

### 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) [53].
- **Motivation:** measured weekly with the Motivated Strategies for Learning Questionnaire (MSLQ) [58].
- **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.

### 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.

### 4.3 Task and Procedure

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 [77]. 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 [77], as well as the exercise questions in the course textbook [26], 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<sup>3</sup>. This weekly cycle was repeated for three weeks (see Figure 5).

**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 [77], textbook exercises [26], and LLM-generated questions, 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.

### 4.4 Data Collection and Analysis

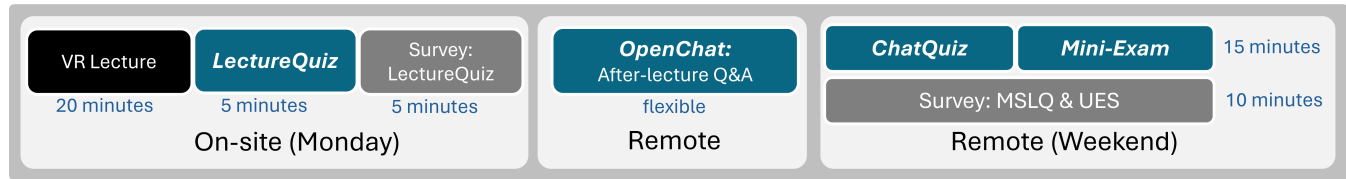
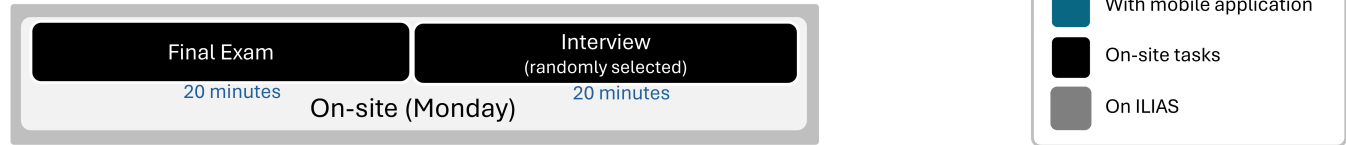
**Data Collection.** We collected multimodal data for mixed-methods analysis: (a) eye-tracking during lectures (AOI coverage percentage, attention switches, ADI), (b) log data from the mobile assistant including *LectureQuiz* scores, (c) on-site survey (two items on *LectureQuiz*) and weekly surveys (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 [58]. The UES-SF measured four engagement dimensions (focused attention, perceived usability, aesthetic appeal, reward factor) on a 5-point Likert scale [53]. 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.

<sup>3</sup>with the built-in survey tool on the ILLIAS platform

**Week 1-3****Week 4**

**Figure 5: 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.**

Qualitative data from the interviews were analyzed using reflexive thematic analysis (TA). Following the six-phase approach [9], 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 [10], 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 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.

### 5.1 In-Lecture Attention

*Overall Attention Distribution.* As shown in Figure 6, we visualize the eye-tracking data collected during the three-week VR lectures. The AOI coverage percentage (see Figure 6a) 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.

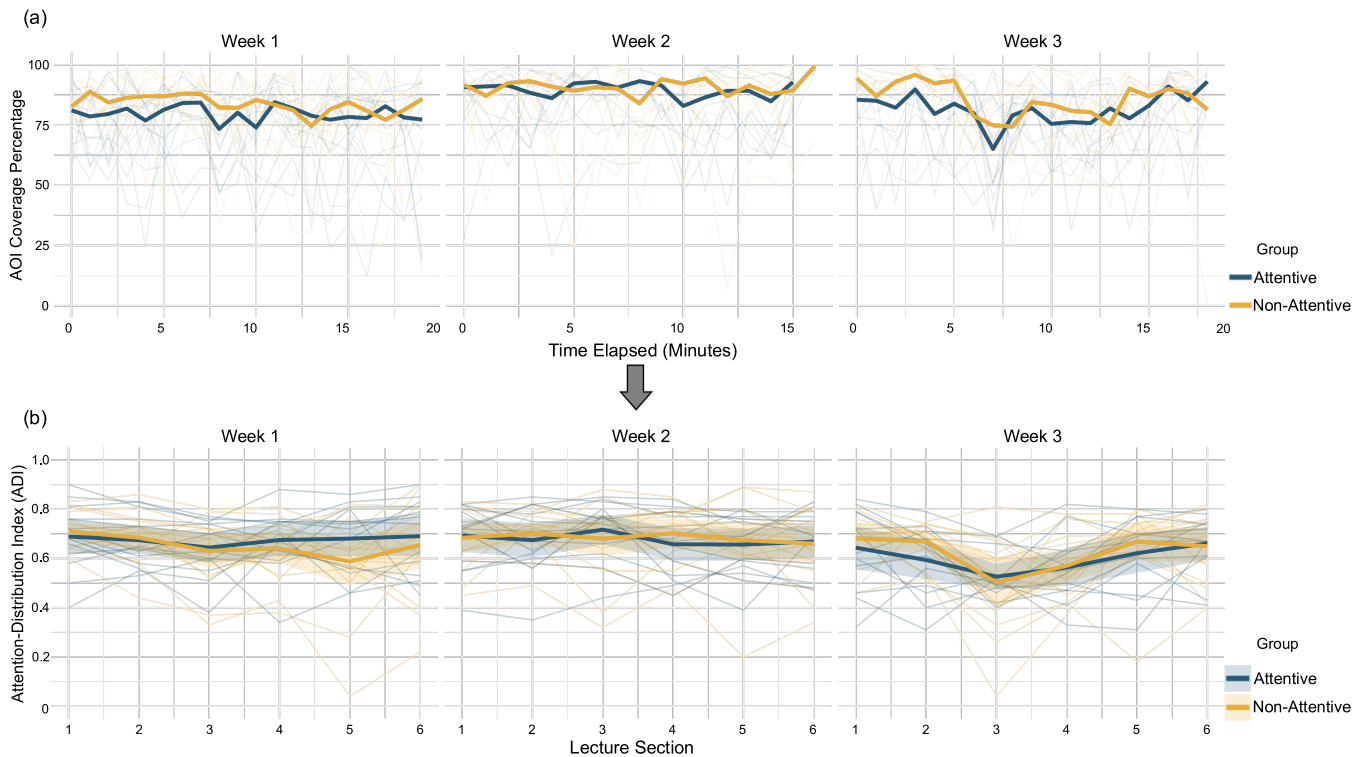
We conducted a mixed ANOVA on the ADI metric (ranging 0–1; see Figure 6b), calculated at the lecture-section level based on AOI

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* ( $BF_{incl} = 6.20$ ). Post-hoc pairwise comparisons (Holm-adjusted) localized the effect to a difference between week 2 and week 3 ( $MD = 0.074, SE = 0.023; t(26) = 3.189, p_{holm} = .011; d = 0.728$ ), 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* ( $F(1, 26) = 0.177, p = .677; BF_{incl} = 0.185$ ) or for a *Week*  $\times$  *Group* interaction ( $F(2, 52) = 0.523, p = .596; BF_{incl} = 0.160$ ).

*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:



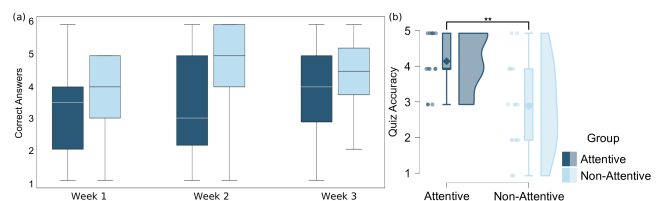
**Figure 6: (a) In-lecture focused percentage over time (minutes) across three weeks for attentive and non-attentive groups; (b) Attention Distribution Index (ADI) across lecture sections for both groups.**

P4, P7), echoing prior work on the benefits of immersive virtual classroom [25]. 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 [77], as well as the observed ADI decrease (Figure 6b). This pattern validates existing findings that attention management remains a challenge in immersive virtual classrooms [27].

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.

## 5.2 Attention-Aware LectureQuiz



**Figure 7: (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 7, 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.

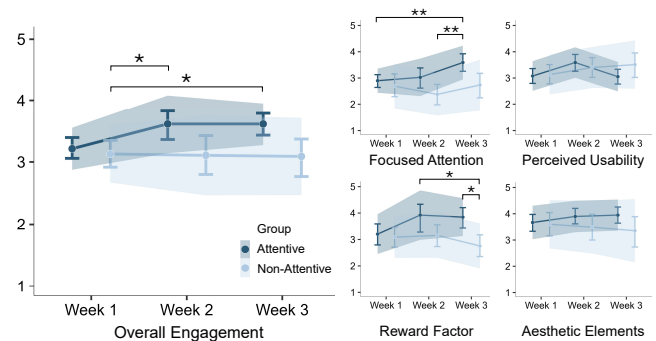
*Qualitative Results.* Thematic analysis provided further insights into participants' experiences with *LectureQuiz*. We generated **Theme 2: Effective Attention-Aware Personalization of LectureQuiz** from 12 codes related to quiz accuracy, timing, and learner perceptions.

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 [27], 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.**

### 5.3 User Engagement



**Figure 8: 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. [53]. 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 *Lecture-Quiz* 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:

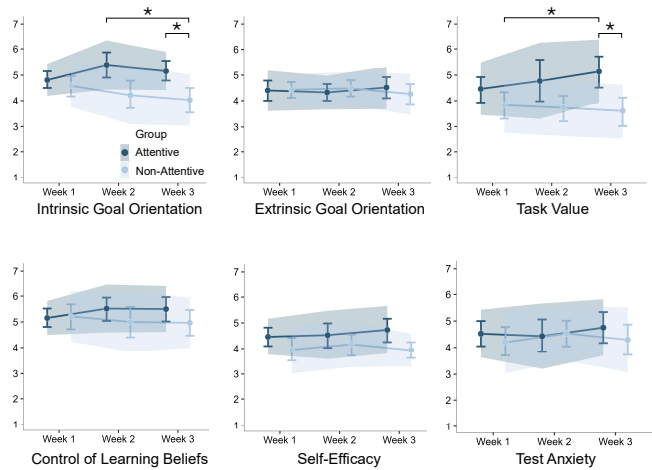
After the lecture, I went home and thought about where I didn’t understand. Then I would chat with it (*OpenChat*) and solve some questions day by day. Over the weekend, when I did the *Mini-Exam*, I felt very confident and it was satisfying to wrap up the week.

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 5.6) 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.**

## 5.4 Learning Motivation



**Figure 9: 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 (\*  $p < 0.05$ ).**

To assess changes in learning motivation, a mixed ANOVA was conducted across the six MSLQ dimensions. Following Pintrich et al. [58], we examined each dimension separately rather than calculating an overall score. Three dimensions showed significant between-group differences: **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 ( $F(2, 52) = 0.752$ ,  $p = 0.477$ ,  $\omega^2 = 0.000$ ;  $BF_{\text{incl}} = 0.096$ ).

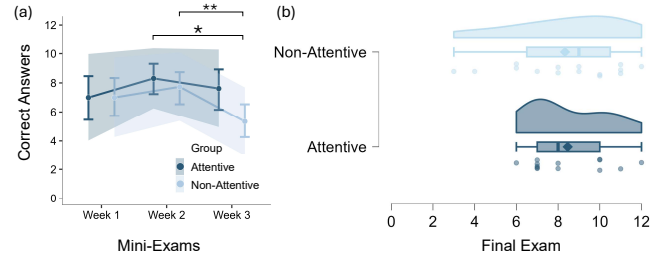
**Task Value.** For task value, there was a significant between-subject effect of *Group*,  $F(1, 26) = 9.586$ ,  $p = 0.005$ ,  $\omega^2 = 0.137$ ,  $p_{\text{Holm}} = 0.025$ , with higher scores in the attentive group ( $M = 4.79$ ,  $SD = 1.25$ ) compared to the non-attentive group ( $M = 3.70$ ,  $SD = 1.05$ ). 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.**



**Figure 10: 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.**

## 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 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.**

## 5.6 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 11b). For *ChatQuiz* scores (Figure 11a), 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*:  $F(2, 58) = 2.17, p = 0.13$ ).

**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.**

## 6 Discussion

### 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 [27], 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.**

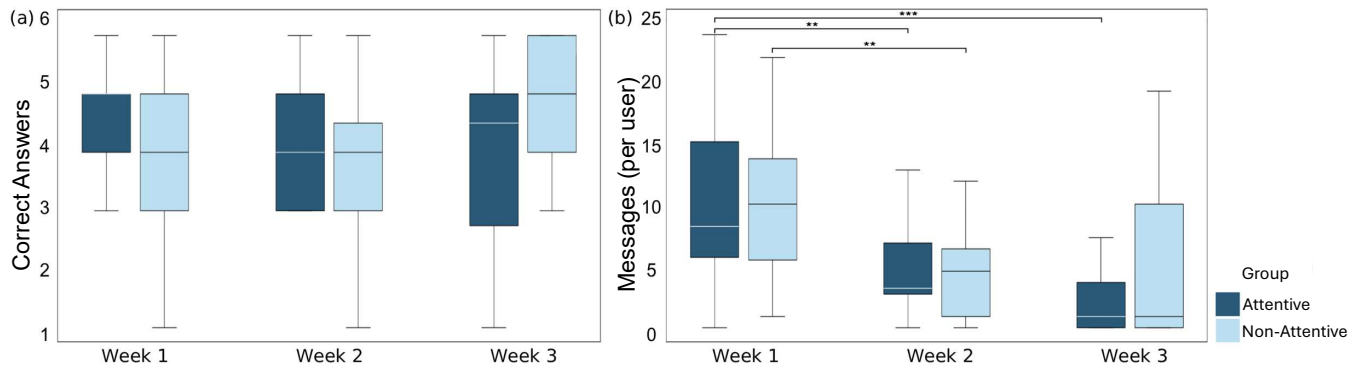
**Motivation and Engagement.** Quantitative measures indicate that *AttentiveLearn* supported higher levels of engagement and motivation, particularly in the motivational dimensions of intrinsic goal orientation, task value, and self-efficacy, as well as the engagement factors of reward and focused attention. 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. [47], 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 [58]. 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 [65, 78].

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



**Figure 11: 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 (\*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).**

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 [46].

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 [3, 14], 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 [57]. **Therefore, *AttentiveLearn* enhanced short-term comprehension and helped sustained progress, but its long-term impact was constrained and requires further investigation.**

## 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 [40].

Second, the lecture length in our study was around 20 minutes. Although this aligns with existing design recommendations [28], 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 *LectureQuiz*. Between-subject comparisons of every design features were beyond the scope of this four-week field study. Future research should examine how different support modalities 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 [33], this assumption does not always hold. As noted in prior work [73], 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, Szaafir and Mutlu [73] 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.

### 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 [17, 48], 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 [36, 73], 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 [14] 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., [11]) 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 [6]. 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 [80], may also serve as valuable input for personalized support. Prior work has demonstrated that biosignal-driven personalization can enhance user experience in diverse contexts [69]. 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 [62], skill-training scenarios with simulation [86], 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 [42].

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 [64]. 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 [44]. Other aspects such as memorization, comprehension, or decision-making may be supported through complementary techniques like spaced-repetition flashcards or gamification [84]. 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 [7, 42, 60, 61, 73], 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 [83]. 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, as in our field study. 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 [8, 73]. 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 [27], 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.

## 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.

## Acknowledgments

We gratefully acknowledge funding by the German Federal Ministry of Research, Technology and Space (BMFTR) for the ABBA project (Grant Number: 16DHBKI004). We also thank Jan Decker for his assistance with the system implementation.

## References

- [1] Yasmeen Abdrabou, Süleyman Özdel, Virmarie Maquiling, Efe Bozkir, and Enkelejda Kasneci. 2025. From Gaze to Data: Privacy and Societal Challenges of Using Eye-tracking Data to Inform GenAI Models. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3715669.3726788
- [2] Yasasi Abeyasinghe, Kevin Cauchi, Vikas Ashok, and Sampath Jayarathna. 2025. Framework for Measuring Visual Attention in Gaze-Driven VR Learning Environments Using Meta Quest Pro. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*. ACM, Tokyo Japan, 1–3. doi:10.1145/3715669.3726795
- [3] Huda S. Alazmi and Ghader 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* 29, 16 (Nov. 2024), 22189–22211. doi:10.1007/s10639-024-12682-3
- [4] Pengcheng An, Kenneth Holstein, Bernice d'Anjou, Berry Eggen, and Saskia Bakker. 2020. The TA Framework: Designing Real-time Teaching Augmentation for K-12 Classrooms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3313831.3376277
- [5] Thushari Atapattu, Katrina Falkner, Menasha Thilakarathne, Lavendini Sivaneasharajah, and Rangana Jayashanka. 2019. An Identification of Learners' Confusion through Language and Discourse Analysis. doi:10.48550/arXiv.1903.03286 arXiv:1903.03286 [cs].
- [6] Jonas Auda, Uwe Gruenefeld, Sarah Faltaous, Sven Mayer, and Stefan Schneegass. 2023. A Scoping Survey on Cross-reality Systems. *ACM Comput. Surv.* 56, 4 (2023), 83:1–83:38. doi:10.1145/3616536
- [7] Sarune Baceviciute, Aske Mottelson, Thomas Terkildsen, and Guido Makransky. 2020. Investigating Representation of Text and Audio in Educational VR using Learning Outcomes and EEG. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376872
- [8] Friederike Blume, Richard Göllner, Korbinian Moeller, Thomas Dresler, Ann-Christine Ehlis, and Caterina 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* 61 (June 2019), 138–147. doi:10.1016/j.learninstruc.2018.10.004
- [9] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. American Psychological Association, Washington, DC, US, 57–71. doi:10.1037/13620-004
- [10] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (Aug. 2019), 589–597. doi:10.1080/2159676X.2019.1628806 Publisher: Routledge \_eprint: https://doi.org/10.1080/2159676X.2019.1628806
- [11] Frederik Brudy, Christian Holz, Roman Rädle, Chi-Jui Wu, Steven Houben, Clemens Nylandsted Klokmose, and Nicolai Marquardt. 2019. Cross-Device Taxonomy: Survey, Opportunities and Challenges of Interactions Spanning Across Multiple Devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–28. doi:10.1145/3290605.3300792
- [12] Andreas Bulling. 2016. Pervasive Attentive User Interfaces. *Computer* 49, 1 (Jan. 2016), 94–98. doi:10.1109/MC.2016.32
- [13] Georg Buscher, Andreas Dengel, Raf Biedert, and Ludger V. Elst. 2012. Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Trans. Interact. Intell. Syst.* 1, 2, Article 9 (Jan. 2012), 30 pages. doi:10.1145/2070719.2070722
- [14] Kun-Hung Cheng and Chin-Chung Tsai. 2019. A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers & Education* 140 (Oct. 2019), 103600. doi:10.1016/j.compedu.2019.103600
- [15] Marvin Chun and Jeremy Wolfe. 2008. Visual Attention. In *Blackwell's Handbook of Perception*. Wiley-Blackwell, Hoboken, New Jersey, 272–310. doi:10.1002/9780470753477.ch9 Journal Abbreviation: Blackwell's Handbook of Perception.
- [16] Monica F. Contrino, Maribell Reyes-Millán, Patricia Vázquez-Villegas, and Jorge 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* 11, 1 (Feb. 2024), 6. doi:10.1186/s40561-024-00292-y
- [17] Chris Dede. 2011. Emerging Technologies, Ubiquitous Learning, and Educational Transformation. In *Towards Ubiquitous Learning*, Carlos Delgado Kloos, Denis Gillet, Raquel M. Crespo García, Fridolin Wild, and Martin Wolpers (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1–8.
- [18] Carrie Demmans Epp, Ben K. Daniel, and Kasia Muldner. 2023. Editorial: Learning analytics for supporting individualization: data-informed adaptation of learning. *Frontiers in Education* 8 (June 2023). doi:10.3389/feduc.2023.1240377 Publisher: Frontiers.
- [19] Sidney D'mello and Art Graesser. 2012. AutoTutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems* 2, 4 (Dec. 2012), 1–39. doi:10.1145/2395123.2395128
- [20] Sidney D'Mello, Andrew Olney, Claire Williams, and Patrick Hays. 2012. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies* 70, 5 (May 2012), 377–398. doi:10.1016/j.ijhcs.2012.01.004
- [21] Eileen du Plooy, Daleen Casteleijn, and Denise Franzsen. 2024. Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement. *Heliyon* 10, 21 (Oct. 2024), e39630. doi:10.1016/j.heliyon.2024.e39630
- [22] Andrew Duchowski. 2007. *Eye Tracking Methodology: Theory and Practice*. Springer, London. doi:10.1007/978-1-84628-609-4 Journal Abbreviation: Eye Tracking Methodology: Theory and Practice Publication Title: Eye Tracking Methodology: Theory and Practice.
- [23] John Dunlosky, Katherine A. Rawson, Elizabeth J. Marsh, Mitchell J. Nathan, and Daniel T. Willingham. 2013. Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest* 14, 1 (Jan. 2013), 4–58. doi:10.1177/1529100612453266
- [24] Sithara H. P. W. Gamage, Jennifer R. Ayres, Monica B. Behrend, and Elizabeth J. Smith. 2019. Optimising Moodle quizzes for online assessments. *International Journal of STEM Education* 6, 1 (Aug. 2019), 27. doi:10.1186/s40594-019-0181-4
- [25] Hong Gao, Efe Bozkir, Lisa Hasenbein, Jens-Uwe Hahn, Richard Göllner, and Enkelejda Kasneci. 2021. Digital Transformations of Classrooms in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3411764.3445596
- [26] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 1995. *Bayesian Data Analysis*. Chapman and Hall/CRC, New York. doi:10.1201/9780429258411
- [27] Yu Han, Yu Miao, Jie Lu, Mei Guo, and Yi Xiao. 2022. Exploring Intervention Strategies for Distracted Students in VR Classrooms. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3491101.3519627
- [28] Meta Horizon. 2024. Building Content to Optimize User Time: 20-40 Minute "Goldilocks" Session Length on Meta Quest. <https://developers.meta.com/horizon/blog/optimize-user-time-goldilocks-session-length-meta-quest/>.
- [29] Stephen Hutt, Kristina Krasich, Caitlin Mills, Nigel Bosch, Shelby White, James R. Brockmole, and Sidney K. D'Mello. 2019. Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction* 29, 4 (Sept. 2019), 821–867. doi:10.1007/s11257-019-09228-5
- [30] Stephen Hutt, Caitlin Mills, Nigel Bosch, Kristina Krasich, James Brockmole, and Sidney D'Mello. 2017. "Out of the Fr-Eye-ing Pan": Towards Gaze-Based Models of Attention during Learning with Technology in the Classroom. In *Proceedings*

- of the 25th Conference on User Modeling, Adaptation and Personalization. ACM, Bratislava Slovakia, 94–103. doi:10.1145/3079628.3079669
- [31] Gwo-Jen Hwang and Chin-Chung 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* 42, 4 (July 2011), E65–E70. doi:10.1111/j.1467-8535.2011.01183.x
- [32] H. Jarodzka and S. Brand-Gruwel. 2017. Tracking the reading eye: towards a model of real-world reading. *Journal of Computer Assisted Learning* 33, 3 (2017), 193–201. doi:10.1111/jcal.12189 \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcal.12189>
- [33] Marcel A. Just and Patricia A. Carpenter. 1980. A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review* 87, 4 (1980), 329–354. doi:10.1037/0033-295X.87.4.329
- [34] Arielle S. Keller, Ido Davidesco, and Kimberly D. Tanner. 2020. Attention Matters: How Orchestrating Attention May Relate to Classroom Learning. *CBE—Life Sciences Education* 19, 3 (Sept. 2020), fe5. doi:10.1187/cbe.20-05-0106 Publisher: American Society for Cell Biology (lse).
- [35] Anam Ahmad Khan. 2019. Gaze assisted voice note taking system. In *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*. ACM, London United Kingdom, 367–371. doi:10.1145/3341162.3349308
- [36] Chris Klavereen, Sebastiaan Vonk, and Ilja Cornelisz. 2017. The Effect of Adaptive versus Static Practicing on Student Learning - Evidence from a Randomized Field Experiment. *Economics of Education Review* 58 (April 2017), 175–187. doi:10.1016/j.econedurev.2017.04.003
- [37] George D. Kuh. 1996. Guiding Principles for Creating Seamless Learning Environments for Undergraduates. *Journal of College Student Development* 37, 2 (1996), 135–48. ERIC Number: EJ527218.
- [38] Moritz Langner, Peyman Toreini, and Alexander Maedche. 2023. Leveraging Eye Tracking Technology for a Situation-Aware Writing Assistant. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications (ETRA '23)*. Association for Computing Machinery, New York, NY, USA, 1–2. doi:10.1145/3588015.3590121
- [39] Shi Liu, Alexander Mädche, and Martin Feick. 2025. GazeClass: Towards Gaze-Adaptive Cross-Device Learning Support for Virtual Classrooms. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3706599.3720232
- [40] Shi Liu, Peyman Toreini, and Alexander Maedche. 2025. AF-Mix: A gaze-aware learning system with attention feedback in mixed reality. *International Journal of Human-Computer Studies* 198 (April 2025), 103467. doi:10.1016/j.ijhcs.2025.103467
- [41] Ziyi Liu, Zhengzhe Zhu, Lijun Zhu, Enze Jiang, Xiyun Hu, Kylie A Peppler, and Karthik Ramani. 2024. ClassMeta: Designing Interactive Virtual Classmate to Promote VR Classroom Participation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3613904.3642947
- [42] Guido Makransky, Niels K. Andreasen, Sarune Baceviciute, and Richard 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* 113, 4 (2021), 719–735. doi:10.1037/edu0000473 Place: US Publisher: American Psychological Association.
- [43] Guido Makransky and Richard 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* 34, 3 (Sept. 2022), 1771–1798. doi:10.1007/s10648-022-09675-4
- [44] Guido Makransky and Gustav 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* 33, 3 (Sept. 2021), 937–958. doi:10.1007/s10648-020-09586-2
- [45] Richard E. Mayer. 2014. Cognitive theory of multimedia learning. In *The Cambridge handbook of multimedia learning, 2nd ed.* Cambridge University Press, New York, NY, US, 43–71. doi:10.1017/CBO9781139547369.005
- [46] Mark A. McDaniel, Pooja K. Agarwal, Barbie J. Huelser, Kathleen B. McDermott, and Henry L. Roediger. 2011. Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology* 103, 2 (May 2011), 399–414. doi:10.1037/a0021782
- [47] Mark A. McDaniel, Kathleen M. Wildman, and Janis 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* 1, 1 (March 2012), 18–26. doi:10.1016/j.jarmac.2011.10.001
- [48] Marcelo Milrad, Lung-Hsiang Wong, Mike Sharples, Gwo-Jen Hwang, Chee-Kit Looi, and Hiroaki Ogata. 2013. Seamless Learning: An International Perspective on Next-Generation Technology-Enhanced Learning. In *Handbook of Mobile Learning* (1 ed.), Zane L. Berge and Lin Muilenburg (Eds.). Routledge, New York, 133–146. doi:10.4324/9780203118764-16
- [49] Brian Moynihan. 2021. Immersive Connections - Teaching, Learning, and Creating in VR at Ithaca College. <https://news.lenovo.com/immersive-connections-teaching-learning-creating-vr-ithaca-college/>
- [50] Stylianos Mystakidis and Vangelis Lymouridis. 2023. Immersive Learning. *Encyclopedia* 3, 2 (June 2023), 396–405. doi:10.3390/encyclopedia3020026 Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.
- [51] Krenare Pireva Nuci, Rabail Tahir, Alf Inge Wang, and Ali Shariq Imran. 2021. Game-Based Digital Quiz as a Tool for Improving Students' Engagement and Learning in Online Lectures. *IEEE Access* 9 (2021), 91220–91234. doi:10.1109/ACCESS.2021.3088583
- [52] Jason Oehlberg. 2018. VR Classroom. <https://github.com/JasonOehlberg/vr-classroom> original-date: 2018-10-25T16:52:36Z.
- [53] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (April 2018), 28–39. doi:10.1016/j.ijhcs.2018.01.004
- [54] Fabrizio Palmas, Jakub Cichor, David A. Plecher, and Gudrun Klinker. 2019. Acceptance and Effectiveness of a Virtual Reality Public Speaking Training. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 363–371. doi:10.1109/ISMAR.2019.00034
- [55] Fabrizio Palmas, Ramona Reinelt, Jakub E. Cichor, David A. Plecher, and Gudrun Klinker. 2021. Virtual Reality Public Speaking Training: Experimental Evaluation of Direct Feedback Technology Acceptance. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 463–472. doi:10.1109/VR50410.2021.00070
- [56] Shannon Palmer, Youn Chu, and Adam M. Persky. 2019. Comparison of Re-watching Class Recordings versus Retrieval Practice as Post-Lecture Learning Strategies. *American Journal of Pharmaceutical Education* 83, 9 (Nov. 2019), 7217. doi:10.5688/ajpe7217
- [57] Gustav Bog Petersen, Aske Mottelson, and Guido Makransky. 2021. Pedagogical Agents in Educational VR: An in the Wild Study. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3411764.3445760
- [58] P Pintrich, D. Smith, Teresa Duncan, and Wilbert Mckeachie. 1991. A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ). *Ann Arbor. Michigan* 48109 (Jan. 1991), 1259.
- [59] Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2022. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-worn Extended Reality. *ACM Comput. Surv.* 55, 3, Article 53 (March 2022), 39 pages. doi:10.1145/3491207
- [60] Jaziar Radianti, Tim A. Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. 2020. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education* 147 (April 2020), 103778. doi:10.1016/j.compedu.2019.103778
- [61] Iulian Radu, Vivek Hv, and Bertrand Schneider. 2021. Unequal Impacts of Augmented Reality on Learning and Collaboration During Robot Programming with Peers. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 245:1–245:23. doi:10.1145/3432944
- [62] Iulian Radu, Tugce Joy, and Bertrand Schneider. 2021. Virtual Makerspaces: Merging AR/VR/MR to Enable Remote Collaborations in Physical Maker Activities. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–5. doi:10.1145/3411763.3451561
- [63] Iulian Radu and Bertrand Schneider. 2019. What Can We Learn from Augmented Reality (AR)? Benefits and Drawbacks of AR for Inquiry-based Learning of Physics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300774
- [64] Annelies Raes, Pieter Vanneste, Marieke Pieters, Ine Windey, Wim Van Den Noortgate, and Fien Depaepe. 2020. Learning and instruction in the hybrid virtual classroom: An investigation of students' engagement and the effect of quizzes. *Computers & Education* 143 (Jan. 2020), 103682. doi:10.1016/j.compedu.2019.103682
- [65] Claudia Roda and Julie Thomas. 2006. Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior* 22, 4 (July 2006), 557–587. doi:10.1016/j.chb.2005.12.005
- [66] Henry L. Roediger III and Jeffrey D. Karpicke. 2006. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science* 17, 3 (2006), 249–255. doi:10.1111/j.1467-9280.2006.01693.x Place: United Kingdom Publisher: Blackwell Publishing.
- [67] Bella Ross, Anne-Marie Chase, Diane Robbie, Grainne Oates, and Yvette 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* 15, 1 (Aug. 2018), 30. doi:10.1186/s41239-018-0113-2
- [68] Benjamin Lukas Schnitzer, Stefano Gampe, Arndt Bieberstein, Isabel Hoffmann, Daniel Görlich, and Oliver Korn. 2025. Language Learning in Virtual Reality: Enhancement of Long-Term Vocabulary Recognition and Understanding through Full-Body Avatars. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3706599.3720187
- [69] Tanja Schultz and Alexander Maedche. 2023. Biosignals meet Adaptive Systems. *SN Applied Sciences* 5, 9 (Aug. 2023), 234. doi:10.1007/s42452-023-05412-w

- [70] Qijia Shao, Amy Sniffen, Julien Blanchet, Megan E Hillis, Xinyu Shi, Themistoklis K Haris, Jason Liu, Jason Lambertson, Melissa Malzkahn, Lorna C Quandt, James Mahoney, David J M Kraemer, Xia Zhou, and Devin Balkcom. 2020. Teaching American Sign Language in Mixed Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4 (2020), 152:1–27. doi:10.1145/3432211
- [71] Kshitij Sharma, Michail Giannakos, and Pierre Dillenbourg. 2020. Eye-Tracking and Artificial Intelligence to Enhance Motivation and Learning. *Smart Learning Environments* 7, 1 (April 2020), 13. doi:10.1186/s40561-020-00122-x
- [72] Songjia Shen, Chek Tien Tan, Hsiang-Ting Chen, William L Raffae, and Tuck Wah Leong. 2025. Educator Perceptions of XRAuthor: An Accessible Tool for Authoring Learning Content with Different Immersion Levels. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3706598.3713982
- [73] Daniel Szafrir and Bilge Mutlu. 2013. ARTful: Adaptive Review Technology for Flipped Learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 1001–1010. doi:10.1145/2470654.2466128
- [74] Yiliu Tang, Jason Situ, Andrea Yaoyun Cui, Mengke Wu, and Yun Huang. 2025. LLM Integration in Extended Reality: A Comprehensive Review of Current Trends, Challenges, and Future Perspectives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–24. doi:10.1145/3706598.3714224
- [75] Santawat Thanyadit, Matthias Heintz, and Effie L-C Law. 2023. Tutor In-sight: Guiding and Visualizing Students' Attention with Mixed Reality Avatar Presentation Tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3544548.3581069
- [76] Hsin-Ruey Tsai, Shih-Kang Chiu, and Bryan Wang. 2025. GazeNoter: Co-Piloted AR Note-Taking via Gaze Selection of LLM Suggestions to Match Users' Intentions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3706598.3714294
- [77] Aki Vehtari. 2024. Bayesian Data Analysis Course. [https://avehtari.github.io/BDA\\_course\\_Aalto/](https://avehtari.github.io/BDA_course_Aalto/)
- [78] Roel Vertegaal, Jeffrey S. Shell, Daniel Chen, and Aadil Mamuji. 2006. Designing for augmented attention: Towards a framework for attentive user interfaces. *Computers in Human Behavior* 22, 4 (July 2006), 771–789. doi:10.1016/j.chb.2005.12.012
- [79] Thiemo Wambsganss, Tobias Kueng, Matthias Soellner, and Jan Marco Leimeister. 2021. ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3411764.3445781
- [80] Thiemo Wambsganss, Matthias Soellner, Kenneth R Koedinger, and Jan Marco Leimeister. 2022. Adaptive Empathy Learning Support in Peer Review Scenarios. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3491102.3517740 event-place: New Orleans, LA, USA.
- [81] Songlin Xu, Dongyin Hu, Ru Wang, and Xinyu Zhang. 2025. PeerEdu: Bootstrapping Online Learning Behaviors via Asynchronous Area of Interest Sharing from Peer Gaze. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 656, 14 pages. doi:10.1145/3706598.3713480
- [82] Thorsten O. Zander, Christian Kothe, Sabine Jatzew, and Matti Gaertner. 2010. Enhancing Human-Computer Interaction with Input from Active and Passive Brain-Computer Interfaces. In *Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction*, Desney S. Tan and Anton Nijholt (Eds.). Springer, London, 181–199. doi:10.1007/978-1-84996-272-8\_11
- [83] Xinli Zhang, Yuchen Chen, Lailin Hu, and Youmei Wang. 2022. The metaverse in education: Definition, framework, features, potential applications, challenges, and future research topics. *Frontiers in Psychology* 13 (Oct. 2022), 1–18. doi:10.3389/fpsyg.2022.1016300 Publisher: Frontiers.
- [84] Yuchong Zhang, Adam Nowak, Andrzej Romanowski, and Morten Fjeld. 2023. Virtuality or Physicality? Supporting Memorization Through Augmented Reality Gamification. In *Companion Proceedings of the 2023 ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 53–58. doi:10.1145/3596454.3597183
- [85] Fengyuan Zhu, Mauricio Sousa, Ludwig Sidenmark, and Tovi Grossman. 2024. PhoneInVR: An Evaluation of Spatial Anchoring and Interaction Techniques for Smartphone Usage in Virtual Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. doi:10.1145/3613904.3642582
- [86] Xiuqi Tommy Zhu, Heidi Cheerman, Minxin Cheng, Sheri R Kiami, Leanne Chukoskie, and Eileen McGivney. 2025. Designing VR Simulation System for Clinical Communication Training with LLMs-Based Embodied Conversational Agents. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3706599.3719693