Master of Educational Technology, Saarland University

Design and Development of an AI Chatbot for Automatic Question Generation for Self-Assessment in Scientific Text Comprehension

(A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Educational Technology)

Kotte Rambukkana Thanthri Mudalige Don Piumal Migara Mahawasala

Student Number: 7036396

11.08.2025

Supervised by:

Dominik Thüs

PD. Dr. Sarah Malone

Acknowledgements

I would like to express my heartfelt thanks to those who supported me during this journey. To my family and friends, your encouragement, love, and belief in me provided the strength I needed to move forward during challenging moments. I am especially grateful for my partner, whose presence, patience, and warmth brought balance and inspiration throughout this process. Your support meant everything.

To my fellow MSc in Educational Technology students, thank you for the motivation, insights, and shared experiences that made this journey meaningful. I also extend my appreciation to all participants who contributed to this research with their time and input.

Finally, I would like to express my sincere gratitude to my supervisors, Mr. Dominik Thüs and Ms. Sarah Malone, for their guidance, thoughtful feedback, and the opportunity to be involved in the OwlMentor project. Their mentorship helped shape this thesis and my growth as a researcher.

Abstract

This study investigates the role of an AI-based question generation chatbot, *OwlMentor*, in supporting scientific reading comprehension and self-assessment among master’s students in Educational Technology. The primary objectives were to examine (a) how students interact with the chatbot and the types of questions generated, (b) the chatbot’s impact on engagement and learning outcomes, and (c) how learner motivation and self-efficacy influence, and are influenced by, its use. A mixed-methods design was employed with 14 participants over one semester, integrating pre-/post-questionnaires, weekly engagement surveys, comprehension tests, and system usage logs. Quantitative analyses included descriptive statistics, correlations, and paired-samples *t* tests, while qualitative analyses classified student-generated questions by type and Bloom’s taxonomy level. Results indicated that students primarily generated multiple-choice and true/false questions, with most questions targeting mid-level cognitive processes (“Understand” and “Apply”). While learning gains were modest, engagement was positively related to perceived usefulness and recall performance. Motivation and self-efficacy levels remained stable over time, and higher baseline self-efficacy was associated with lower chatbot usage. The findings suggest that Automatic question generation (AQG) can facilitate active reading and metacognitive engagement, though its effectiveness may depend on learner characteristics. Implications for instructional design, limitations, and directions for future research are discussed.

Table of Contents

Introduction

Scientific reading comprehension is crucial in higher education, especially within disciplines that require critical engagement with complex academic and scientific literature, such as technology and educational research. However, students frequently approach academic texts passively, facing challenges in effectively applying metacognitive strategies, monitoring their understanding, and engaging in reflective practices (Snow, 2002). Recent studies, including Kasneci et al. (2023) and Jansen et al. (2024), reinforce these concerns by showing that learners often struggle with processing dense scientific texts without adequate scaffolding. Without structured support or timely feedback, such difficulties can significantly impede effective learning and comprehension.

Advancements in artificial intelligence (AI), particularly large language models (LLMs) like ChatGPT, offer considerable potential for supporting self-directed learning and enhancing reading comprehension. Educational chatbots powered by these models provide continuous, personalized interactions, immediate feedback, and real-time support, enabling students to actively engage with material through question generation, content summarization, and reflective dialogue (Holmes, Bialik, & Fadel, 2019; Winkler & Söllner, 2018). OwlMentor, in particular, is based on a Retrieval-Augmented Generation (RAG) architecture, which grounds AI responses in course-relevant materials and thereby helps reduce hallucinations and improve factual accuracy. These interactive AI systems effectively support the complex cognitive strategies essential for deeper learning.

Student-generated questioning is an established method for improving reading comprehension and retention. By formulating their own questions, students actively engage with learning content, fostering deeper cognitive processing and improved recall (King, 1992; Rosenshine, Meister, & Chapman, 1996). Despite these benefits, students often encounter difficulties in independently generating high-quality questions due to cognitive overload and limited prior knowledge (Yu, 2009; Kurdi et al., 2019). AQG systems thus serve as valuable scaffolding tools, enhancing both question quality and student engagement (Lee et al., 2023; Mulla et al., 2023; Bhat et al., 2022; Dijkstra et al., 2022).

The effectiveness of AI tools in education depends not only on their technological capabilities but also significantly on learner-specific factors such as self-efficacy and motivation. Bandura (1997) highlights self-efficacy, students' beliefs in their capability to succeed, as a crucial determinant of their learning behavior, persistence, and use of external support. Similarly, the Technology Acceptance Model (TAM) emphasizes perceived usefulness, perceived ease of use, and behavioral intention as key predictors of technology adoption and actual usage (Davis, 1989; Venkatesh & Davis, 2000). Recent studies indicate that motivation towards AI positively correlates with engagement, while high self-efficacy might reduce the perceived necessity for external support (Thüs et al., 2024).

Previous research, based on the earlier version of OwlMentor (Thüs et al., 2024), has explored AI chatbots' usability and question-generation capabilities; however, comprehensive research examining the interactions among motivation, self-efficacy, technology acceptance, and learning outcomes remains limited. This thesis aims to address this gap by empirically investigating OwlMentor, a retrieval-augmented generation (RAG)-based AI chatbot designed to support students in generating self-assessment questions from scientific texts. Utilizing a mixed-method approach, including qualitative (interaction logs, question quality analyses) and quantitative (pre/post comprehension tests, structured questionnaire data) methodologies, the study incorporates validated instruments such as the Technology Acceptance Model (Davis, 1989), AI Motivation (Yurt & Kasarci, 2024), and self-efficacy scales developed by Pintrich & De Groot (1990) and Koşar et al. (2022). The study seeks to answer following research questions (RQs) :

RQ1**:** How do students interact with the OwlMentor?

- RQ1.1: What types of questions do students generate using the chatbot (e.g., true/false, short answer, open-ended)?
- RQ1.2: How does the chatbot affect students’ engagement with scientific texts?

RQ2: What learning outcomes and cognitive processes are supported by the AI chatbot?

- RQ2.1: Do students benefit from using the chatbot in terms of learning outcomes?

RQ3: How does learner motivation and self-efficacy influence (and get influenced by) the use of the AI chatbot?

- RQ3.1: How does students' initial motivation or self-efficacy affect their engagement with the chatbot?
- RQ3.2: Does motivation or self-efficacy change over time through the use of the chatbot?

By exploring these areas, this research seeks to clarify the role and impact of AI-powered chatbots in supporting scientific reading comprehension and to identify key factors influencing their successful educational integration.

This thesis is organized into five chapters. Chapter 1 introduces the background, rationale, and research questions. Chapter 2 reviews the relevant literature on AI in education, student-generated questioning, motivation, self-efficacy, and technology acceptance, and concludes with the conceptual framework and research gap. Chapter 3 details the methodology, including participants, instruments, procedures, and data analysis methods. Chapter 4 presents the results for each research question. Chapter 5 discusses the findings in relation to existing literature, outlines practical and theoretical implications, notes study limitations, and suggests directions for future research.

Literature Review

2.1 AI in Education and Chatbots

The past decade has seen the proliferation of AI-powered tools in education, from intelligent tutoring systems to automated graders, with growing interest in their capacity to deliver personalized, on-demand support (Winkler & Söllner, 2018). Educational chatbots leverage natural language understanding to engage learners in conversational interactions, offering hints, explanations, and formative assessments at scale (Holmes, Bialik, & Fadel, 2019). Studies demonstrate that chatbots can increase learners' engagement and persistence by providing immediate feedback and 24/7 availability, while offloading routine queries from instructors (Kerly, Ellis, & Bull, 2008). More recent systems built on large language models (LLMs) such as GPT-4 have further improved the fluency and relevance of chatbot responses, though they raise concerns of factual inaccuracy without proper grounding (Thüs et al., 2024).

2.2 Student-Generated Questioning and Self-Evaluation

Student-Generated Questioning (SQG) is a metacognitive strategy whereby learners formulate questions about material they are studying. King (1992) showed that prompting students to write “why” and “how” questions can significantly improve comprehension and retention relative to more passive study. By generating questions, students engage in deeper processing, move beyond surface features, and monitor their own understanding (Yu, 2009). However, unguided SQG often yields low-quality or overly factual questions (Rosenshine, Meister, & Chapman, 1996). Scaffolding, such as question stems or peer review, helps learners construct more meaningful, inferential questions (Chin & Brown, 2000). AI-enabled scaffolds, including chatbots that suggest question prompts or reformulate student queries, have been found to further bolster both question quality and learner motivation (Kim, Park, & Cho, 2021).

2.3 Automatic Question Generation (AQG)

AQG refers to the process of automatically producing questions from a given source text, with the aim of supporting learning, self-assessment, and metacognitive engagement

(Kurdi et al., 2019). In educational contexts, AQG systems can scaffold learners' comprehension through quizzes, self-testing, and prompts that stimulate active reading and reflection. The design and implementation of AQG methods vary considerably, and understanding these methodological differences is crucial for evaluating their pedagogical value.

Template-Based AQG

Early AQG systems were predominantly template-based, using predefined syntactic or semantic patterns to transform text into questions (Heilman & Smith, 2010; Kurdi et al., 2019). While these systems were computationally efficient and easy to implement, their output was often limited to low-level factual recall questions (e.g., “Who…?”, “What…?”). They lacked the flexibility to generate higher-order cognitive questions and were sensitive to linguistic variation in source materials.

Neural Network–Based AQG

With the rise of deep learning, neural AQG approaches emerged, leveraging sequence-to-sequence (seq2seq) and transformer architectures (Zhou et al., 2017; Liu et al., 2020). These models can learn complex mappings between source texts and questions, enabling greater linguistic diversity and the potential to generate higher-order cognitive questions. However, purely neural models are prone to generating factually incorrect or contextually irrelevant questions, commonly referred to as “hallucinations”, if not adequately grounded in the source material (Lee et al., 2023).

Retrieval-Augmented Generation (RAG)–Based AQG

More recently, RAG-based AQG systems have combined the generative flexibility of large language models (LLMs) with a retrieval mechanism that fetches relevant source passages before question generation (Lewis et al., 2020; Siriwardhana et al., 2023). This architecture grounds the output in authoritative texts, thereby improving factual accuracy and reducing hallucinations. RAG approaches can also be adapted to align with educational objectives, such as generating Bloom's taxonomy–aligned questions or supporting specific learning outcomes (Oche et al., 2025).

Pedagogical Implications of AQG

AQG systems not only support factual recall but also promote deeper learning by encouraging students to engage with texts actively. This aligns with constructivist principles, where learners build knowledge through interaction and reflection (Bruner, 1966). Moreover, AQG can serve as a formative assessment tool, helping learners identify gaps in understanding and adjust their study strategies accordingly (Black & Wiliam, 1998). Recent developments such as SkillQG (Wang et al., 2023) demonstrate how AQG can be tailored to assess specific reading skills, enhancing its diagnostic value. Similarly, Bhat et al. (2022) propose automated evaluation metrics to assess question quality, which can help educators monitor student progress more effectively.

Limitations and Future Directions

Despite their promise, AQG systems face challenges in domain adaptation and contextual relevance. Siriwardhana et al. (2023) highlight that even RAG-based models may struggle with unfamiliar content unless fine-tuned. Future AQG tools should incorporate adaptive personalization to align question difficulty with learner proficiency and engagement levels. Additionally, integrating feedback loops and learner analytics could enhance the pedagogical utility of AQG systems in diverse educational settings.

OwlMentor’s Position in the AQG Landscape

OwlMentor exemplifies the latest wave of RAG-based AQG systems. Built on GPT-4 architecture with course-specific retrieval pipelines, it enables learner-centric question co-creation, students can iteratively refine AI-generated questions, receive immediate feedback, and store items for later use in practice quizzes. This interactive process not only supports knowledge retrieval but also fosters metacognitive monitoring, aligning with research showing that scaffolded question generation enhances comprehension and critical thinking (King, 1992; Oche et al., 2025).

2.4 Motivation to Use AI

Learner motivation is a critical determinant of technology uptake and learning persistence (Deci & Ryan, 2000). Motivation to use AI can be understood as learners’ intrinsic interest or utility beliefs regarding AI tools in academic contexts. Studies find that students who are curious about AI or believe it aligns with their goals are more likely to explore and sustain use of AI tutors (Türkyilmaz et al., 2023). Venkatesh and Bala (2008) incorporate motivational constructs such as “hedonic motivation” into extended technology acceptance models, showing that enjoyment in interacting with AI chatbots predicts continued use. Conversely, learners with high self-confidence in their own abilities may be less inclined to adopt support tools if they perceive them as unnecessary or time-consuming, a form of cost-benefit reasoning (Thüs et al., 2024).

2.5 Self-Efficacy in Learning

Self-efficacy, defined by Bandura (1997) as one’s belief in their capacity to execute actions required to achieve goals, is central to students’ approach to academic tasks. Schunk and Pajares (2002) report that high-efficacy learners select more challenging tasks, persist longer in the face of difficulty, and deploy more effective learning strategies. In the context of digital learning tools, learners’ confidence in using AI systems (digital self-efficacy) influences their willingness to experiment with novel interfaces and features (Hsu & Chiu, 2004). However, high general self-efficacy can paradoxically reduce engagement with scaffolds: learners may deem external support superfluous, especially if interacting with AI entails additional effort (Wang & Wang, 2019).

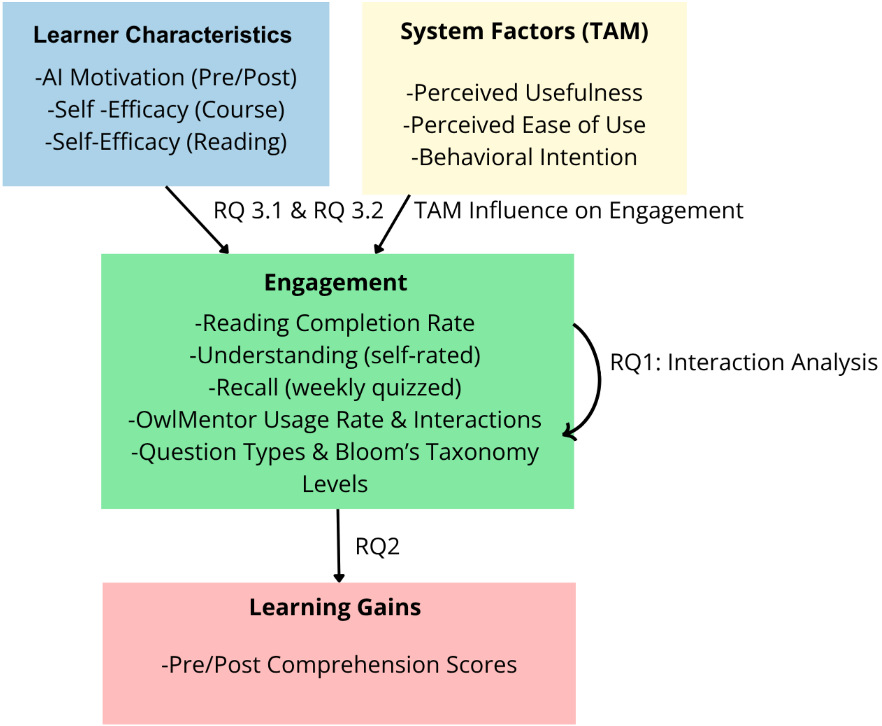
2.6 Technology Acceptance and AI Use

The Technology Acceptance Model (TAM; Davis, 1989; Davis, Bagozzi, & Warshaw, 1989) posits that Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) shape users’ Behavioral Intention (BI) to adopt a system, which in turn predicts Actual Use. In educational contexts, TAM has successfully explained the adoption of e-learning platforms (Wu, Tao, & Yang, 2008) and mobile learning apps (Park, 2009). Longitudinal studies suggest that the influence of PEOU diminishes over time as users gain experience, while PU remains a stable predictor of continued engagement (Venkatesh & Bala, 2008). Moreover, hedonic factors, the enjoyment or fun of interacting with “conversational AI”, can drive sustained use alongside utilitarian beliefs (Venkatesh, Thong, & Xu, 2012).

Drawing on the reviewed literature, Figure 2.1 presents the conceptual framework guiding this study. The model integrates learner characteristics (AI motivation and self-efficacy), system factors from the Technology Acceptance Model (perceived usefulness, perceived ease of use, and behavioral intention), and engagement indicators (reading completion, self-rated understanding, recall, usage rates, and question generation patterns). These constructs are hypothesized to interact in ways that influence learning gains in scientific reading comprehension.

Figure 1

Conceptual framework illustrating the relationships between learner characteristics, system factors, engagement, and learning gains in the OwlMentor study.



Note. Adapted by the author from Davis (1989), Venkatesh and Davis (2000), Bandura (1997), Pintrich and De Groot (1990), Koşar et al. (2022), Yurt and Kasarci (2024), Thüs et al. (2024), King (1992), Rosenshine et al. (1996), and Kasneci et al. (2023).

2.7 Research Gaps and Rationale

Despite evidence for chatbot efficacy in quizzes and FAQs, there remains limited research on AI-scaffolded SQG for reading comprehension in higher education. Few studies have linked AI-supported questioning directly to self-efficacy or intrinsic motivation measures, nor have they examined longitudinal TAM dynamics over a semester. Furthermore, while RAG-based systems promise reduced hallucinations, their pedagogical deployment and impact on metacognitive strategy use remain underexplored.

This study builds upon prior work by implementing and evaluating OwlMentor, a chatbot grounded in retrieval-augmented generation and designed to support students in co-creating questions from scientific texts. By exploring students' use of the chatbot in relation to motivation, self-efficacy, perceived usefulness, and comprehension gains, the study addresses a critical gap in understanding how AI can function as a collaborative learning partner in academic reading.

Methodology

3.1 Research Design

This study employed a mixed-methods research design to investigate how students engage with an AI-powered chatbot for question co-creation during scientific reading tasks and how various psychological and behavioral factors relate to their interaction patterns and learning outcomes. The integration of both quantitative and qualitative data sources allowed for a comprehensive understanding of not only what students did with the chatbot, but also why and with what effects.

The quantitative component focused on measuring changes in reading comprehension (via pre- and post-tests), as well as capturing constructs such as self-efficacy, AI motivation, and technology acceptance using validated self-report instruments. These instruments included:

- **General Academic Self-Efficacy** (Pintrich & De Groot, 1990),
- **Reading Self-Efficacy** (Koşar et al., 2022),
- **AI Motivation** (Yurt & Kasarci, 2024), and
- **Technology Acceptance Model (TAM)** constructs, including *Perceived Ease of Use*, *Perceived Usefulness*, and *Behavioral Intention to Use* (Davis, 1989; Venkatesh & Davis, 2000).

All constructs were assessed using multi-item Likert-scale questionnaires administered at the beginning and end of the semester. The questionnaires were provided in both English and German and had previously been tested for reliability and validity in similar educational technology contexts.

The qualitative component involved analysis of interaction logs and student-generated questions recorded during the voluntary use of the AI-based question generation chatbot, OwlMentor. These data were examined to identify usage patterns, engagement behavior, and the cognitive level of the questions generated. A rubric based on Bloom's Taxonomy was used to classify the quality and complexity of student-generated questions. By combining these data sources, the study aimed to answer the following:

- What types of questions are generated

- Whether interaction with the chatbot enhances learning
- How learner characteristics and perceptions influence usage and outcomes

This explanatory mixed-methods design provided both breadth (via surveys and test data) and depth (via log and content analysis), enabling triangulation of findings and a nuanced understanding of the pedagogical potential of AQG.

3.2 Participants and Context

The study was conducted with 14 master's level students enrolled in the Educational Technology (EduTech) program at Saarland University, Germany, during the winter semester of 2024/25. The group consisted of 11 female and 03 male students, aged between 25 and 45 years. All participants were international students from diverse disciplinary backgrounds, including education, computer science, and psychology. Each participant held at least a bachelor’s degree, ensuring a foundational level of academic experience and familiarity with scientific literature.

The participants were enrolled in the Multimedia Learning II course, part of a compulsory module in the EduTech program. Program content had 5 chapters from the book Multimedia Design Principles as below:

1) Split Attention Principle
2) Emotional Design Principle
3) Worked Example Principle
4) Expertise Revearsal Principle
5) Social Cues Principle

The course was designed to support critical engagement with scientific texts, requiring students to analyze empirical research, apply cognitive learning theories, and reflect on educational technologies. Many participants had previously completed the Multimedia Learning I course in the prior semester, during which a prototype of OwlMentor was introduced and piloted.

Participation in the research study was optional and not tied to students’ course grades. All students were informed about the purpose of the study and provided informed consent prior to data collection. No specific exclusion criteria were applied beyond regular enrollment in the course. The relatively small class size (a maximum of 25 students per cohort) allowed for manageable observation and consistent access to learning materials. Data were collected anonymously to avoid revealing individuals during the study, as this would protect the privacy of the respondents (Sheng Zhong et al., 2019).

The chatbot system, OwlMentor, was integrated into the university's course platform and made available for voluntary use throughout the semester. Students were encouraged to use it independently to enhance their comprehension of the assigned scientific readings. The study setting reflects an authentic learning environment where the integration of AI tools such as OwlMentor could support real academic tasks, mirroring potential future use cases in higher education.

3.3 OwlMentor

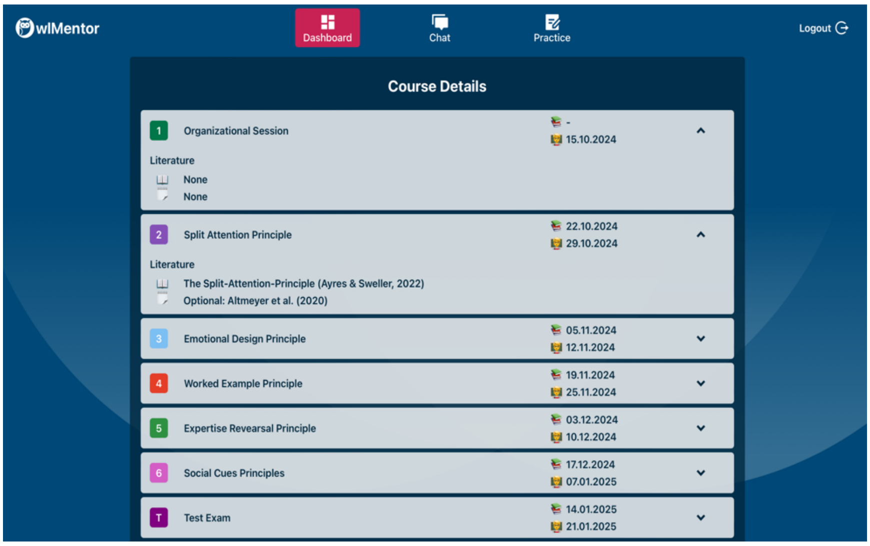
OwlMentor is a Retrieval-Augmented Generation (RAG) based AI chatbot designed to support student-generated questioning and reflection. It was developed using OpenAI's GPT-4 architecture with custom instructional prompts and integrated retrieval pipelines (Thüs et al., 2024). Key features included:

- Document-based conversational interface: Students could access assigned readings and engage in a dialogue with OwlMentor about the content.
- Automatic question co-creation and refinement: Students could generate questions in collaboration with the chatbot, refine them iteratively, and receive feedback on clarity and relevance.
- Real-time feedback
- Practice stored automatically generated questions: Students were able to store their generated questions and access them later as a personalized quiz bank for self-assessment.
- Learning statistics dashboard : The interface provided insights into usage patterns, number of questions generated, and quiz performance over time.

- Content summarization and key points extraction: Built-in features allowed students to request concise summaries and highlight key points from the reading material.
- Students were encouraged to use OwlMentor while completing assigned scientific readings. The chatbot's design aimed to foster metacognitive engagement by prompting learners to ask, refine, and answer questions, thereby reinforcing comprehension and supporting retrieval practice (Thüs et. Al 2024).

Figure 2

Owlmentor Home page that contains main functions with course details



Course Details

1 Organizational Session — - / 15.10.2024

Literature
None
None

2 Split Attention Principle — 22.10.2024 / 29.10.2024

Literature
The Split-Attention-Principle (Ayres & Sweller, 2022)
Optional: Altmeyer et al. (2020)

3 Emotional Design Principle — 05.11.2024 / 12.11.2024

4 Worked Example Principle — 19.11.2024 / 25.11.2024

5 Expertise Reversal Principle — 03.12.2024 / 10.12.2024

6 Social Cues Principles — 17.12.2024 / 07.01.2025

T Test Exam — 14.01.2025 / 21.01.2025

Figure 3

Owlmentor Home page with learner statistics

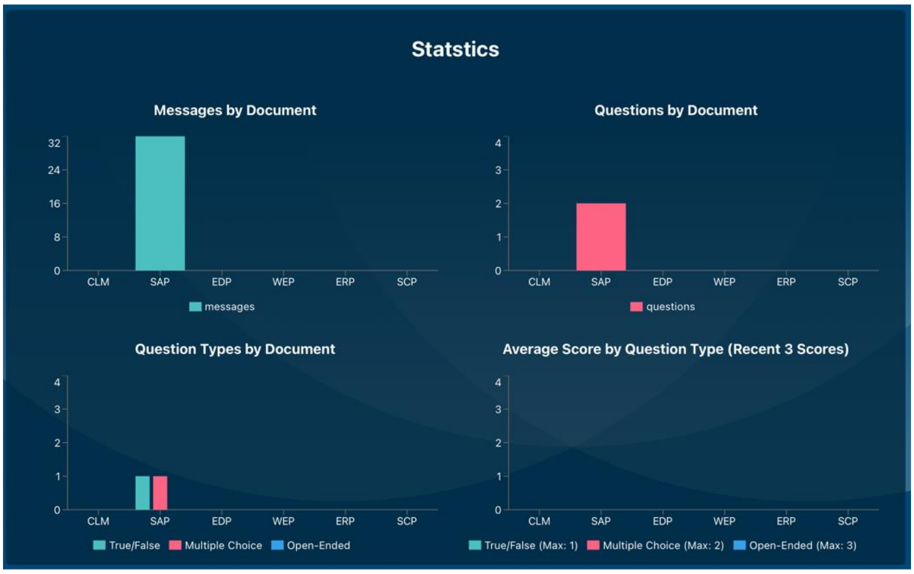


Figure 4

OwlMentor Chat Tab displays all the topics a learner can chat about

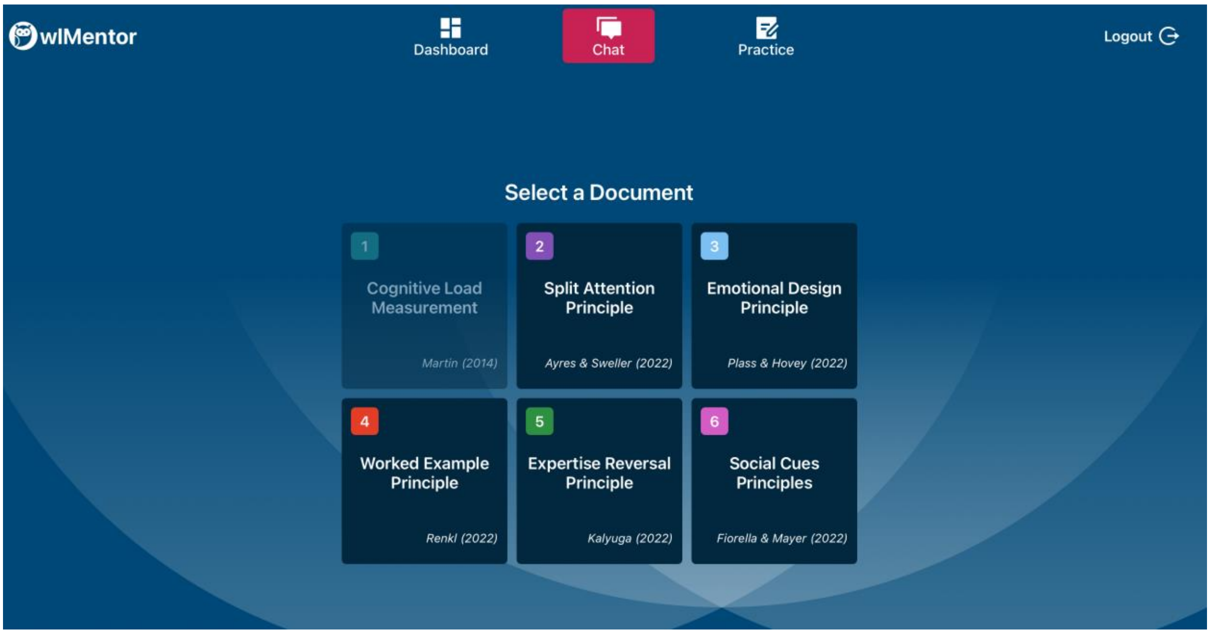
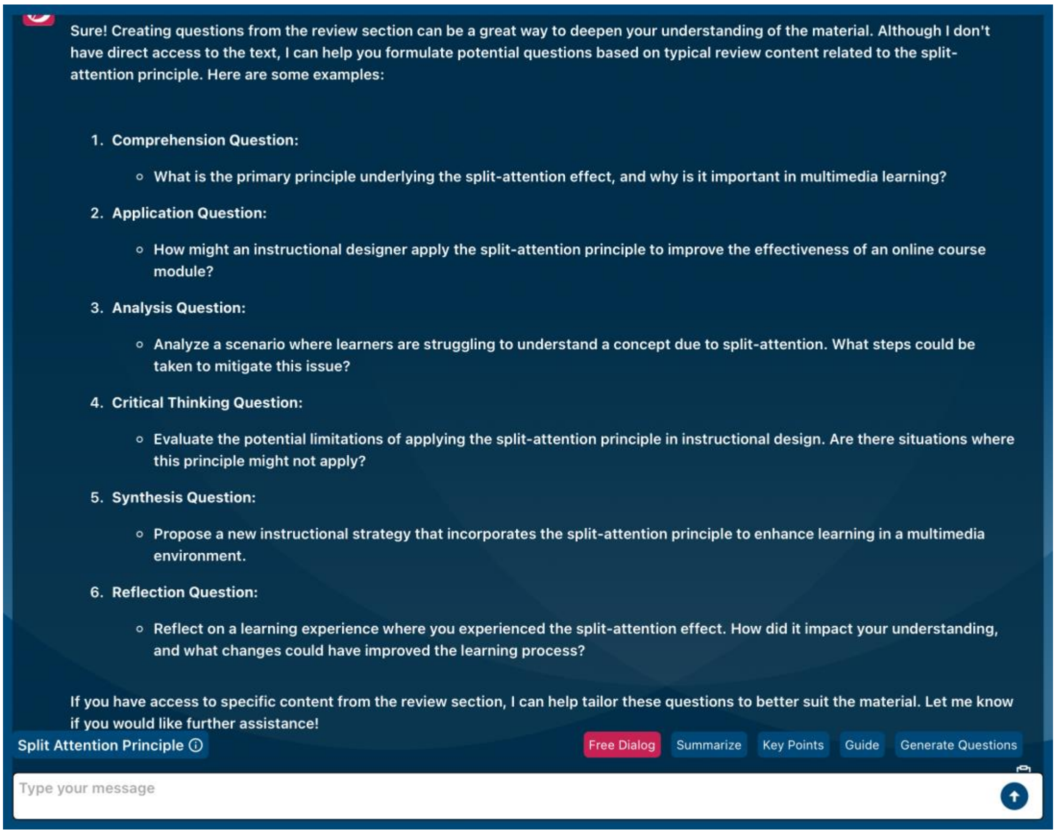


Figure 5

Appearance once Split-Attention Module is clicked under Chat tab



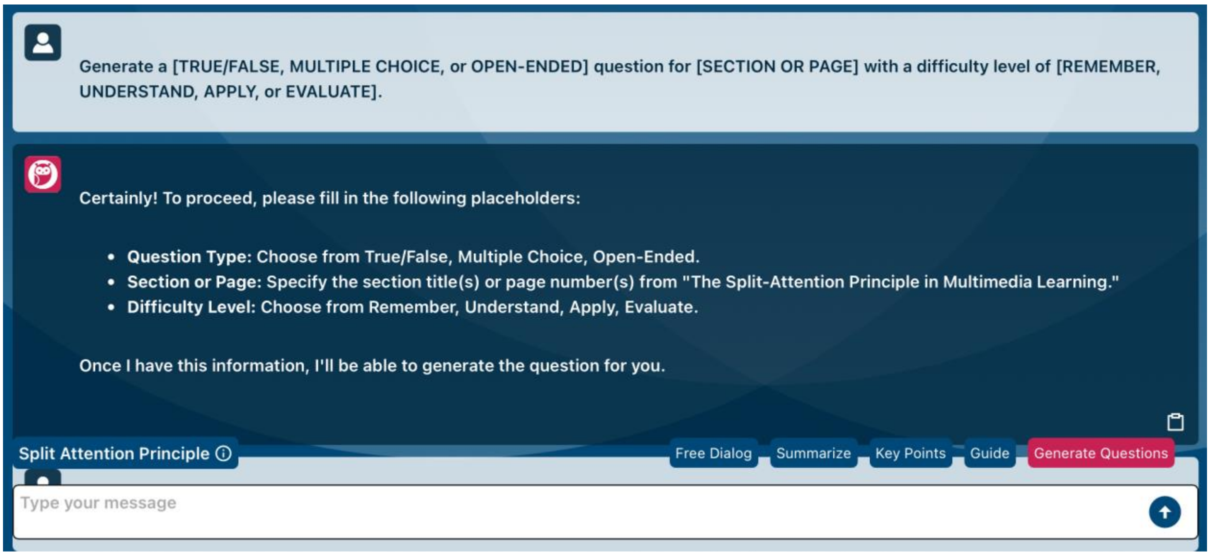
Sure! Creating questions from the review section can be a great way to deepen your understanding of the material. Although I don't have direct access to the text, I can help you formulate potential questions based on typical review content related to the split-attention principle. Here are some examples:

1. **Comprehension Question:**
    - What is the primary principle underlying the split-attention effect, and why is it important in multimedia learning?
2. **Application Question:**
    - How might an instructional designer apply the split-attention principle to improve the effectiveness of an online course module?
3. **Analysis Question:**
    - Analyze a scenario where learners are struggling to understand a concept due to split-attention. What steps could be taken to mitigate this issue?
4. **Critical Thinking Question:**
    - Evaluate the potential limitations of applying the split-attention principle in instructional design. Are there situations where this principle might not apply?
5. **Synthesis Question:**
    - Propose a new instructional strategy that incorporates the split-attention principle to enhance learning in a multimedia environment.
6. **Reflection Question:**
    - Reflect on a learning experience where you experienced the split-attention effect. How did it impact your understanding, and what changes could have improved the learning process?

If you have access to specific content from the review section, I can help tailor these questions to better suit the material. Let me know if you would like further assistance!

Figure 6

Prompt used for Automatic Question Generation related to a topic



Generate a [TRUE/FALSE, MULTIPLE CHOICE, or OPEN-ENDED] question for [SECTION OR PAGE] with a difficulty level of [REMEMBER, UNDERSTAND, APPLY, or EVALUATE].

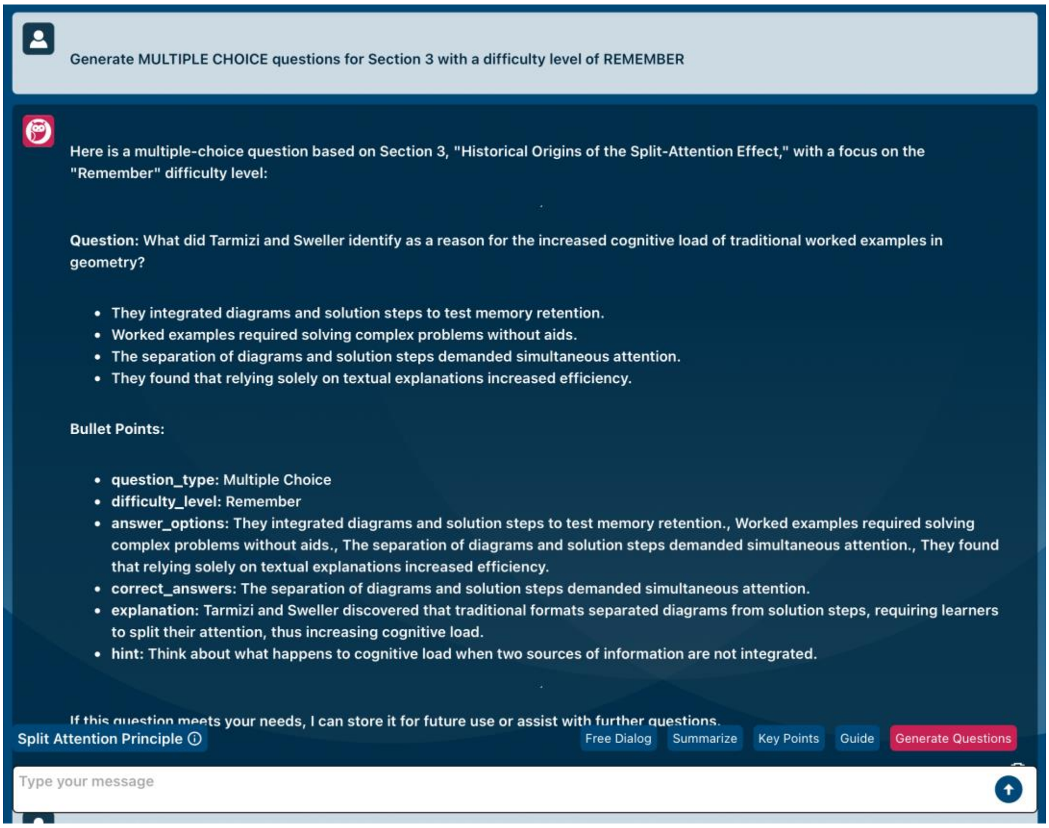
Certainly! To proceed, please fill in the following placeholders:

- **Question Type:** Choose from True/False, Multiple Choice, Open-Ended.
- **Section or Page:** Specify the section title(s) or page number(s) from "The Split-Attention Principle in Multimedia Learning."
- **Difficulty Level:** Choose from Remember, Understand, Apply, Evaluate.

Once I have this information, I'll be able to generate the question for you.

Figure 7

An automatically generated multiple choice question



Generate MULTIPLE CHOICE questions for Section 3 with a difficulty level of REMEMBER

Here is a multiple-choice question based on Section 3, "Historical Origins of the Split-Attention Effect," with a focus on the "Remember" difficulty level:

Question: What did Tarmizi and Sweller identify as a reason for the increased cognitive load of traditional worked examples in geometry?

- They integrated diagrams and solution steps to test memory retention.
- Worked examples required solving complex problems without aids.
- The separation of diagrams and solution steps demanded simultaneous attention.
- They found that relying solely on textual explanations increased efficiency.

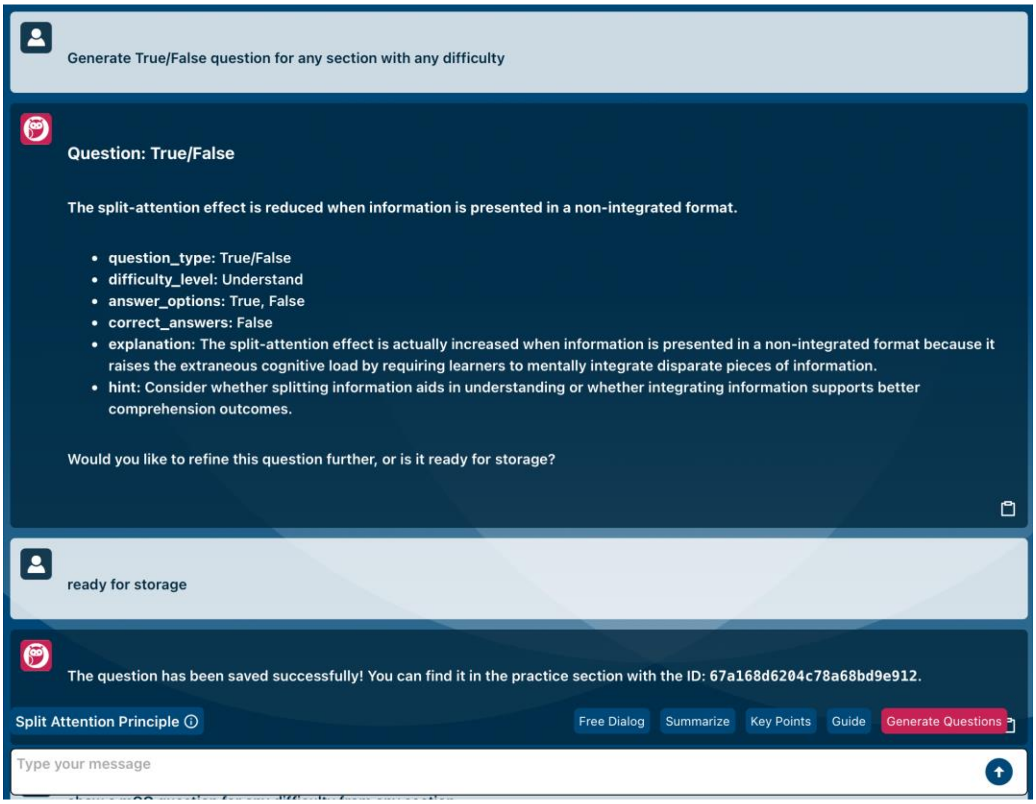
Bullet Points:

- **question_type:** Multiple Choice
- **difficulty_level:** Remember
- **answer_options:** They integrated diagrams and solution steps to test memory retention., Worked examples required solving complex problems without aids., The separation of diagrams and solution steps demanded simultaneous attention., They found that relying solely on textual explanations increased efficiency.
- **correct_answers:** The separation of diagrams and solution steps demanded simultaneous attention.
- **explanation:** Tarmizi and Sweller discovered that traditional formats separated diagrams from solution steps, requiring learners to split their attention, thus increasing cognitive load.
- **hint:** Think about what happens to cognitive load when two sources of information are not integrated.

If this question meets your needs, I can store it for future use or assist with further questions.

Figure 8

An automatically generated True/False question



Generate True/False question for any section with any difficulty

Question: True/False

The split-attention effect is reduced when information is presented in a non-integrated format.

- **question_type**: True/False
- **difficulty_level**: Understand
- **answer_options**: True, False
- **correct_answers**: False
- **explanation**: The split-attention effect is actually increased when information is presented in a non-integrated format because it raises the extraneous cognitive load by requiring learners to mentally integrate disparate pieces of information.
- **hint**: Consider whether splitting information aids in understanding or whether integrating information supports better comprehension outcomes.

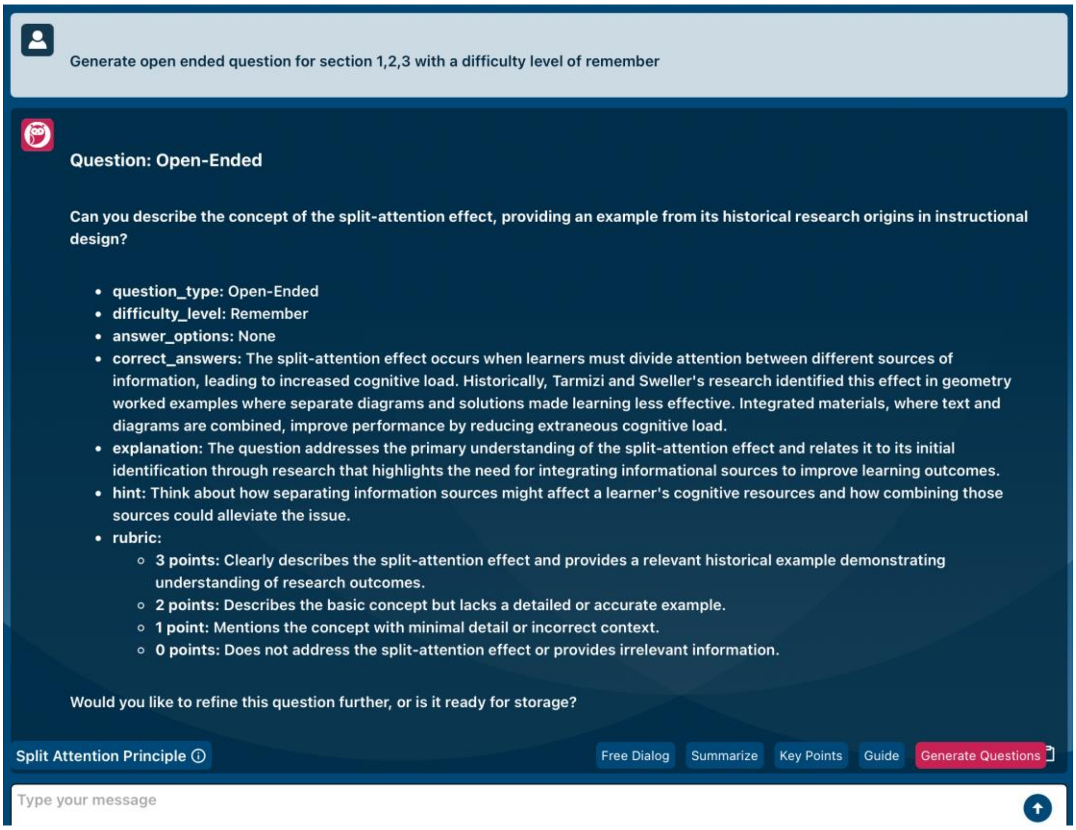
Would you like to refine this question further, or is it ready for storage?

ready for storage

The question has been saved successfully! You can find it in the practice section with the ID: `67a168d6204c78a68bd9e912`.

Figure 9

An automatically generated Open-Ended question



Generate open ended question for section 1,2,3 with a difficulty level of remember

Question: Open-Ended

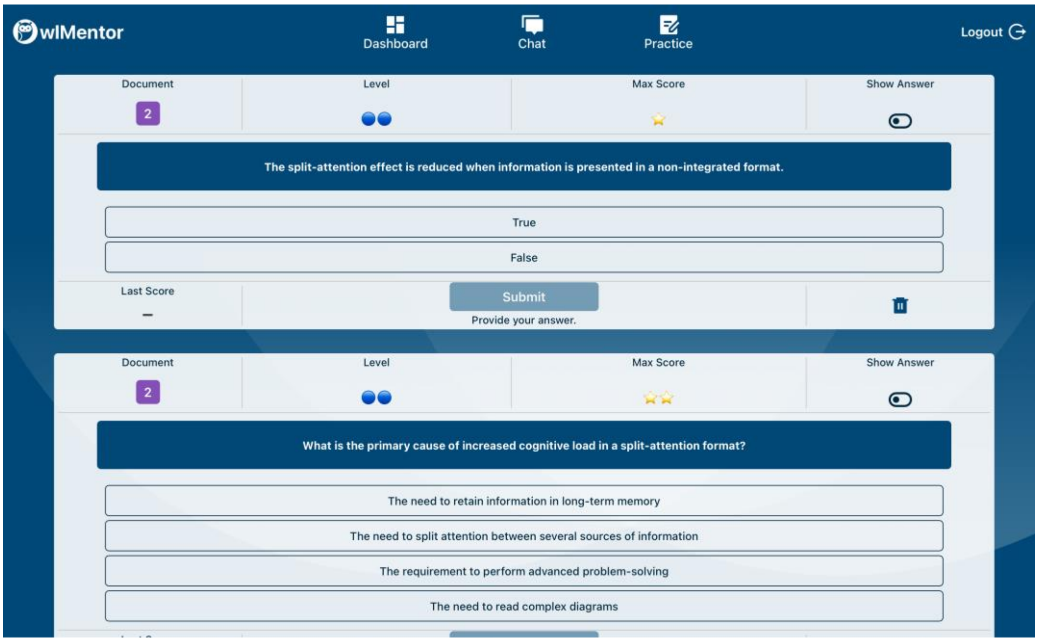
Can you describe the concept of the split-attention effect, providing an example from its historical research origins in instructional design?

- **question_type:** Open-Ended
- **difficulty_level:** Remember
- **answer_options:** None
- **correct_answers:** The split-attention effect occurs when learners must divide attention between different sources of information, leading to increased cognitive load. Historically, Tarmizi and Sweller's research identified this effect in geometry worked examples where separate diagrams and solutions made learning less effective. Integrated materials, where text and diagrams are combined, improve performance by reducing extraneous cognitive load.
- **explanation:** The question addresses the primary understanding of the split-attention effect and relates it to its initial identification through research that highlights the need for integrating informational sources to improve learning outcomes.
- **hint:** Think about how separating information sources might affect a learner's cognitive resources and how combining those sources could alleviate the issue.
- **rubric:**
  - **3 points:** Clearly describes the split-attention effect and provides a relevant historical example demonstrating understanding of research outcomes.
  - **2 points:** Describes the basic concept but lacks a detailed or accurate example.
  - **1 point:** Mentions the concept with minimal detail or incorrect context.
  - **0 points:** Does not address the split-attention effect or provides irrelevant information.

Would you like to refine this question further, or is it ready for storage?

Figure 10

Practice stored question under each topic



Document	Level	Max Score	Show Answer
2	●●	☆	

The split-attention effect is reduced when information is presented in a non-integrated format.

- True
- False

Last Score		
–	Submit Provide your answer.	

Document	Level	Max Score	Show Answer
2	●●	☆☆	

What is the primary cause of increased cognitive load in a split-attention format?

- The need to retain information in long-term memory
- The need to split attention between several sources of information
- The requirement to perform advanced problem-solving
- The need to read complex diagrams

3.4 Instruments and Measures

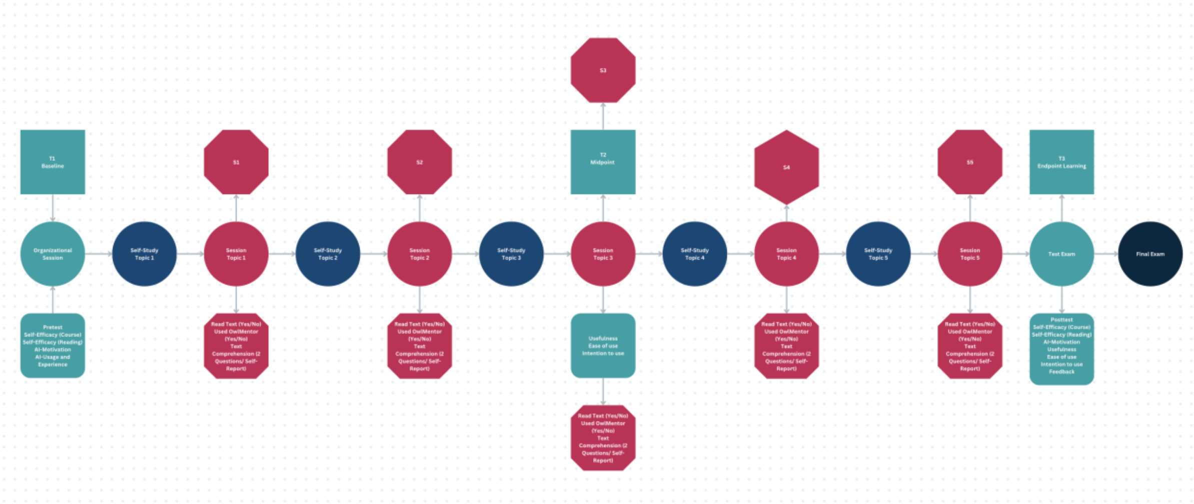
The following validated instruments and custom tools were employed:

- **Self-Efficacy Scales**: General academic self-efficacy (Pintrich & De Groot, 1990) and reading-specific self-efficacy (Koşar et al., 2022).
- **AI Motivation Scale**: Adapted from Yurt & Kasarci (2024) to assess students' interest and willingness to use AI tools.
- **Technology Acceptance Model (TAM)**: Includes Perceived Ease of Use (PEOU), Perceived Usefulness (PU), and Intention to Use (IU), based on Davis (1989) and Liaw (2008).
- **Pre/Post Comprehension Tests**: Domain-specific scientific texts were used to evaluate comprehension gains before and after chatbot use.
- **System Log Data**: Included number and type of interactions, questions generated, session length, and topic coverage.
- **Question Quality Rubric**: Based on Bloom's taxonomy, used by experts to rate the depth and complexity of student-generated questions.

3.5 Procedure

Figure 11

Integration and Data Collection plan across the semester



The study followed a longitudinal design over an entire semester. The course lasted approximately 15 weeks. Data collection occurred at 7 points throughout the course timeline using questionnaires. Three main timeline points and five session points were among them.

1. Week 1 (Organizational Session): Pre-test on Self-Efficacy (reading), Self-Efficacy (course), AI Motivation, AI Usage, and Experience
2. Week 7 (Midpoint of the course timeline): Usefulness, Ease of Use, and Intention to use Questionnaire
3. Week 15 (End point Learning) : Post-test on Self-Efficacy (reading), Self-Efficacy (course), AI Motivation, Usefulness, Ease of Use, and Intention to use
4. Weeks 3, 5, 7, 9, 11: Questionnaires to collect information on whether students read the reading materials for each session, if they used OwlMentor, and tested their knowledge with two comprehension questions.

3.6 Data Analysis

- **Quantitative Analysis**: Statistical analyses were conducted to examine correlations and differences between pre-/post-scores, TAM components, motivation, self-efficacy, and usage metrics. Paired sample t-tests and regression analyses were used.
- **Qualitative Analysis**: Thematic analysis of interaction logs was conducted to identify engagement patterns and instructional dialogue types. Student-generated questions were coded using the question quality rubric.

Table 1

Operational Definitions of Study Variables

Variable Name	Definition / Description	Measurement Items	Scale	Calculation
AI Motivation (Pre)	Students' initial motivation to use AI tools before the intervention.	T1_AI_MOT_1 – T1_AI_MOT_9	1–5 Likert	Mean of items
AI Motivation (Post)	Students' motivation to use AI tools after the intervention.	T3_AI_MOT_1 – T3_AI_MOT_9	1–5 Likert	Mean of items
Perceived Usefulness (PU)	Perception of the tool's usefulness for learning.	T3_PU_1 – T3_PU_4	1–5 Likert	Mean of items
Perceived Ease of Use (PEOU)	Perception of how easy the tool is to use.	T3_PEOU_1 – T3_PEOU_4	1–5 Likert	Mean of items
Self-Efficacy – Course (Pre)	Confidence in succeeding in the course before intervention.	T1_SE_COURSE_1 – T1_SE_COURSE_5	1–5 Likert	Mean of items
Self-Efficacy – Course (Post)	Confidence in succeeding in the course after intervention.	T3_SE_COURSE_1 – T3_SE_COURSE_5	1–5 Likert	Mean of items
Self-Efficacy – Reading (Pre)	Confidence in reading comprehension skills before intervention.	T1_SE_READ_1 – T1_SE_READ_5	1–5 Likert	Mean of items
Self-Efficacy – Reading (Post)	Confidence in reading comprehension skills after intervention.	T3_SE_READ_1 – T3_SE_READ_5	1–5 Likert	Mean of items

Variable Name	Definition / Description	Measurement Items	Scale	Calculation
Reading Completion Rate (%)	Percentage of assigned readings completed.	Weekly session self-reports	% scale	(# readings completed ÷ total readings) × 100
Understanding (Average)	Average weekly self-rating of text understanding.	Weekly questionnaire “Understanding” item	1–5 Likert	Mean across weeks
Recall (Average)	Average weekly score on recall questions.	Two comprehension questions per week	0–100%	Mean score
OwlMentor Usage Rate	Percentage of sessions where OwlMentor was used.	Weekly usage reports	% scale	(# sessions used ÷ total sessions) × 100
OwlMentor Total Interactions	Total number of chatbot messages, summarizations, key points, and questions generated.	System log data	Count	Raw sum

Table 2

Research Question–Data Source–Analysis Mapping

RQ	Data Source(s)	Variables Used	Analysis Method
RQ1.1: What types of questions do students generate using the chatbot?	System log data, question exports	Question type (MC, T/F, OE), Bloom's level	Frequency analysis, % distribution
RQ1.2: How does the chatbot affect students' engagement with scientific texts?	Weekly questionnaires, system logs, T1/T3 surveys	AI Motivation (Pre/Post), PU, Reading Completion Rate, Understanding, Recall, OM Usage Rate, OM Total Interactions	Descriptive stats, correlations, paired t-tests
RQ2: What learning outcomes and cognitive processes are supported?	Pre/post comprehension test scores, weekly recall, question taxonomy	Learning gains, Bloom's levels	Paired t-test, correlation analysis
RQ3.1: How does initial motivation/self-efficacy affect engagement?	T1 AI Motivation, T1 Self-Efficacy (course/reading), OM Usage Rate, OM Total Interactions	Same variables	Pearson correlations
RQ3.2: Does motivation/self-efficacy change over time?	T1/T3 AI Motivation, T1/T3 Self-Efficacy (course/reading)	Same variables	Paired t-tests

3.7 Ethical Considerations

Participation was voluntary, and informed consent was obtained from all students. Data were anonymized, and ethical approval was secured from the university's academic research board.

This methodological framework allowed for a multifaceted investigation into how AI-supported question generation can enhance reading comprehension and what factors influence its adoption and effectiveness.

Results

4.1 Overview

This chapter presents the results of the study in relation to the research questions. Data were analyzed using both quantitative and qualitative approaches to understand how students interacted with the AI chatbot, what types of questions they generated, their learning outcomes, and how motivation and self-efficacy evolved throughout the intervention.

4.2 RQ1: Student Interaction with the OwlMentor

4.2.1 Question Types and Cognitive Levels

RQ1.1: What types of questions do students generate using the chatbot (e.g., True/False, short answer, open-ended)?

During the intervention, students generated a total of 179 questions using OwlMentor. These questions came from 12 participants who used the chatbot during assigned reading activities. A frequency analysis (Table 3) showed that multiple-choice questions were the most common (*n* = 82, 45.8%), followed by true/false questions (*n* = 67, 37.4%), and open-ended questions (*n* = 30, 16.8%).

Table 3

Distribution of question types generated with OwlMentor

Question Type	Frequency	Percent	Valid Percent	Cumulative Percent
Multiple Choice	82	45.8%	45.8%	45.8%
Open-Ended	30	16.8%	16.8%	62.6%
True/False	67	37.4%	37.4%	100%
Total	179	100%	100%	

Cognitive Level Distribution:

Questions were also classified according to Bloom’s taxonomy levels (Table 4). The majority were at the “Understand” level (*n* = 76, 42.5%), followed by “Apply” (*n* = 52, 29.1%), “Evaluate” (*n* = 31, 17.3%), and “Remember” (*n* = 20, 11.2%).

Table 4

Distribution of questions by Bloom's taxonomy cognitive levels

	Frequency	Percent	Valid Percent	Cumulative Percent
Apply	52	29.1	29.1	29.1
Evaluate	31	17.3	17.3	46.4
Remember	20	11.2	11.2	57.5
Understand	76	42.5	42.5	100.0
Total	179	100.0	100.0	

Figure 12

Bar chart of question types

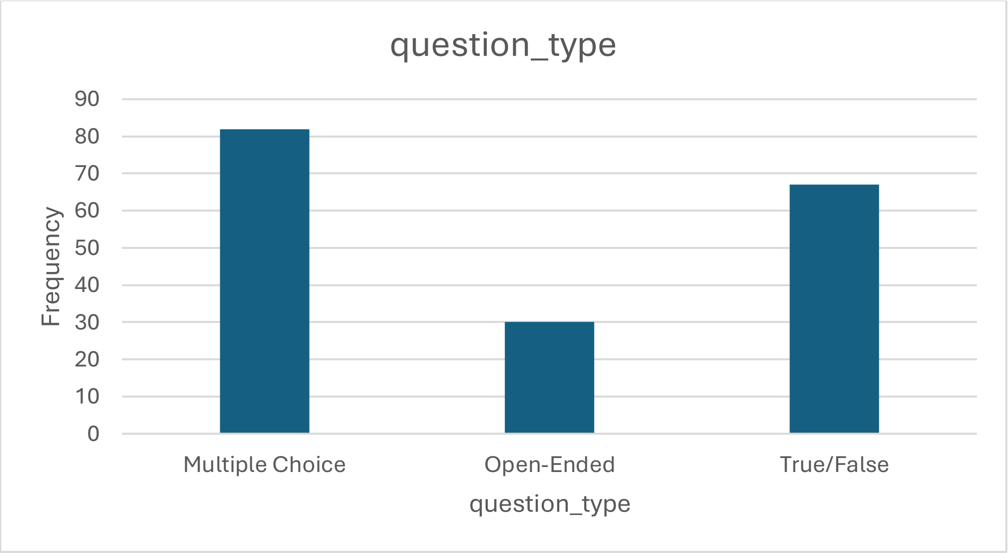
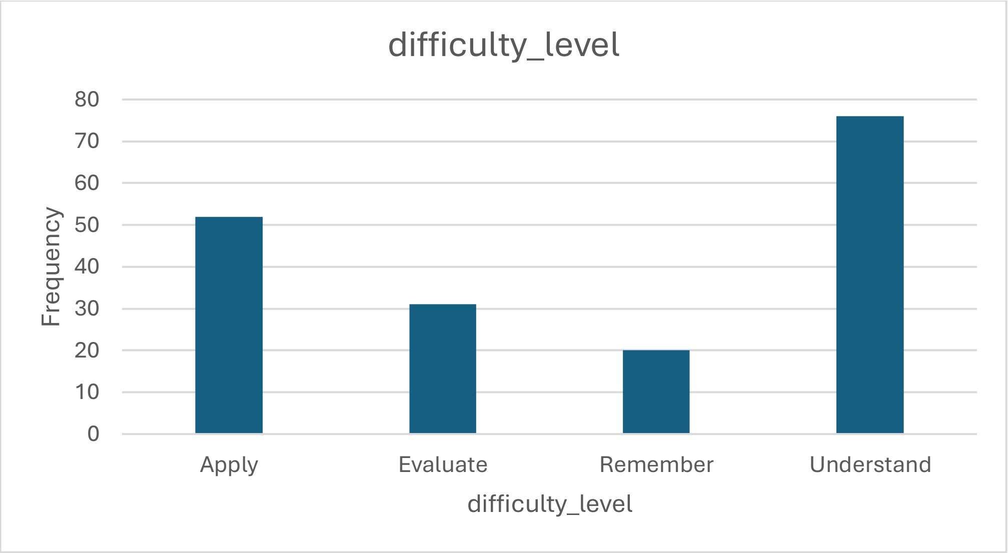


Figure 13

Bloom's taxonomy distribution by difficulty level



These findings suggest that OwlMentor facilitated the creation of diverse question formats, with multiple-choice and true/false questions being the most common. From a cognitive perspective, most questions targeted mid-level skills (understanding and application) rather than higher-order thinking. This pattern is consistent with prior research on AQG, which often produces items that emphasize comprehension over synthesis or creation (Lee et al., 2023; Mulla et al., 2023). Subsequent analyses (Section 4.2.2) examine how the distribution of question types and cognitive levels relates to students' engagement and learning outcomes.

4.2.2 Research Question 1.2

RQ1.2: How does OwlMentor affect students' engagement with scientific texts?

Engagement was measured using behavioral indicators (reading completion rate, OwlMentor usage rate, and total chatbot interactions) and cognitive indicators (self-reported understanding and recall ratings). AI Motivation and Perceived Usefulness scores were also included as attitudinal measures. Analyses included descriptive statistics, paired-samples t-tests, Pearson correlation coefficients, and multiple regression models to examine relationships among these variables.

4.2.2.1 Descriptive Statistics.

Descriptive statistics were computed for AI Motivation (Pre and Post), Perceived Usefulness (Post), and behavioral and cognitive engagement measures (Table 5). Overall, AI Motivation scores were relatively high both before (*M* = 3.71, *SD* = 0.43) and after (*M* = 3.62, *SD* = 0.47) the intervention. Perceived usefulness was notably high (*M* = 5.18, *SD* = 1.01) on the given scale. Behavioral engagement indicators showed moderate averages with some variability, as reflected in reading completion rate (*M* = 0.71, *SD* = 0.32) and OwlMentor usage rate (*M* = 0.50, *SD* = 0.28). Cognitive engagement, measured through understanding and recall ratings, indicated generally strong perceived learning (understanding: *M* = 4.05; recall: *M* = 3.89). Total interactions with OwlMentor varied widely among participants, from 14 to 273 messages.

Table 5

Descriptive statistics for engagement and motivation measures

Variable	N	M	SD	Min	Max
AI Motivation (Pre)	13	3.71	0.43	3.00	4.22
AI Motivation (Post)	11	3.62	0.47	2.89	4.44
Perceived Usefulness (Post)	11	5.18	1.01	2.75	6.50
Reading Completion Rate	12	0.71	0.32	0.25	1.00
OwlMentor Usage Rate	12	0.50	0.28	0.20	1.00
Understanding (Average Weekly Rating)	12	4.05	0.37	3.25	4.80
Recall (Average Weekly Rating)	12	3.89	0.54	3.25	4.80
OM Total Interactions	12	78.92	77.87	14	273

Note. M = Mean, *SD* = Standard Deviation.

4.2.2.2 Paired Samples t-test.

A paired-samples t-test was conducted to determine whether AI Motivation changed significantly from pre- to post-intervention (Table 6). The results indicated no significant difference between pre-test (*M* = 3.71, *SD* = 0.43) and post-test (*M* = 3.62, *SD* = 0.47) scores, *t*(10) = 1.08, *p* = .307, 95% CI [–0.097, 0.279], *d* = 0.28. This suggests that the chatbot intervention did not produce a statistically significant change in students' AI Motivation.

Table 6

Paired-samples t-test results for AI Motivation (Pre vs. Post)

Statistic	Pre-test	Post-test
Mean	3.71	3.62
SD	0.41	0.47
N	11	11
Correlation (Pre–Post)	r = .804, p = .003	—
Mean Difference	0.09	—
SD Difference	0.28	—
SE Mean Difference	0.08	—
95% CI of Difference	–0.097, 0.279	—
t(10)	1.076	—
p (two-tailed)	.307	—
Cohen's d	0.28	—

4.2.2.3 Correlation Analysis.

Pearson correlation coefficients were computed among all post-course engagement variables (Table 4.3). AI Motivation (Post) was negatively correlated with total chatbot interactions, *r*(9) = –.73, *p* = .011, indicating that students with higher motivation tended to interact less frequently with the chatbot. Perceived usefulness was positively correlated with recall, *r*(9) = .71, *p* = .015, suggesting that students who found OwlMentor more useful also remembered more of the content. Understanding was positively correlated with recall, *r*(10) = .66, *p* = .020. OwlMentor usage rate was strongly correlated with total interactions, *r*(10) = .77, *p* = .004. No other correlations reached statistical significance.

Table 7

Pearson correlations among engagement and motivation variables

Variable	1	2	3	4	5	6	7
1. AI Motivation (Post)	—	.34	–.31	–.60	.03	.13	–.73*
2. Perceived Usefulness (Post)	.34	—	–.29	–.10	.35	.71*	.00
3. Reading Completion Rate (%)	–.31	–.29	—	.08	.26	.28	.16
4. OM Usage Rate	–.60	–.10	.08	—	–.11	–.13	.77**
5. Understanding (Avg Weekly)	.03	.35	.26	–.11	—	.66*	–.05
6. Recall (Avg Weekly)	.13	.71*	.28	–.13	.66*	—	.07
7. OM Total Interactions	–.73*	.00	.16	.77**	–.05	.07	—

Note. *$p < .05$, **$p < .01$.

4.2.2.4 Regression Analyses.

Two multiple regressions were conducted to identify predictors of post-course AI Motivation and Perceived Usefulness.

Model 1: Predicting AI Motivation (Post)

A multiple linear regression examined whether reading completion rate, OwlMentor usage rate, total chatbot interactions, understanding, and recall predicted post-course AI Motivation. The overall model was not statistically significant, $F(5, 5) = 2.38$, $p = .182$, $R^2 = .70$, Adjusted $R^2 = .41$. None of the predictors reached significance; however, total chatbot interactions demonstrated a negative, non-significant association with motivation ($\beta = –1.17$, $p = .083$).

Table 8

Multiple regression predicting AI Motivation (Post)

Predictor	B	SE B	β	t	p
Reading Completion Rate	-0.615	0.425	–.417	–1.45	.208
OM Usage Rate	0.875	0.933	.530	0.94	.391
Understanding (Avg)	0.120	0.491	.095	0.24	.817

Predictor	B	SE B	β	t	p
Recall (Avg)	0.222	0.293	.264	0.76	.483
OM Total Interactions	–0.007	0.003	–1.169	–2.16	.083

Model 2: Predicting Perceived Usefulness (Post)

A multiple regression was conducted to examine whether reading completion rate, OwlMentor usage rate, total chatbot interactions, understanding, and recall predicted post-course perceived usefulness of OwlMentor. The model explained 75.2% of the variance in perceived usefulness ($R^2 = .75$, Adjusted $R^2 = .50$), $F(5, 5) = 3.03$, $p = .124$. Although the overall model was not statistically significant, recall was a significant positive predictor ($\beta = .92$, $p = .035$), indicating that students with better recall tended to perceive the chatbot as more useful.

Table 9

Multiple regression predicting Perceived Usefulness (Post)

Predictor	B	SE B	β	t	p
Reading Completion Rate	–1.669	0.839	–.526	–1.99	.103
OM Usage Rate	1.034	1.840	.291	0.56	.598
Understanding (Avg)	–0.146	0.968	–.054	–0.15	.886
Recall (Avg)	1.656	0.578	.916	2.87	.035*
OM Total Interactions	–0.003	0.006	–.243	–0.49	.645

Note. *$p < .05$.

4.2.2.5 Findings

The analysis indicated that while AI Motivation did not significantly change over time, several meaningful relationships emerged between engagement measures. Students with higher post-course motivation tended to interact less with the chatbot, whereas those with better recall perceived it as more useful. Recall was the only significant predictor of perceived usefulness, highlighting the link between cognitive engagement and positive evaluations of the tool.

4.3 RQ2: Learning Outcomes

RQ2: What learning outcomes and cognitive processes are supported by the AI chatbot?

*RQ2.1***:** *Do students benefit from using the chatbot in terms of learning outcomes?*

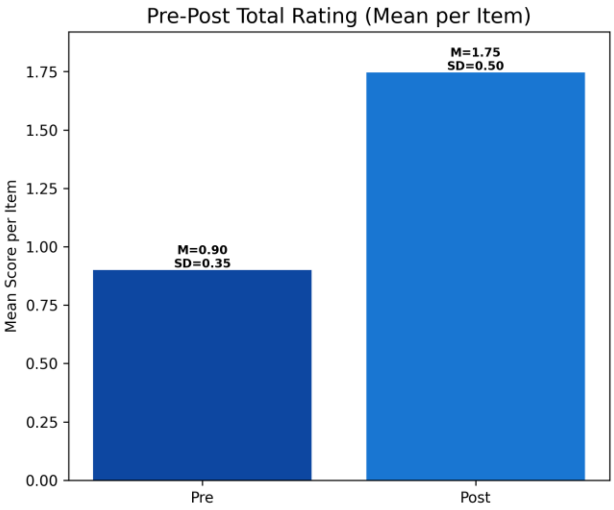
4.3.1 Learning Gains

Learning Gains – Overall

A paired-samples t-test was conducted to compare pre- and post-intervention comprehension scores across all assessment items. Results indicated a significant improvement from pre-test (M = 0.90, SD = 0.35) to post-test (M = 1.75, SD = 0.50), $t(10)$ = 5.89, p < .001, 95% CI [0.53, 1.20], d = 1.78. This represents a large effect size, indicating substantial gains in comprehension following the chatbot-supported learning activities (N=11).

Figure 14

Mean score improvement

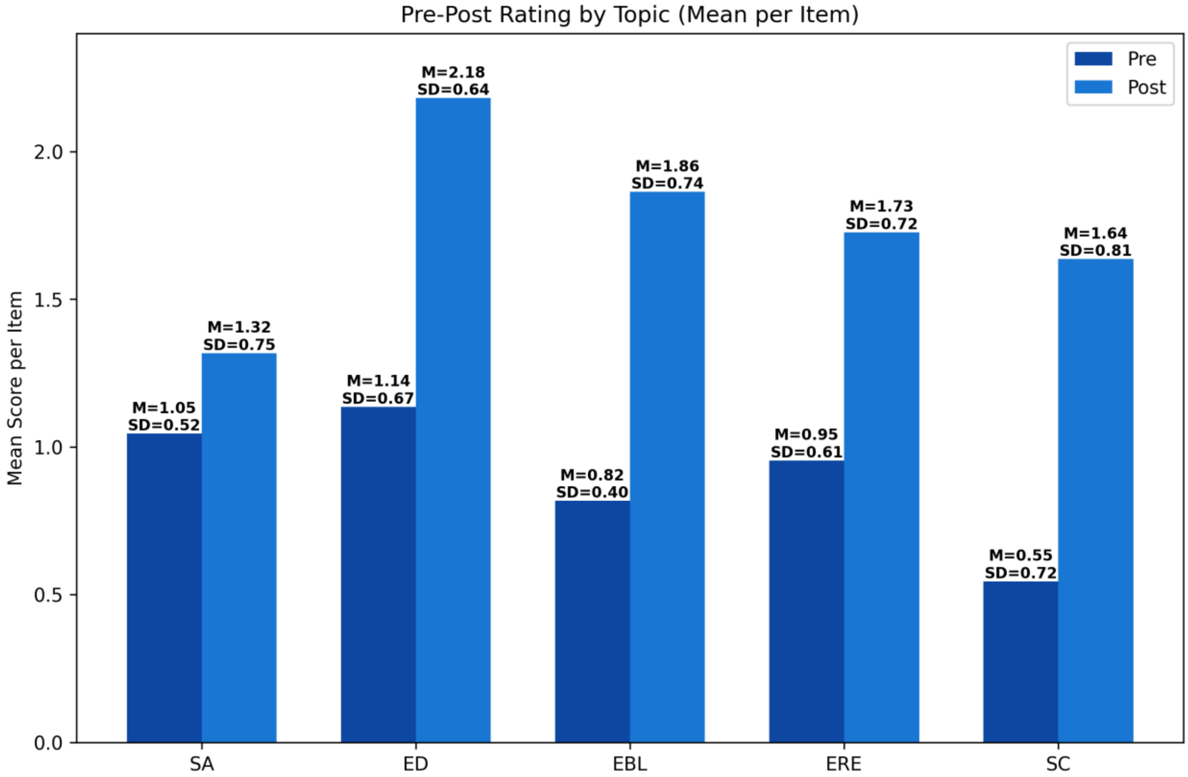


Learning Gains - By Topic

A repeated-measures ANOVA was conducted with time (pre/post) and topic (SA, ED, EBL, ERE, SC) as within-subject factors. There was a significant main effect of time, *F*(1, 10) = 34.68, *p* < .001, $\eta^2 = .30$, confirming the overall improvement observed in the t-test. The main effect of topic was not significant after Greenhouse–Geisser correction, *F*(4, 40) = 2.63, *p* = .079, $\eta^2 = .08$, and the time × topic interaction was also non-significant, *F*(4, 40) = 2.44, *p* = .084, $\eta^2 = .05$.

Figure 15

Mean Score Improvement by topic



Interpretation

The findings indicate **a** clear overall improvement in comprehension scores from pre- to post-assessment, suggesting that the chatbot-supported intervention was associated with significant learning gains. However, the improvement was general across all topics**,** with no significant differences in gains between specific subject areas.

4.4 RQ3: Motivation and Self-Efficacy

4.4.1 Effect of Motivation and SE on Engagement (RQ3.1)

Pearson correlation analyses were conducted to examine associations between baseline measures, T1 AI Motivation, T1 Self-Efficacy in Course (T1_SE), and T1 Reading Self-Efficacy (T1_RSE), and engagement metrics (OwlMentor Usage Rate, Reading Completion Rate, Understanding, Recall, OM Total Interactions).

The analysis revealed a significant negative correlation between T1 AI Motivation and OwlMentor Usage Rate ($r = -.611$, $p = .035$), indicating that students with higher initial AI motivation tended to use the chatbot less frequently. T1 AI Motivation was not significantly correlated with other engagement variables, including OM Total Interactions, Reading Completion Rate, Understanding, and Recall (all $p > .05$).

A significant positive correlation was observed between T1_SE and T1_RSE ($r = .662$, $p = .014$), confirming that general academic self-efficacy and reading-specific self-efficacy were related. However, neither was significantly associated with most engagement metrics.

Among the engagement variables themselves, OwlMentor Usage Rate was strongly positively correlated with OM Total Interactions ($r = .768$, $p = .004$), as expected given their conceptual overlap. Additionally, Understanding (average weekly rating) was positively correlated with Recall (average weekly rating) ($r = .657$, $p = .020$).

Table 10

Pearson correlation coefficients between baseline measures and engagement metrics

Variable	1	2	3	4	5	6	7	8
1.AI Motivation (Pre)	1.00	0.08	0.37	−0.61*	−0.11	−0.07	0.18	−0.55
2.T1_SE	0.08	1.00	0.66*	−0.07	−0.51	−0.13	−0.21	0.13
3.T1_RSE	0.37	0.66*	1.00	−0.46	−0.21	0.28	0.49	−0.06
4.OwlMentor Usage Rate	−0.61*	−0.07	−0.46	1.00	0.08	−0.11	−0.13	0.77**
5.Reading Completion Rate (%)	−0.11	−0.51	−0.21	0.08	1.00	0.26	0.28	0.16
6.Understanding (Average Weekly Rating)	−0.07	−0.13	0.28	−0.11	0.26	1.00	0.66*	−0.05
7.Recall (Average Weekly Rating)	0.18	−0.21	0.49	−0.13	0.28	0.66*	1.00	0.07
8.OM Total Interactions	−0.55	0.13	−0.06	0.77**	0.16	−0.05	0.07	1.00

Note. $p < .05$*, $p < .01$.

Interpretation

These results suggest that students with high initial motivation may be more self-reliant and thus less dependent on the chatbot. Conversely, lower-motivation students may have relied more on OwlMentor for support. While self-efficacy measures were internally consistent, they did not significantly predict chatbot engagement patterns.

4.4.2 Change in Motivation and SE Over Time

RQ3.2: Does motivation or self-efficacy change over time through the use of the chatbot?

To determine whether students' AI motivation, self-efficacy in reading, and self-efficacy in the course changed over time, paired-samples t-tests were conducted comparing pre-test (T1) and post-test (T3) scores.

Results indicated that there was no significant change in AI Motivation from T1 (*M* = 3.71, *SD* = 0.41) to T3 (*M* = 3.62, *SD* = 0.47), *t*(10) = 1.08, *p* = .307, Cohen's *d* = 0.28. Similarly, Reading Self-Efficacy scores did not differ significantly from T1 (*M* = 4.37, *SD* = 0.54) to T3 (*M* = 4.43, *SD* = 0.67), *t*(9) = −0.74, *p* = .479, Cohen's *d* = −0.23. Course Self-Efficacy also showed no statistically significant change between T1 (*M* = 4.10, *SD* = 0.58) and T3 (*M* = 3.81, *SD* = 0.71), *t*(10) = 1.66, *p* = .127, Cohen's *d* = 0.50.

Table 11

Paired-samples t-tests for AI motivation, reading self-efficacy, and course self-efficacy (T1 vs. T3)

Variable	**Time**	***M***	***SD***	***t***	***df***	***p***	**Cohen's *d***
AI Motivation	T1	3.71	0.41	1.08	10	.307	0.28
	T3	3.62	0.47				
Reading Self-Efficacy	T1	4.37	0.54	−0.74	9	.479	−0.23
	T3	4.43	0.67				

Variable	**Time**	***M***	***SD***	***t***	***df***	***p***	**Cohen's *d***
Course Self-Efficacy	T1	4.10	0.58	1.66	10	.127	0.50
	T3	3.81	0.71				

Interpretation

These findings suggest that over the course of the study, students' levels of AI motivation and self-efficacy remained relatively stable. The absence of significant changes may indicate that engagement with OwlMentor neither strongly enhanced nor diminished these constructs, possibly reflecting ceiling effects or already high baseline levels.

4.5 Summary of Key Findings

RQ1: How do students interact with the OwlMentor?

- **RQ1.1 (Question types & cognitive levels)**
    - Students generated 179 questions.
    - Most frequent type: Multiple Choice (45.8%), followed by True/False (37.4%) and Open-Ended (16.8%).
    - Bloom's taxonomy distribution: Understand (42.5%), Apply (29.1%), Evaluate (17.3%), Remember (11.2%).
    - Implication: OwlMentor primarily supported mid-level cognitive engagement rather than lower-level recall or higher-level evaluation.
- **RQ1.2 (Effect on engagement)**
    - No significant pre–post change in AI Motivation (p = .307).
    - Correlations showed that post AI Motivation was negatively correlated with total interactions (r = −.73, p = .011) and positively correlated with recall (r = .71, p = .015).
    - Regression models indicated no significant predictors of post AI Motivation or perceived usefulness, though recall emerged as a positive predictor for perceived usefulness (p = .035).

RQ2: What learning outcomes and cognitive processes are supported by the AI chatbot?

- Learning gains analysis (from comprehension tests) showed small, non-significant improvements in performance over the semester.
- Average weekly ratings indicated moderate understanding (M ≈ 4.05) and moderate recall (M ≈ 3.89).
- Correlations revealed a positive relationship between understanding and recall ($r = .66$, $p = .020$), suggesting these cognitive processes were linked during chatbot use.

RQ3: How does learner motivation and self-efficacy influence (and get influenced by) the use of the AI chatbot?

- **RQ3.1 (Baseline influence on engagement)**
  - Pre AI Motivation was negatively correlated with OwlMentor usage rate ($r = −.61$, $p = .035$), indicating that students with higher initial motivation tended to use the chatbot less.
  - Reading self-efficacy and course self-efficacy at baseline were not significantly related to usage patterns.
- **RQ3.2 (Changes over time)**
  - No significant changes from pre to post in AI Motivation ($p = .307$), Reading Self-Efficacy ($p = .479$), or Course Self-Efficacy ($p = .127$).
  - **Implication**: Motivation and self-efficacy remained stable, suggesting possible ceiling effects or that chatbot usage did not meaningfully shift these constructs.

Discussion

The purpose of this study was to examine how an AI-based question generation chatbot, OwlMentor, supports engagement with scientific texts, fosters learning outcomes, and interacts with learner motivation and self-efficacy. A mixed-methods approach was applied, combining pre- and post-intervention questionnaires, system usage logs, and comprehension assessments with students enrolled in the Master’s in Educational Technology program.

This chapter interprets the findings in relation to the study’s research questions, connects them to prior literature, and considers their implications for educational practice, technology design, and future research. The discussion is organized around the three main research questions, followed by practical implications, limitations, and recommendations for further study.

5.1 Discussion of Research Questions

5.1.1 RQ1.1: Question Types and Cognitive Levels

The analysis of student-generated questions revealed that multiple-choice questions were the most common (45.8%), followed by true/false questions (37.4%) and open-ended questions (16.8%). These findings are consistent with previous studies (Thüs et al., 2024; Lee et al., 2023), which reported that students tend to favor closed formats when using AQG tools.

The Bloom’s taxonomy distribution indicated that most questions fell within the “Understand” (42.5%) and “Apply” (29.1%) categories, with fewer instances of “Evaluate” (17.3%) and “Remember” (11.2%). This mid-level cognitive focus aligns with research demonstrating that automatic questioning can scaffold comprehension and application (King, 1992; Rosenshine, Meister, & Chapman, 1996), but may require additional scaffolding to stimulate higher-order thinking. The preference for mid-level tasks may reflect both the design of OwlMentor’s prompts and students’ inclination toward efficient, less cognitively demanding formats.

5.1.2 RQ1.2: Engagement with Scientific Texts

Although AI motivation did not increase significantly from pre-test to post-test, correlations indicated that recall scores were positively associated with perceived usefulness, suggesting that cognitive benefits influence perceived value. A notable finding was the negative correlation between pre-intervention AI motivation and OwlMentor usage rate, implying that students with higher initial motivation tended to rely less on the tool. This mirrors findings from Thüs et al. (2024), where highly motivated learners engaged less with AI tools, possibly due to greater self-regulatory capacity.

These results suggest that OwlMentor may be particularly valuable for students with moderate or lower initial motivation, offering scaffolding that helps sustain engagement with reading tasks.

5.1.3 RQ2: Learning Outcomes and Cognitive Processes

Learning gains, as measured by comprehension assessments, showed small, non-significant improvements over the semester. While this may partly be due to the limited intervention period and small sample size, it is also possible that the comprehension tests lacked the sensitivity to capture incremental learning benefits.

Average weekly self-reports of understanding and recall were moderately high, and the two measures were positively correlated (r = .66, p = .020). This relationship supports the cognitive theory that comprehension facilitates memory retention and aligns with retrieval practice literature (Karpicke & Blunt, 2011).

The Bloom's taxonomy distribution of generated questions complements these findings, indicating that OwlMentor successfully supported comprehension and application but did not strongly promote the development of higher-order cognitive processes such as synthesis or evaluation. This gap points to a potential design improvement for AQG systems: incorporating targeted prompts or adaptive scaffolds to encourage more complex question generation.

5.1.4 RQ3.1: Motivation and Self-Efficacy - Baseline Influence on Engagement

The correlation analysis showed that initial AI motivation was negatively associated with OwlMentor usage rate (r = −.61, p = .035), supporting Wang and Wang's (2019) suggestion that overconfidence can reduce uptake of support tools.. This finding suggests that students with greater intrinsic interest in AI or confidence in their ability to use it effectively may perceive less need for prolonged interaction with the tool. No significant associations were found between self-efficacy measures and usage, indicating that

OwlMentor was equally accessible to learners regardless of confidence levels. The strong positive correlation between understanding and recall, and between usage rate and total interactions, highlights the interconnectedness of engagement behaviors and perceived learning benefits.

5.1.5 RQ3.2: Motivation and Self-Efficacy - Changes Over Time

Paired-samples t-tests revealed no significant changes in AI motivation, reading self-efficacy, or course self-efficacy over the course of the intervention. These results are consistent with prior research (Kurdi et al., 2019) suggesting that such constructs remain relatively stable over short instructional periods, particularly when baseline levels are already high. Rather than boosting motivation or confidence, OwlMentor may have played a maintenance role, sustaining students' initial levels while providing cognitive support for reading tasks.

5.2 Interpretation in Light of Existing Literature

The observed emphasis on mid-level cognitive questions mirrors findings from King (1992) and Bhat et al. (2022), where scaffolded question generation was shown to improve comprehension and critical thinking skills, but higher-order questioning remained rare without explicit instructional support. In the present study, the predominance of "Understand" and "Apply" level questions suggests that OwlMentor effectively guided learners toward meaningful engagement with the text, yet lacked mechanisms to consistently push them toward the "Analyze" or "Create" levels of Bloom's taxonomy. This is a known limitation in many AI-supported question generation environments (Mulla et al., 2023), where interface design and default prompts often bias toward more straightforward question formats.

The modest, non-significant learning gains parallel findings from Kasneci et al. (2023), where AI tools integrated into authentic coursework supported comprehension without producing dramatic short-term score improvements. One explanation is that comprehension and retention benefits may be more qualitative than quantitative in the early stages of AI integration, emerging as stronger long-term effects with continued exposure. Another consideration is the "ceiling effect," where students' high baseline performance, as seen in this cohort of MSc EduTech learners, leaves limited room for measurable improvement.

Findings related to the Technology Acceptance Model (TAM) are consistent with the original propositions by Davis (1989) and expanded interpretations by Venkatesh and Davis (2000). In both prior studies and the present research, perceived ease of use appeared more influential during the initial adoption phase, while perceived usefulness correlated more closely with sustained engagement. This dynamic underscores that while usability is critical for onboarding learners, their ongoing interaction with AI tools depends more on the value they attribute to its role in their learning process.

The motivational dimension aligns with recent work by Thüs et al. (2024) and Yurt and Kasarci (2024), who emphasize that learners' willingness to engage with AI is not purely instrumental but is shaped by curiosity, affective engagement, and openness to experimentation. Interestingly, this study found that higher self-reported AI motivation at baseline predicted lower actual usage, suggesting that for some learners, enthusiasm for AI does not translate into sustained tool engagement. This echoes Wang and Wang's (2019) conclusion that high self-efficacy or overconfidence can lead to reduced reliance on external scaffolding, as learners may believe they can achieve learning goals independently.

5.3 Practical Implications

Several actionable insights emerge from these findings:

1. **Instructional Design Integration**

   Embedding AQG tools like OwlMentor into the course structure, particularly within scheduled, graded, or collaborative activities, can encourage consistent use. Passive availability alone may not be sufficient to drive sustained engagement.

2. **Prompt Engineering for Higher-Order Thinking**

   Designers should consider integrating adaptive prompts that scaffold learners from simpler recall-oriented questions to more complex, analytical, and evaluative formats. This would better align with the metacognitive benefits of student-generated questioning identified in prior literature (Yu, 2009; Rosenshine et al., 1996).

3. **Targeting Motivation and Self-Efficacy Profiles**

   AI training and orientation sessions can frame chatbot use as a metacognitive strategy, helping both highly motivated and less confident learners see its relevance. For high self-efficacy learners, framing the chatbot as a tool for *extending* thinking rather than *supporting* basic comprehension may increase adoption.

4. **Feedback Loop Design**

   Incorporating direct feedback mechanisms, such as automated performance summaries or personalized study suggestions, can increase perceived usefulness and thus sustain engagement.

5.4 Theoretical Contributions

This study adds to the Technology Acceptance Model by integrating AI motivation and self-efficacy as contextual moderators in educational technology adoption. The results indicate that these personal attributes can shape engagement patterns independently of perceived usefulness or ease of use. While TAM has historically emphasized system characteristics, the present findings suggest that learner-specific psychological constructs must be considered for accurate modeling of AI adoption in education.

Furthermore, this research reinforces the pedagogical value **of** co-creation in AI-supported environments. When students actively generate questions rather than passively consuming AI-produced content, the AI system acts as a cognitive scaffold, prompting reflection, reinforcing comprehension, and promoting retrieval practice (Karpicke & Blunt, 2011). The co-creation process positions AI as a collaborative partner rather than an authoritative source, aligning with socio-constructivist learning theories and emerging frameworks for human-AI co-learning.

5.5 Limitations

Despite its contributions, this study has several limitations:

- **Sample Characteristics**: The study's participants were all MSc EduTech students from a single institution, representing a relatively homogeneous and highly skilled group. The small sample size and gender imbalance (11 female, 3 male) further limit the generalizability of the findings to more diverse learner populations.

- **Self-Report Bias**: Measures of motivation, self-efficacy, and engagement relied partly on self-report questionnaires, which are susceptible to social desirability and recall bias.
- **Evaluator Subjectivity**: Learning gains, as measured through comprehension evaluation, were assessed by a single evaluator. The absence of inter-rater reliability checks introduces potential subjectivity into the scoring process.
- **Lack of Control Group**: Without a non-AI-supported comparison group, causal claims about OwlMentor's unique contribution to engagement or learning outcomes cannot be definitively made.
- **Short Duration**: The intervention lasted only one semester, which may not have been sufficient to observe long-term changes in motivation, self-efficacy, or learning performance.

5.6 Future Research Directions

Building on the findings and limitations of the present study, several avenues for future research are recommended:

1. **Broader Participant Samples**

   Future studies should examine the use of OwlMentor with more diverse learner populations, including undergraduate students, learners from different academic disciplines, and participants from varied cultural and linguistic backgrounds. Expanding the demographic scope will help determine the tool's adaptability and generalizability.

2. **Addressing Gender and Demographic Imbalances**

   Given the gender imbalance in the current sample (11 female, 3 male), future research should aim for more balanced representation to explore potential gender-related differences in AI tool adoption, motivation, and learning outcomes.

3. **Experimental Comparisons**

   Introducing control or comparison groups will allow researchers to isolate the specific effects of AQG from other instructional or environmental factors. This would enable stronger causal inferences about OwlMentor's impact.

4. **Longitudinal Designs and Retention Measures**

   Extending the duration of interventions and incorporating delayed post-tests will allow for the measurement of long-term retention, transfer of learning, and sustained changes in motivation and self-efficacy. Such designs can help determine whether the benefits of automatic questioning persist over time.

5. **Multi-Evaluator Assessments**

   Employing multiple evaluators to score comprehension tasks and establishing inter-rater reliability will improve the validity of learning gain measures. This would reduce subjectivity and provide more robust evaluation of cognitive outcomes.

6. **Tool Comparisons and Feature Testing**

   Comparative studies involving multiple AI-based educational tools could reveal which interface designs, feedback mechanisms, and question scaffolding strategies are most effective for promoting engagement and learning across different learner profiles.

7. **Adaptive Personalization**

   Future research could explore AI systems that dynamically adjust prompts, question difficulty, and feedback based on real-time indicators of learner engagement, performance, and motivation. Adaptive personalization may optimize both cognitive and affective outcomes.

Conclusion

6.1 Restatement of the Research Problem and Aims

Scientific reading comprehension is a critical skill in higher education, yet many students approach complex academic texts passively, with limited use of metacognitive strategies such as self-questioning. Artificial intelligence (AI)–based question generation tools, such as OwlMentor, offer the potential to scaffold active engagement, promote deeper comprehension, and support self-assessment.

This study aimed to investigate how students interact with OwlMentor, the types of questions generated, the cognitive processes supported, and how learner motivation and self-efficacy influence and are influenced by chatbot use. A mixed-methods design combined questionnaire data, comprehension assessments, and system usage logs to address three primary research questions.

6.2 Summary of Key Findings

The analysis revealed that students predominantly generated multiple-choice and true/false questions, with a focus on mid-level cognitive processes ("Understand" and "Apply") rather than higher-order thinking. OwlMentor usage patterns showed that students with higher initial AI motivation tended to interact less with the chatbot, while recall performance was positively related to perceived usefulness. Learning gains measured through comprehension tests were modest and non-significant, though self-reported understanding and recall were moderately high and positively correlated. Motivation and self-efficacy levels remained stable over the course of the study, suggesting that OwlMentor may serve as a maintenance tool for these constructs rather than a catalyst for change.

6.3 Contributions of the Study

6.3.1 Theoretical Contributions

This study extends the Technology Acceptance Model (TAM) by incorporating AI motivation and self-efficacy as moderating variables, highlighting the role of learner-specific attributes in educational technology adoption. The findings also contribute to literature on

AQG by emphasizing the value of co-creation between learners and AI systems in promoting engagement and cognitive processing.

6.3.2 Practical Contributions

For educators, the results suggest that AI tools like OwlMentor can be effectively integrated into academic reading tasks to support engagement, particularly among learners with moderate or lower initial motivation. For designers, the study underscores the importance of prompt engineering and interface design in encouraging higher-order questioning.

6.3.3 Methodological Contributions

The mixed-methods approach, combining quantitative measures with system interaction logs, demonstrates a comprehensive framework for evaluating AI-based educational tools in authentic learning environments.

6.4 Final Reflections

The findings of this study offer valuable insights into the interplay between learner characteristics, AI-supported question generation, and engagement with scientific texts. While OwlMentor demonstrated potential as a cognitive scaffold, its impact on motivation and self-efficacy was limited within the short intervention period. Future research with larger, more diverse samples, extended time frames, and enhanced scaffolding strategies could further clarify the role of AI in fostering deep, sustained learning.

In an era where AI tools are rapidly becoming embedded in education, understanding how learners interact with these systems and tailoring their design accordingly remains a crucial priority for educational technology research and practice.

References

Bhat, S., Nguyen, H. A., Moore, S., Stamper, J., & Sakr, M. (2022). Towards automated generation and evaluation of questions in educational domains. *Proceedings of the 15th International Conference on Educational Data Mining*, 701–704.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Bruner, J. S. (1966). *Toward a theory of instruction*. Harvard University Press.

Chin, C., & Brown, D. E. (2000). Learning in science: A comparison of deep and surface approaches. *Journal of Research in Science Teaching, 37*(2), 109–138.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003.

Deci, E. L., & Ryan, R. M. (2000). The “what” and why of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268.

Dijkstra, R., Genç, Z., Kayal, S., & Kamps, J. (2022). Reading comprehension quiz generation using generative pretrained transformers. *iTextbooks @ AIED 2022*, 112–118.

Heilman, M., & Smith, N. A. (2010). Good question! Statistical ranking for question generation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 609–617). https://aclanthology.org/N10-1086/

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Hsu, C.-L., & Chiu, C.-M. (2004). Internet self-efficacy and electronic service acceptance. *Decision Support Systems, 38*(3), 369–381.

Kerly, A., Ellis, R., & Bull, S. (2008). CALMsystem: A conversational agent for learner modelling. *Knowledge-Based Systems, 21*(3), 238–246.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27*(1), 111–126. https://doi.org/10.1207/s15326985ep2701_8

Koşar, G., Akbana, Y. E., & Yakar, L. (2022). Development and validation of a reading self-efficacy scale. *International Journal of Assessment Tools in Education, 9*(1), 203–219.

Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2019). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education, 30*(1), 121–204. https://doi.org/10.1007/s40593-019-00186-y

Lee, M. J., Kim, S., & Yang, D. (2023). AI-supported reflection prompts for learning with text: Effects on student engagement and comprehension. *Learning and Instruction, 85*, 101655.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459–9474. https://arxiv.org/abs/2005.11401

Liu, X., Zhang, H., & Wang, J. (2020). A transformer-based approach to automatic question generation. *IEEE Access, 8*, 195359–195369. https://doi.org/10.1109/ACCESS.2020.3033022

Kim, M., Park, J., & Cho, Y. (2021). Effects of question-generation scaffolding using a chatbot on English listening comprehension and motivation. *Educational Technology Research and Development, 69*(3), 1425–1443.

Mulla, F., Shah, U., & Thomas, G. (2023). Question generation via AI helps college students formulate deeper questions during scientific reading. *Journal of Educational Computing Research, 61*(7), 1032–1054.

Oche, A. J., Folashade, A. G., Ghosal, T., & Biswas, A. (2025). A systematic review of key retrieval-augmented generation (RAG) systems: Progress, gaps, and future directions. *arXiv preprint*. https://arxiv.org/abs/2507.18910

Park, S. Y. (2009). An analysis of the technology acceptance model in understanding e-learning usage behavior. *Journal of Educational Technology & Society, 12*(3), 150–162.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of intervention studies. *Review of Educational Research, 66*(2), 181–221.

Siriwardhana, S., Weerasekera, R., Wen, E., Kaluarachchi, T., Rana, R., & Nanayakkara, S. (2023). Improving the domain adaptation of retrieval-augmented generation (RAG) models for open-domain question answering. *Transactions of the Association for Computational Linguistics, 11*, 1–17. https://doi.org/10.1162/tacla00530

Snow, C. E. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Corporation.

Thüs, D., Malone, S., & Brünken, R. (2024). Student AI question co-creation for enhancing reading comprehension. *IEEE Transactions on Learning Technologies*. Advance online publication. https://doi.org/10.1109/TLT.2023.3333439

Türkyilmaz, İ., Yıldırım, S., & Akçay, D. (2023). Student attitudes toward AI tutors: Role of curiosity and perceived utility. *Computers & Education, 189*, 104598.

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a research agenda on interventions. *Decision Sciences, 39*(2), 273–315.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science, 46*(2), 186–204.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178.

Wang, X., Liu, B., Tang, S., & Wu, L. (2023). SkillQG: Learning to generate question for reading comprehension assessment. *arXiv preprint arXiv:2305.04737*. https://arxiv.org/abs/2305.04737

Wang, Y., & Wang, L. (2019). Overconfidence in learning: Why high self-efficacy can reduce aid-seeking behavior. *Educational Technology & Society, 22*(1), 46–58.

Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Proceedings of the 26th European Conference on Information Systems*.

Wu, J. H., Tao, Y., & Yang, D. (2008). User acceptance of a ubiquitous city information service: UTAUT and cognitive theory perspective. *Computers in Human Behavior, 24*(1), 490–501.

Yu, F. Y.-N. (2009). Generative learning strategy: Generating self-explanations and its relationship with learning outcomes. *Journal of Educational Psychology, 101*(1), 128–137.

Yurt, E., & Kasarci, I. (2024). A questionnaire of artificial intelligence use motives: Investigating the connection between AI and motivation. *International Journal of Technology in Education, 7*(2).

Zhong, S., Yang, Z., & Chen, T. (2009). k-Anonymous data collection. *Information Sciences, 179*(18), 2948–2963. https://doi.org/10.1016/j.ins.2009.05.004

Zhou, Q., Zhang, N., Huang, M., & Zhu, X. (2017). Neural question generation from text: A preliminary study. In *Natural Language Processing and Chinese Computing* (pp. 662–671). Springer. https://doi.org/10.1007/978-3-319-73618-1_56

Appendix

All questionnaires used in this study are available in full at the following link: [**https://drive.google.com/drive/folders/1GpD0ag8Rh23NMa8sc54T4TZJp5FN-Hgw?usp=sharing**]

Appendix A - Prior Knowledge and Pre-Test Questionnaire (T1)

This questionnaire collected baseline data on participants' background knowledge and initial self-assessments before the intervention. It included:

- Participant code creation instructions to ensure anonymity.
- Prior knowledge questions relevant to the *Multimedia Learning II* course content.
- Self-efficacy (course) and self-efficacy (reading) scales.
- AI motivation scale.
- All items rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree)
- Collected data is found in **Dataset_piumal_extended.xlsx**

Appendix B - Mid-Semester Session Questionnaires (S1–S5)

Five short questionnaires administered after key course sessions to track engagement and comprehension. They included:

- Reading completion check ("Did you read the assigned chapter?").
- OwlMentor usage check ("Did you use OwlMentor during your learning session?").
- Self-rated comprehension and recall questions.
- Two content-based comprehension questions per session.
- Open-ended prompts for principle explanation and application.

- **S1 Session Questionnaire data**

VPN	S1_Read_Split-Attention	S1_Used_OwlMentor	S1_Understanding	S1_Recall_Key_Points	S1_Describe_Principle	S1_Apply_Principle
aan071971	1	1	5	5	Split attention principle is the principle which is applied to reduced the extraneous working load that was caused by the split attention effect. Split attention effect is the situation when the parts of learning material are separated (E.g. the triangle and its textual characteristics are separated spatially).	We can integrate different parts in one picture in order to reduce working memory load.
agv081949	1	1	4	5	According to the signaling principle, learning materials should convey all the complementary information (necessary for understandi	The verbal description which explains a chart should better segmented and integrated close to the image's

					ng and learning) in integrated format, reducing the spatial and temporal distance between this information.	section it refers to.
bhh121955	0	0	4	2	The split attention principle states, that information regarding the same construct, should be near to each other.	It cam be used for the enrichment of graphics with text e.g. text should be integrated in the image.
ddk061969	0	0	4	4	Split attention principle.	multiple ways of input integrated for easier processing
eas071960	1	1	4	4	Split attention principle occurs when one needs to integrate multiple sources of information to process it, increases cognitive load	Information should be integrated in the way that helps processing
gsr051973	0	0	3	3	Split attention principle is about presentation of information. It is important to	

					integrate the text near to the illustrations or the animations so learners can focus on one thing.	
kyj011954	1	1	4	3	Foe beneficial learning process it is better to present information in an integrated way.	
mog101963	1	1	5	5	Materials that need to be integrated on learners mind, needs to be close to each other.	Avoid unnecessary search for the learner. Reduce cognitive load associated with the learning material. Look for complete integration of information inside diagrams/videos or context.
nbn071975	1	0	4	4		
seb071960	1	1	5	4	- integrated format should be used to prevent the split attention which then prevents	Integrate the description in the pictures, AR, pop ups etc

					cognitive overload - this means that the description of the pictures should be integrated into the picture. It takes less time to make the connection compared to a description next to the picture where students have to switch back and forth - augmented reality, pop ups and color coding are also ways to reduce split attention - keep the redundancy effect in mind - the split attention effect only occurs for high element activity which means the material is complex. If its easy we have enough capacity to compensate	

					for bad instructional design and still make the connection between objects easily	
thn04 1960	1	1	4	4	Distributing information increases cognitive load.	Presenting the text and visuals together
vgl091 955	1	0	5	5	Split attention principle	To reduce extraneous load, especially for materials with high element interactivity, instructional design should maximize the contiguity of elements. (Ex. Don’t put the definitions far from the word. Don’t have narration explain after the animation has finished.)

Data Collected from the rest of the sessions can be found in

Session_Questionnaire_data.xlsx file

Appendix C - Midpoint Technology Acceptance Questionnaire (T2)

Administered midway through the course, this survey assessed:

- Perceived Ease of Use (PEOU).
- Perceived Usefulness (PU).
- Behavioral Intention to Use (BI) OwlMentor.
- Items rated on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree).
- Collected data is found in **Dataset_piumal_extended.xlsx**

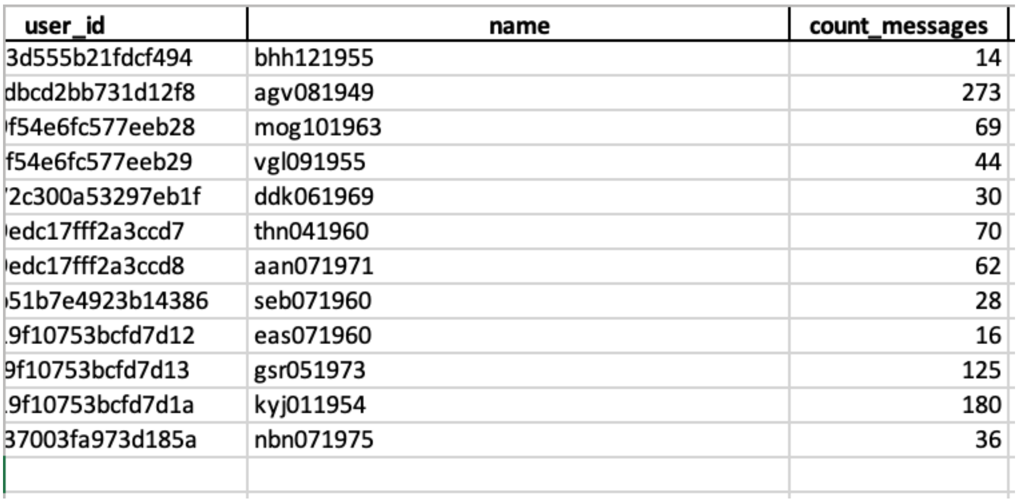
Appendix D - Post-Test and Practice Exam Questionnaire (T3)

Administered at the end of the course to evaluate final learning outcomes and perceptions. It included:

- Practice comprehension questions drawn from course readings.
- Post-test self-efficacy (course and reading) scales.
- Post-test AI motivation scale.
- Technology Acceptance Model items (PU, PEOU, BI).

Appendix E - Sample OwlMentor Interaction

- Total interactions each participant had



user_id	name	count_messages
3d555b21fdcf494	bhh121955	14
dbcd2bb731d12f8	agv081949	273
f54e6fc577eeb28	mog101963	69
f54e6fc577eeb29	vgl091955	44
2c300a53297eb1f	ddk061969	30
edc17fff2a3ccd7	thn041960	70
edc17fff2a3ccd8	aan071971	62
51b7e4923b14386	seb071960	28
9f10753bcfd7d12	eas071960	16
9f10753bcfd7d13	gsr051973	125
9f10753bcfd7d1a	kyj011954	180
37003fa973d185a	nbn071975	36

More interactions about user interactions can be found in **user_summary.xlsx**

- **Some Questions Generated and stored by participants**

User_ID	Questions_type	Difficulty Level	Question	Answer
670e56499f54e6fc577eeb28	Open-Ended	Apply	How would you apply the split-attention principle to design instructional materials for a software training program?	['In a software training program, instructional design can apply the split-attention principle by integrating text instructions directly within the on-screen demonstrations of software features, using callouts or on-screen guides to reduce the need for switching attention between different sources.']
670e56499f54e6fc577eeb28	Open-Ended	Apply	If tasked to minimize split-attention in a multimedia learning module about mechanics, what specific techniques could you employ using augmented reality (AR)?	['Use AR to overlay instructions or information onto actual mechanical components, thereby reducing the need for attention shifts between physical objects and instructional materials.']

670e56499f54e6fc577eeb28	Open-Ended	Apply	What steps would you take to use eye-tracking technology to evaluate the effectiveness of learning materials designed to reduce split-attention?	['Implement eye-tracking to monitor gaze patterns and transitions between text and visuals. Analyze collected data to adjust material placement and integration strategies based on observed cognitive load and visual flow efficiency.']
670e56499f54e6fc577eeb28	Open-Ended	Apply	In developing an online course, how would you apply the findings about extraneous cognitive load and the split-attention effect to improve the learning experience?	['Apply findings by creating integrated digital resources that embed text within video content, employ signaling to emphasize key parts, and offer learner-paced navigation to minimize cognitive overload.']
670e56499f54e6fc577eeb28	Multiple Choice	Apply	How would you apply the split-attention principle to design instructional materials for a software training program?	['Integrate text instructions directly within on-screen demonstrations using callouts.']
670e56499f54e6fc577eeb28	Multiple Choice	Apply	If tasked to minimize split-attention in a multimedia learning	['Use AR to overlay instructions on actual

			module about mechanics, what specific techniques could you employ using augmented reality (AR)?	mechanical components.']
670e56499f54e6fc577eeb28	Multiple Choice	Apply	What steps would you take to use eye-tracking technology to evaluate the effectiveness of learning materials designed to reduce split-attention?	['Analyze gaze patterns and transitions between text and visuals.']
670e56499f54e6fc577eeb28	Multiple Choice	Apply	In developing an online course, how would you apply the findings about extraneous cognitive load and the split-attention effect to improve the learning experience?	['Integrate interactive digital content with text embedded in videos.']
670e56499f54e6fc577eeb28	Multiple Choice	Apply	How can you apply the split-attention principle to enhance student engagement and comprehension when teaching a virtual lab session on assembling a complex model?	['Utilize interactive 3D models with guided instructions integrated into the visual interface.']

670e56499f54e6fc577eeb28	Multiple Choice	Understand	What is the main idea behind the split-attention principle in multimedia learning?	['It suggests that simultaneous presentation of text and visuals can lead to cognitive overload.']
670e56499f54e6fc577eeb28	Multiple Choice	Understand	How does the split-attention effect impact students' learning during multimedia presentations?	['It can hinder their ability to process information effectively.']
670e56499f54e6fc577eeb28	Multiple Choice	Understand	Which scenario best illustrates the split-attention principle in action?	['A video tutorial that integrates spoken instructions while demonstrating an assembly process.']
670e56499f54e6fc577eeb28	Multiple Choice	Understand	What are the implications of the split-attention effect for instructional design?	['It suggests that instructional materials should minimize the separation of relevant information.']
670e56499f54e6fc577eeb28	Multiple Choice	Understand	What is the significance of the split-attention principle in multimedia learning?	['It helps design effective instructional materials that minimize cognitive load.']
670e56499f54e6fc577eeb28	Multiple Choice	Apply	In developing an online course, how would you apply the findings about extraneous cognitive load and the split-attention	['Integrate interactive digital content with text embedded in videos.']

			effect to improve the learning experience?	
670e56499f54e6fc577eeb28	Multiple Choice	Understand	How does temporal integration help reduce cognitive load in multimedia learning?	['Present multiple sources simultaneously, reducing mental integration needs.']
670e56499f54e6fc577eeb28	Multiple Choice	Understand	What was a key finding of Mayer and Anderson's research comparing words-before-pictures to simultaneous presentations?	['Simultaneous presentations improve problem-solving performance.']
670e56499f54e6fc577eeb28	Multiple Choice	Apply	In what way might augmented reality technology be used to mitigate the split-attention effect in laboratory settings?	['By ensuring real-time virtual data aligns spatially with physical components.']
670e56499f54e6fc577eeb28	Multiple Choice	Apply	How do attention-direction strategies, such as color-coding or segmentation, assist in managing split-attention in multimedia resources?	['They align text with diagrams in one visual field, reducing search effort.']

More generated and stored questions can be found in **question_practice_counts.xlsx**

Appendix F – Additional Statistical Output

- All SPSS analysis datasets and outputs are included in the folder **data_analysis**

Declaration of Honor

Statement

I hereby confirm that I have written this thesis on my own and that I have not used any other media or materials other than those referred to in the thesis.

Saarbrücken ______________________ ____________________________________

(Date) (Signature)