修士学位論文

AI Writing Assistants for Supporting Post-Class Summary Note Creation

（授業後の要約ノート作成支援 **AI** アシスタントの研究）

令和 6 年度

工学系研究科電気系工学専攻融合情報学コース

37-236479

香取 浩紀

指導教員 矢谷 浩司 准教授

令和 7 年 1 月 23 日提出

Acknowledgements

Abstract

Summary note learning is rewriting notes while reviewing the lecture's content post-class. Learners can deepen their understanding and remember the content through writing. However, there are problems such as a decline in motivation and a decrease in the quality of notes because learners usually take notes alone. Many studies have pointed out that interventions during writing summary notes are essential, but previous methods required other's labor and were less accessible for interventions. In this research, I implemented a note-taking system with writing assistants using generative AI and conducted a user study to survey whether supporting summary note-taking by assistants' interventions is effective. The writing assistants generate explanations or questions based on the content in learners' summary notes and of the lecture. The user study was a between-subjects experiment with three groups - a group with the assistant that generates explanations (the explaining group), a group with the assistant that generates questions (the questioning group), and a group without assistants (the control group). They reviewed their notes and took a test one week later. I did not find significant differences in almost all the quantitative results, such as the number of characters, the quality of summary notes, and the test scores across the three groups. However, through interviews with the participants, I found positive and negative points about the writing assistants and the format of explanations and questions. Thanks to the assistants, participants noticed missing points in their notes and felt they understood the lecture content well. On the other hand, they could not concentrate on note-taking and remember the content the assistants did not men-

tion. Finally, I described the design implications of writing assistants for summary note-taking.

概要

要約ノート学習とは、授業後に授業の内容を振り返りながらノートにまとめ直す学習である。書くことを通じて理解や記憶が深まる一方、一人で作成することが多いため、モチベーションの低下、ノートの質の低下が問題となっている。そのような課題に対処するため、要約ノートの執筆中の介入が重要であると指摘されてきたが、これまでの研究は他者の労力を必要としていることが多い。そこで、本研究では、生成 AI を利用したライティングアシスタントを搭載したノートシステムを実装し、アシスタントによる介入が要約ノート作成の支援として効果的か調査するユーザスタディを実施した。ライティングアシスタントは、要約ノートに書かれた内容と講義内容をもとに説明または質問を生成する。ユーザスタディでは、アシスタントが説明を生成する説明群、アシスタントが質問を生成する質問群、アシスタントなしの対照群の被験者間実験で、参加者は同一の録画講義を視聴して要約ノートを書き、その 1 週間後にまとめノートを復習した上でテストを受験した。ノートの文字数や質、テストの点数といった多くの定量的指標で、グループ間での統計的差異は確認できなかったものの、インタビューを通じて、アシスタントそのものや説明形式と質問形式それぞれの良し悪しが明らかになった。見落としている点に気付いたり、理解が深まったという肯定的な意見が得られた一方で、ノートを書くことに集中できなかったや、アシスタントに依存し過ぎてアシスタントが言及しなかった点の理解が疎かになったといった否定的な意見も見られた。最後に、今後の要約ノート学習におけるライティングアシスタントの設計指針について述べる。

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 The importance and challenges of note-taking

Note-taking is one of the most critical processes in learning. It helps learners retain and understand the content they are studying [46]. Research shows note-taking enhances academic performance compared to non-note-takers [14, 18, 40, 51]. Previous studies revealed two reasons for the importance of note-taking for learning, and they are "encoding function" and the "storage function" [6, 30–32, 52]. The encoding function involves promoting learners' thinking and understanding through the act of note-taking itself. The storage function allows learners to use their notes as a resource for review, particularly before exams. Regarding the storage function, it is essential not only to take notes but also to actively review them [5, 18, 29, 32, 55].

There are two primary types of note-taking identified in previous research: "in-class" and "post-class" [6, 7, 11]. Because note-taking in-class is performed in an environment where learning is enforced, such as during lectures, the willingness to take notes is less affected by learners' spontaneity or autonomy. However, it requires multitasking, as learners must simultaneously listen to lectures and write notes. This increases cognitive workload and stresses learners' mental resources [25, 43]. Con-

versely, post-class note-taking allows learners to focus entirely on writing notes without the added burden of listening to a lecture, reducing cognitive demands [55]. In this thesis, I call the note taken post-class "summary note."

Because learners often write their summary notes alone without any support, the quality of notes was not desirable [6]. Also, psychological issues stemming from human nature arise, such as individuals slacking off when writing by themselves and feeling the perceived effort of creating summary notes due to the low awareness of the effectiveness of summary note creation [54, 55].

1.2 Technical support with interventions for note-taking

Many researchers have sought to address these challenges, such as learners' low motivation and the low quality of summary notes. For example, guided summary notes, where learners complete notes with prewritten titles and subtitles, have been shown to improve note quality and reduce the psychological barriers to starting [8]. Other systems support structured note-taking [36] or enable learners to refer to notes written by others [15, 16]. There is also research about collaborative writing in note-taking [17, 27, 47, 49]. These works have tried to resolve problems with summary notes caused by doing it alone by using systems support. These studies highlight the importance of interventions during the note-taking process, as emphasized by several researchers [3, 32, 55].

Interventions in learning provide interactivity and the feeling of being with someone and reduce the isolation of studying alone. Through technology and automation, interventions will be more accessible to many learners. There are studies about learning supports using technology such as AI chat-bot for reviewing [35], a chat-bot for memorizing English words [45], and a system that AI pretends to a student to train

beginner teachers [37]. AI can process long texts sophisticatedly and generate output as a natural language. These AI features will give learners insight, which they cannot notice alone, by scanning the content of lectures and summary notes. Then, integrating interactive learning technologies like AI into summary note-taking can significantly enhance learning outcomes.

1.3 The main idea of this research

In this research, I implemented a system that creates a learning cycle that note-taking assistants generate explanations or questions based on the content of the summary notes written by learners and they then utilize these explanations or answer to these questions to improve the quality of their notes. By using the questioning format, learners have to think of answers mentally. If they cannot find the correct answers, they need to search for information using resources such as the Internet or textbooks. I anticipate that AI-generated questions will stimulate learners to take proactive actions rather than directly providing answer. Indeed, prior research has demonstrated that questioning-style AI enhances critical thinking skills more effectively than explaining-style AI or a control group [13]. While there is a distinction between critical thinking and note-taking, both share common processes, such as evaluating the necessity of information and logically understanding content during summary note-taking. This is because I adopted the system's explaining and questioning interface.

Moreover, this system offers additional benefits: it does not require the presence of other learners and demands minimal effort from instructors. These advantages address the challenges of previous works about the intervention in summary note-taking. To examine the effectiveness of the system with explaining-based and question-oriented assistants in writing summary notes, I pose the following research questions:

RQ1 How do note-taking assistants affect the quantity and quality of summary notes and learning outcomes?

RQ2 How do note-taking assistants influence the process of note-taking and learners' psychology?

To answer these two research questions, I conducted the between-subjects user study with three groups (N=30): the group used explanation-oriented assistant (**explaining group**), the group used question-oriented assistant (**questioning group**), and the group wrote notes without any assistants (**control group**). In the user study, participants wrote memos without any assistants while watching a recorded video lecture. They then created summary notes based on their memos, with the explaining and questioning groups utilizing their respective assistants. One week later, they reviewed their summary notes and completed a post-test based on the lecture content. I also conducted interviews on both days. This study examines note-taking assistants' quantitative and qualitative impact on summary notes, learning outcomes, and the note-taking process. The findings inform design implications for future note-taking assistants.

Chapter 2

Related Work

2.1 Interventions using technology during note-taking

In many cases, learners independently write summary notes, which can make it challenging to receive feedback and often leads to a decline in the quality of the notes [6]. To address these issues, providing interventions during the note-taking process is critical [3, 32, 55]. Technology-based automatic and generic interventions are particularly valuable in contexts with limited access to resources, such as teachers and learning materials, because learners tend to write a summary note alone.

A substantial body of research has focused on systems designed to support collaborative note-taking [17, 27, 47, 49]. For example, there is a system in which students can add comments on shared slides during class and a platform to do discussion among students [27, 47].

In self-study environments, NoteCoStruct, a note-taking system developed by Fang et al., is particularly notable [15, 16]. This system enables learners to browse key points, explanations, and writing styles from notes created by others who attended the same lecture. It also supports the creation of structured notes, which go beyond a sequential transcription of the teacher's explanations and board writing.

Furthermore, in the domain of video lectures and note-taking, Cuong et al. developed a system that supports note-taking while watching video lectures by detecting points that users think essential through analyzing learners' gaze in real-time and adjusting playback speed [38]. Similarly, there is another system that allows users to search for videos containing content similar to their notes [48]. In terms of referring to the content written in notes, the system that generates search queries based on the content of users' notes and their search history has been shown to improve the efficiency of finding relevant information compared to manual searches [39].

Collaborative learning and systems providing access to resources created by others have proven effective However, addressing individual issues, such as discovering new information or refining personal knowledge based on the content in notes, is equally important. Despite these advancements, there remains a lack of interactive interventions tailored to individual circumstances in the field of summary note-taking.

Interventions are effective because they allow learners to reflect on their notes and learn from a meta-perspective, counteracting the self-contained nature of the summary note-writing process. In this research, I propose a system in which a note-taking assistant powered by generative AI provides explanations or questions based on the content of summary notes. By iteratively generating explanations and questions, searching for new information, reflecting, and answering questions, I verify that learners can improve both the quality and quantity of their summary notes and enhance learning outcomes.

2.2   Methods for evaluating notes

Evaluating notes and summary notes and understanding their relationship to learning outcomes is an important area of research. Previous studies have shown that there are many kinds of methods for evaluation and assessment indicators.

In terms of quantitative analysis, the number of characters and words are representative indicators [15, 55]. Combining quantitative and qualitative aspects, many studies have adopted a matrix approach that evaluates the comprehensiveness of topics covered and the depth of content for each topic [6, 8, 20, 28, 41]. In this method, we first prepare several items that are the topic of the summary note. After that, we score from 0 to 3 points based on the richness of content for each item and regard the sum of scores as the indicator of the note. For example, the criteria for each score are "There is no description of the item. (0 points)", "There is only the name of the item (1 point)", "There are descriptions of the item, but they are insufficient. (2 points)", and "There are sufficient descriptions of the topic. (3 points)".

The other study has analyzed additional elements, such as main topics, supplementary information, examples, and images [20]. Indicators such as the number of ideas, prioritization of items, clarity, readability, and accuracy of notes have also been proposed [15].

Another critical study area is the relationship between learning outcomes and note quality/quantity. For instance, the research by Flanigan et al. reported a correlation between the number of ideas in notes and test scores [19]. Chen et al. found that while note quality significantly predicted learning outcomes, quantity alone did not [6]. Inuzuka et al. analyzed five aspects of notes - (1) amount of description, (2) structuring (a concept similar to summarization of relating information, organization of information, and taking structured notes), (3) use of figures, (4) use of examples, and (5) copying text—and their relationship to test performance [55].

Overall, prior research suggests a strong relationship between note quality and learning outcomes, while the relationship between note quantity and outcomes remains unclear. In this research, I analyze the quality and quantity of notes produced by the user study participants and examine their relationship to learning outcomes.

Chapter 3

System Design

In this chapter, I describe the design, functionality, and implementation of the note-taking system that participants used in the user study. I illustrate the procedure and details of the user study in Chapter 4.

3.1 Features of the system

3.1.1 Overview of the system

I will show an example system interface with a question-oriented assistant in Figure 3.1. Participants typed their notes on the document sheet in Figure 3.1. While participants were writing their summary notes, the note-taking assistant generated explanations or questions based on the notes and the lecture content as mentioned in Section 3.1.2. Red, bold, and italicized texts in Figure 3.1 are generated questions. The assistant positioned the generated explanations and questions near the relevant sections of the notes to ensure contextual alignment. In the explaining group, the assistants provide explanations in the same way as questions regarding position, font, and color.
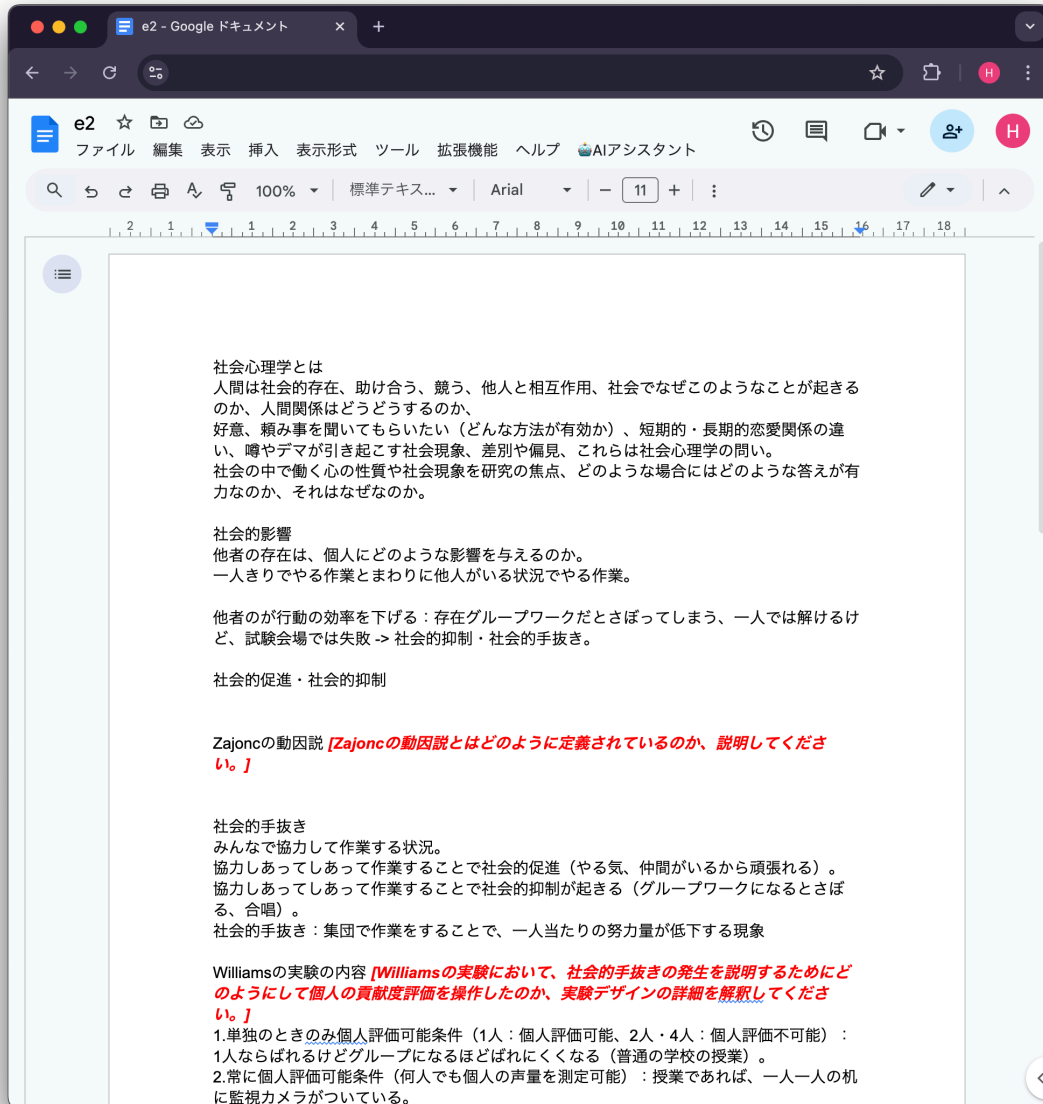
Figure. 3.1 Overview of the note-taking system with a question-oriented assistant. Participants wrote their notes on the document. Red, bold, and italicized texts are generated questions. Participants could choose to answer or ignore the questions. In the user study, the notes and the assistants' generated explanations and questions were in Japanese, as all participants were Japanese speakers.

Participants in the explaining and questioning groups had full control over handling the generated content. Specifically:

- **The explaining group**: Participants could erase, copy and paste, or partially use the generated explanations.

- **The questioning group**: Participants could choose to answer or ignore the generated questions.

However, there was one constraint for each group. I asked participants in the questioning group to delete all generated questions by the end of the summary note-writing process. Similarly, I instructed participants in the explaining group to delete generated explanations or change their color from red to black. I adopted these rules to calculate note quantity and quality accurately across the three groups. Participants in the control group did not use note-taking assistants and just wrote their summary notes directly on the document.

3.1.2 Prompt of the system

This system's core functionality is note-taking assistants' ability to generate explanations and questions. I used a generative AI API to implement this functionality. The prompt to generate explanations and questions included the content of notes, a list of previously generated explanations and questions, the transcript of the video lecture, and instructions.

I instructed assistants to generate explanations and questions based on the notes, the transcript, and the four steps of Bloom's Taxonomy. Bloom's Taxonomy is a framework that categorizes six levels of learners' understanding: Remember, Understand, Apply, Analyze, Evaluate, and Create. Create is the highest level of understanding.

I adopted this framework because it is famous and widely used in many prior research in the field of education [12, 23, 26, 42]. Although Bloom published its

original work [2] in 1956, I referred to the revised version from 2001 [1, 34]. I used expressions from prior studies that analyzed the relationship between exam questions and Bloom's Taxonomy [26]. In this system, assistants generated questions randomly from the four lower levels of Bloom's Taxonomy: Remember, Understand, Apply, and Analyze. I omitted the question regarding Evaluate and Create because these two steps are too advanced levels for writing summary notes, which aims to input and understand the lecture content through writing. The prompt also included the following instructions

- Generate a question encouraging a learner to supplement information mentioned in the transcript but missing from the notes.

- Avoid generating questions whose answers are already present in both the transcript and the notes.

- Avoid generating duplicate questions with the same content.

For the explaining group, the prompt was almost identical to that used for the questioning group. Systemically, the explanation-oriented assistant generates a question, the same as the question-oriented assistant. The difference between them is that the explanation-oriented assistant generates an answer to the question, and the system places the answer as an explanation on the note.

The following sentences are examples of explanations and questions generated by the assistants at the user study, translated from Japanese into English. The alphabetical and numeric combination following the statement represents the participant's ID. For example, "EP10" means the person who is a participant ID No.10 in the explaining group. In addition, "Q" denotes the questioning group, and "C" represents the control group.

"Social cutting refers to the phenomenon of a decrease in the amount of effort per person when working in a group. On the other hand, social inhibition refers

to a phenomenon in which the efficiency of individual actions declines in the presence of others, especially when the assignment is complex, or individuals lack the required skills. Clarifying responsibilities, enhancing task attractiveness, and boosting motivation are effective measures to prevent these problems. For example, it is recommended to avoid social cutting by specifying each person's responsibilities and establishing a system to measure individual contributions." (Explanation, step: Analyze, EP10)

「社会的手抜きは、集団で作業をする際に、一人あたりの努力量が低下する現象を指す。一方、社会的抑制は、他者の存在によって個人の行動の効率が下がる現象を指し、特に課題が複雑または熟練度が低い場合に起こる。これらを防ぐための対策として、責任の所在を明確にすることや、課題を魅力的なものとし、動機付けを高める工夫が有効である。例えば、誰がどの部分を担当しているかを明確にすることで手抜きを防ぎ、個人の貢献度を測る仕組みを整えることが推奨される。」 (Explanation, step: Analyze, EP10)

"Based on the result of the social cutting experiment by Williams et al., please explain how social cutting can be prevented in a real-world group." (Question, step: Understand, QP2)

「ウィリアムズらの社会的手抜き実験の結果を基に、どのようにすれば現実のグループで社会的手抜きを防ぐことができるか説明してください。」 (Question, step: Understand, QP2)

3.1.3 Mechanisms of generating explanations and questions

The system provides two mechanisms for generating explanations or questions. The first mechanism is the generation by user's instruction. This mechanism means the assistant generates one explanation or question when a user clicks the "Run the generation" button in Figure 3.2. The second mechanism is an automatic regular generation
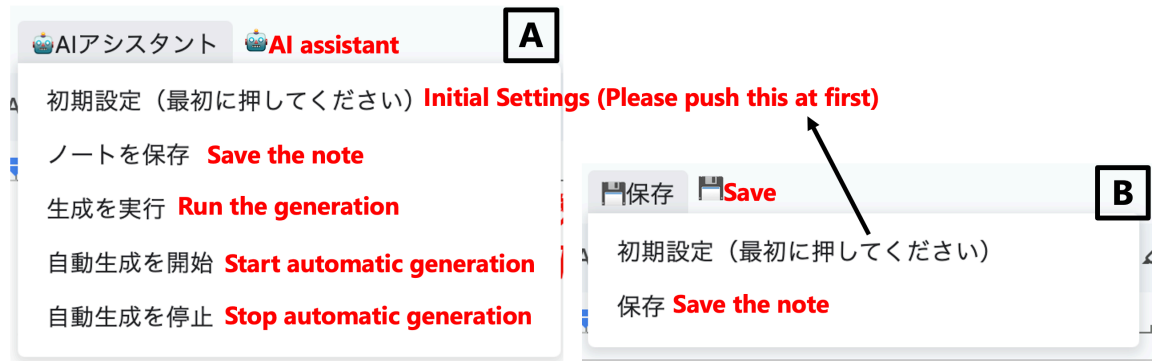
Figure. 3.2 The figure compares buttons displayed in three groups. Panel A displays the buttons for explaining and question-oriented assistants. This panel appears when a user pushes the button named "AI assistant." Panel B shows the buttons used in the control group's system. This panel appears when a user pushes the button named "Save." Red texts are translations of the names of buttons.

without the user's instruction. This feature allows the assistant to generate explanations or questions automatically at regular intervals. Users can activate this feature by clicking the "Start Automatic Generation" button and deactivate it by clicking the "Stop Automatic Generation" button in Figure 3.2. Once activated, the assistant generates one explanation or question every five minutes.

Automatic regular generation can address situations where users might forget to click the button named "Run the generation" while focusing on writing their summary notes. This feature ensures that assistants expose participants in the explaining and questioning group to generated explanations and questions in order to make a clear distinction between these groups and the control group. I determined the interval of the regular generation based on observations of user behavior before the user study.

3.2   Implementation of the system

This note-taking system uses Google Document [1] and participants write their notes on this document. I used Google Apps Script [2] to run processes of reading the content

---

[1]https://www.google.com/intl/ja_jp/docs/about/
[2]https://www.google.com/script/start/

of the note, calling the API to generate explanations and questions, and placing generated explanations and questions onto the document. For generating explanations and questions, I utilized the Assistants API [3] provided by OpenAI API [4] with GPT-4o [5]. Additionally, the system logged data, including the content of the notes and timestamps corresponding to when the system read the content and when the system placed generated explanations or questions on the notes.

---

[3] https://platform.openai.com/docs/assistants/overview/
[4] https://openai.com/index/openai-api/
[5] https://openai.com/index/hello-gpt-4o/

Chapter 4

Experimental Design

In this chapter, I present a user study using the note-taking system with note-taking assistants that I introduced and detailed in Chapter 3.

4.1 Study Procedure

I conducted a between-subjects user study with three groups: a questioning group, an explaining group, and a control group. Each group used the corresponding note-taking system described in Chapter 3. I concealed the fact that the user study was a control experiment and the true purpose of the user study from the participants. Instead, I told them the dummy purpose of the study, which was to collect data from summary notes to analyze note-taking behaviors and outcomes quantitatively and qualitatively. The deception aims to mitigate the possibility of participants favoring the questioning or explaining group in their responses or behaviors. I did a debriefing session at the end of the study to disclose the actual purpose.

I proceeded with the user study, as shown in the below lists. The user study consists of two stages and took place entirely online via one-on-one Zoom sessions. I required the participants to join both stages, and all completed them.

Stage1 User study at Day1

> **Stage1-1** The onboarding session for explaining the procedure of the user study and writing a consent form. (15 minutes)
>
> **Stage1-2** Participants listened to the video lecture about social psychology while taking notes. (30 minutes)
>
> **Stage1-3** Participants wrote a summary note by overwriting the note taken in Stage1-2. People in the explaining and questioning group used their note-taking assistants. (30minutes)
>
> **Stage1-4** The first interview. (25 minutes)

Stage2 User study at Day8

> **Stage2-1** Participants reviewed the summary note written in Stage1-3. (5 minutes)
>
> **Stage2-2** Participants took a multiple-choice test whose questions were based on the content of the video lecture in Stage1-2. (15 minutes)
>
> **Stage2-3** The second interview. (5 minutes)
>
> **Stage2-4** The offboarding session for debriefing. (5 minutes)

4.1.1 The first stage

The first stage takes approximately 100 minutes. While the duration of the onboarding session and the initial interview varied among participants, I strictly controlled and accurately timed the time allocated for watching the video lecture and writing a summary note.

At first, participants joined the onboarding session and submitted a consent form to join the user study. I also explained the flow of the user study.

Next, participants watched the 28-minute pre-recorded video at a standard play-back speed. They also took notes while watching the video, but I firmly instructed participants to concentrate on watching the video rather than note-taking.

In this phase, none of the three groups utilized the note-taking assistants.

The theme of the video lecture is an introductory class on social psychology. This lecture [1] is part of a series of psychology lectures for high school students provided by the official YouTube channel [2] of The Japanese Psychological Association [3]. The reason I selected social psychology as a theme of the user study is prior research on summary notes often utilized psychology lecture [6, 15, 55]. Additionally, the primary purposes for writing summary notes are remembering and understanding the lecture's content through writing, and subjects regarding arts, which need memorizing, are more suitable.
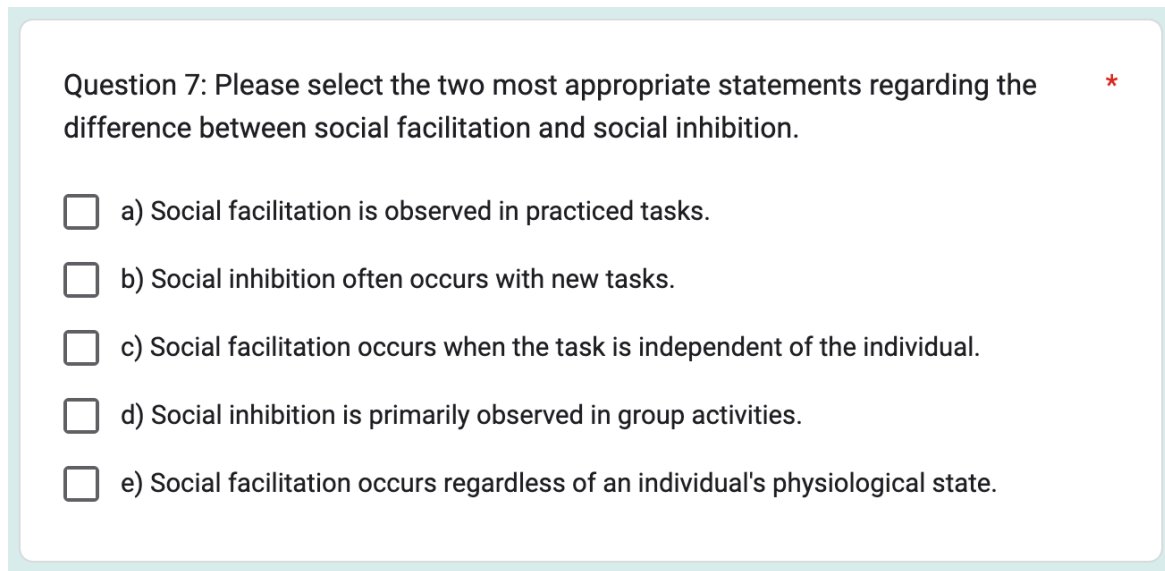
Third, participants spent 30 minutes revising and expanding their summary notes based on the notes they had taken during the lecture. Participants in the questioning and explaining groups used the corresponding note-taking assistants, while the control group did not. Instead, participants in the control group just wrote their notes on Google Docs without any AI support. Before the note-taking session began, I instructed the two groups on using the note-taking assistants and several usage guidelines. All participants can watch the video again and adjust the playback speed (pause, speed up, or slow down). I also permitted them to search for additional information online. However, I prohibited them from using any generative AI tools besides the note-taking assistants.

Finally, I conducted the first interview. This interview lasted about 25 minutes. I show the details of the interview, such as the list of questions in Section 4.3.4.

---

[1] 社会心理学入門＿真島理恵＿高校生のための心理学講座 (日本心理学会)＿1. https://www.youtube.com/watch?v=W0KzIr4YVP4

[2] https://www.youtube.com/@user-yx3ql3wt4i

[3] https://psych.or.jp/

Figure. 4.1 A question in the post-test in Stage2-2. Participants have to select two correct answers from five options. The proper options are "a" and "b". Participants solved the test written in Japanese. The question in this figure is translated into English.

4.1.2 The second stage

The second stage took place one week after the first stage, an average duration of the university classes. During this stage, I measured the time participants spent reviewing their notes and completing the test accurately.

First, participants had 5 minutes to review only their summary notes. To standardize conditions, participants were not allowed to review their summary notes or access any materials related to social psychology between the first and second stages.

Next, I assigned the test to them, and they had to submit it within 15 minutes. I forbade them from accessing any materials, including their notes. I created the test questions based on the lecture content with Google Forms. The total number of questions is 12. Each question requires participants to select multiple correct answers from several options. An example question is shown in Figure 4.1.

Following the test, I conducted the second interview in about 5 minutes. Section 4.3.4 shows the details of this interview. Lastly, I informed participants of the true purpose of the user study.

4.2 Participant

I recruited participants through CrowdWorks [4]. I asked candidates whether they had knowledge of social psychology or not. All selected participants answered either "I don't have the knowledge of social psychology completely" or "I have heard the word' social psychology' but don't know its content." They must also be fluent in Japanese and accustomed to using Google Docs and Zoom.

A total of 30 participants (15 male and 15 female) took part in the study. Of these, 6 are students, and 24 are non-students. They are all adults, with an average age of 36.4 years (SD = 9.804). I randomly assigned them to one of three groups, with 10 participants in each group. Each participant received a compensation of 2601 yen upon completing the study.

Additionally, I submitted the institutional review of this user study to the Environmental Health and Safety Office of the Graduate School of Engineering and the Graduate School of Information Science and Technology of the University of Tokyo. The office approved the review.

4.3 Measurement

4.3.1 Notes quantity

Previous studies on post-class note-taking have used the number of characters or words in notes as an evaluation metric [6, 8, 55]. In this research, I focused on two aspects:

---

[4]CrowdWorks is one of the largest crowdsourcing platforms in Japan. https://crowdworks.jp/

the total number of characters in the completed summary notes and the change in character count from the beginning to the end of the note-taking process. I compared these metrics across the three groups.

This analysis addresses the quantity of notes in RQ1, as outlined in Chapter 1. I hypothesize the following: **H1: The quantity of notes in the explaining and questioning groups will increase compared to the control group**. This hypothesis is based on two assumptions. First, participants in the explaining group will likely copy and paste generated explanations, directly increasing the character count. Second, generated questions will guide participants in the questioning group in identifying additional content to include in their notes.

4.3.2   Notes quality

In addition to quantity, another critical metric for evaluating notes is their quality. I employed a rubric evaluation, as described in Section 2.2 to assess quality. Four raters independently scored all notes, evaluating them both at the beginning and at the completion of the summary note-writing process. The raters included three Japanese first-year master's students from our lab and myself. The rubric consisted of eight items shown in Table 4.1. The raters scored each item from 0 to 3 points based on the unified criteria agreed upon by all raters. Therefore, the total score range was from 0 to 24, with a step size of 1 point. The final score for each set of notes is the average of the total scores across the four raters. I analyzed the rubric score when completing the summary note and the change in score from the beginning to the end of the note-writing process.

This analysis addresses the quality of notes as part of RQ1 in Chapter 1. My hypothesis is as follows: **H2: The quality of notes in the explaining and questioning groups will increase compared to the control group**. In the questioning group, question-oriented assistants will prompt participants to complete omitted

Table 4.1 The list of items in the video lecture for the rubric evaluation.

number	item
1	What is social psychology
2	Social facilitation
3	Social inhibition
4	Zajonc's drive theory
5	Social cutting
6	The procedure of the social cutting experiment by Williams et al.
7	The result and discussion of the social cutting experiment by Williams et al.
8	Methods to prevent social cutting

parts and stimulate deeper thinking through their questions, enhancing both the notes' comprehensiveness and depth. In the explaining group, the note-taking assistants will likely generate detailed explanations, which will improve the notes' quality.

4.3.3 Learning outcome

I conducted the post-test, described in Section 4.1.2, to evaluate participants' learning outcomes. The result of this test addresses the learning outcome component of RQ1 in Chapter 1. I hypothesize the following: **H3: The learning outcomes will be highest in the questioning group, followed by the explaining group, and lowest in the control group**. This is because people in the questioning group will achieve deeper learning due to engaging in reflective thinking and producing well-organized notes facilitated by the question-oriented assistants compared to the other two groups. Participants in the explaining group are less likely to engage in deep thinking, as they primarily copy and paste generated explanations. However, the quality of their notes, as discussed in Section 4.3.2, will be higher than that of the control group. This leads to a difference in learning outcomes between the explaining and control groups.

4.3.4   Interview

As mentioned in Section 4.1, I conducted two semi-structured interviews with each participant. The interview questions differed across the three groups (explaining, questioning, and control) to reflect the note-taking assistant's presence or absence and its form (explaining or questioning). This interview addresses RQ2 in Chapter 1. For the explaining and questioning groups, I asked about their experiences of writing notes with the note-taking assistants and the generated explanations or questions. The control group's questions focused on their habits and opinions regarding summary note-taking.

In the two interviews, I asked the following questions. "IQ1-X" represents a question in the first interview, and "IQ2-X" represents a question in the second interview. The alphabet in parentheses following the question text indicates the group for which the question was asked (E for the explaining group, Q for the questioning group, and C for the control group).

IQ1-1 What did you think, and how did you behave regarding this paragraph when writing a summary note? (Think-aloud question: A Paragraph refers to a section of the video lecture, such as the item lists for the rubric evaluation in Section 4.3.2.) (E, Q, C)

IQ1-2 What did you think, and how did you behave regarding this generated explanation or question? (Think-aloud question, Asked for every generated explanation or question.) (E, Q)

IQ1-3 What do you think about the influence of the note-taking assistant's function of generating explanations or questions on writing a summary note? (E, Q)

IQ1-4 What features would you like to have in a note-taking assistant? (E, Q)

IQ1-5 What do you think about the difference between the presence of note-taking assistants and not? Please compare your past experiences of summary note-taking and your experience with this user study. (E, Q)

IQ1-6 What are helpful and unhelpful questions, and why? (Q)

IQ1-7 Do you take notes during class? (C)

IQ1-8 Do you take notes after class? (C)

IQ1-9 What do you think about the disadvantages and challenges of writing summary notes? (C)

IQ1-10 What kind of support would you like when writing summary notes? (C)

IQ1-11 How often, for how long, and in which contexts (e.g., timing, subjects) do you write summary notes? (C)

IQ1-12 Do you prefer digital documents or paper for note-taking? What are the advantages and disadvantages of each? (C)

IQ1-13 What is the purpose of writing summary notes? (C)

IQ1-14 What do you think about the benefits of writing summary notes? (C)

IQ2-1 What is the influence of the summary notes on your understanding and learning of social psychology? (E, Q, C)

IQ2-2 What is the influence of note-taking assistants on your understanding and learning of social psychology? (E, Q)

IQ2-3 What are the points to improve your behavior when writing a summary note or using note-taking assistants? (E, Q, C)

Chapter 5

Result

5.1 Notes quantity

Figure 5.1 presents the results for note quantity across three groups. I compared the total number of characters in the completed summary notes among the three groups, applying a log transformation to the data. Shapiro's test did not show a violation of normality ($W = 0.947, p = 0.140$), and Bartlett's test confirmed homogeneity of variances ($\chi^2(2) = 0.448, p = 0.799$). With one-way ANOVA, I did not find any significant effect of the pretense or absence of an assistant and the difference of assistants' formats on the total number of characters across three conditions ($F(2, 28) = 0.662, p = 0.524 > 0.05$).

I also analyzed the change in character count from the beginning to the end of the note-taking process across three groups, applying a square root transformation to the data. Shapiro's test indicated no violation of normality ($W = 0.935, p = 0.068$), and Bartlett's test did not show a violation of homogeneity of variances ($\chi^2(2) = 4.10, p = 0.129$). A one-way ANOVA revealed a significant effect of the pretense or absence of an assistant and the difference of assistants' formats on the change in character count ($F(2, 28) = 4.824, p = 0.016 < 0.05$). A Dunnett's pairwise comparison showed a
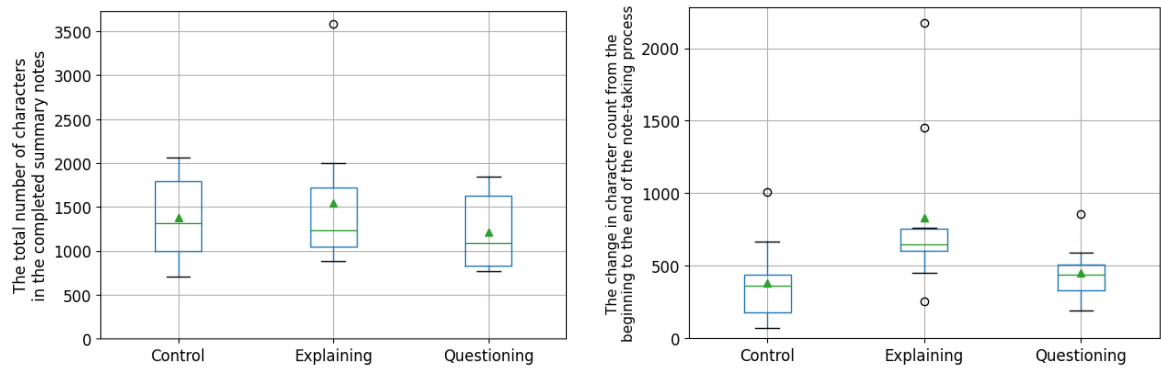
(a) The box plot of the total number of characters in the completed summary notes across the three groups.

(b) The box plot of the change in character count from the beginning to the end of the note-taking process across the three groups.

Figure. 5.1 The box plots of the number of characters of summary notes across the control, explaining, and questioning groups. A green line in the box is a median, and a green triangle is an average.

significant difference of the difference of assistants' formats on the change in character count between the explaining group and the control group ($p = 0.011 < 0.05$). Hence, H1 is partially supported based on this result.

5.2   Notes quality

I calculated the rubric scores to evaluate the quality of summary notes. Four raters individually scored all notes at the beginning and the completion of the summary note-writing process. I calculated correlation coefficients for all two pairs of four rates to confirm inter-rater reliability. A Pearson's chi-squared test revealed that p-values of all pairs were less than 0.05, indicating significant correlations among all raters.

Figure 5.2 presents the results for note quality across three groups. I compared the rubric scores, which I mentioned in Section 4.3.2, at the completion of the summary note-writing process across three groups. Shapiro's test did not show a violation of normality ($W = 0.975, p = 0.676$), and Bartlett's test did not reveal a violation of

(a) The box plot of the rubric scores at the completion of the summary note-writing process across the three groups.

(b) The box plot of the change in rubric scores from the beginning to the end of the note-writing process across the three groups.

Figure. 5.2 The box plots of rubric scores of summary notes across the control, explaining, and questioning groups. A green line in the box is a median, and a green triangle is an average.

homogeneity of variances $(\chi^2(2) = 0.370, p = 0.832)$. With one-way ANOVA, I did not find any significant effect $(F(2, 28) = 2.27, p = 0.123 > 0.05)$.

Next, I compared the change in rubric scores from the beginning to the end of the note-writing process across three groups. Shapiro's test did not reveal a violation of normality $(W = 0.930, p = 0.050)$, and Bartlett's test did not show a violation of homogeneity of variances $(\chi^2(2) = 5.38, p = 0.068)$. A one-way ANOVA did not reveal any significant effect $(F(2, 28) = 0.268, p = 0.767 > 0.05)$. Thus, H2 is not supported.

5.3   Learning outcome

The results for learning outcomes are presented in Figure 5.3. I compared the post-test scores across the three groups. Because the assumption of data normality was violated, I applied a Kruskal-Wallis H-test. This test did not reveal any significant effect $(\chi^2(2) = 1.02, p = 0.600 > 0.05)$. Therefore, H3 is not supported.
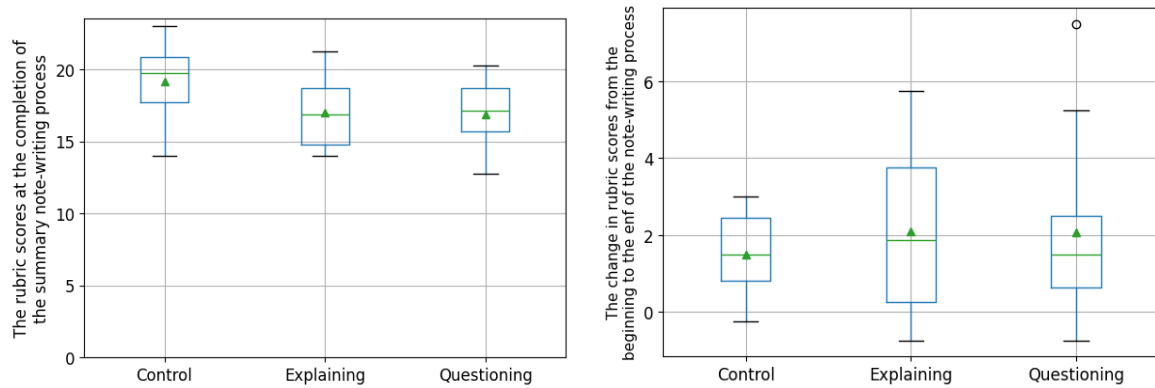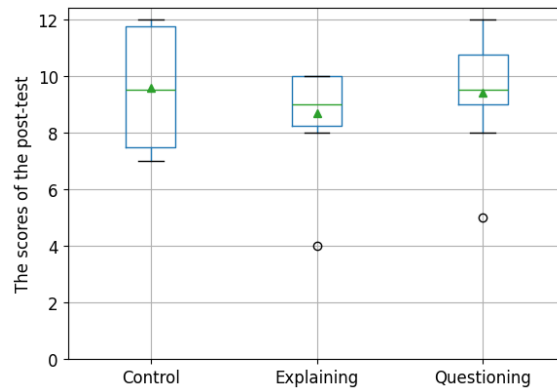
Figure. 5.3 The box plot of the post-test scores across the control, explaining, and questioning groups. A green line in the box is a median, and a green triangle is an average.

5.4   Interview

In this section, I summarize the interview script and address RQ2. The alphanumeric combinations following citations represent a participant's ID and an interview question number, as explained in Section 3.1.2 and Section 4.3.4.

5.4.1   Participants' behaviors while writing summary notes

Resources used for writing summary notes

Most participants relied on the lecture video to gather information for their summary notes. They frequently adjusted the video speed, pausing or fast-forwarding as needed. A few participants rarely referred to the video; instead, they depended on their existing notes and memory. Conversely, some participants searched for supplementary information using Google or websites like Wikipedia. The resources did not significantly differ across the three groups.

> *"The right note (the note at the end of Stage1-2 mentioned in Section 4.1) is*
> *pretty much like a memo, so I wrote my summary note by reviewing the lecture*

video once and learned points that I didn't fully understand when I watched the video at first." (CP10, IQ1-1)

「右（4.1 節の Stage1-2 終了時点でのノート）はかなりメモという感じだったので、左（4.1 節の Stage1-3 終了時点でのノート）に行くのに動画を一度ちょっと見直して、よくわからなかったところを理解してまとめたというような感じです。」（CP10, IQ1-1）

"To recall the details of the social cutting experiment by Williams et al., I watched the video again at a fast-forward speed and understood the experiment's procedure."(EP8, IQ1-1)

「実際にどんな実験が行われたかを思い出すために、もう 1 回ビデオを見直して早送りで見直して実験の手順とかを確認しました。」（EP8, IQ1-1）

"I found a website while searching for Social Cutting on the Internet. I added the content, which did not appear on the video, to my summary note referring to explanations on the site."(CP4, IQ1-1)

「ネットの方で社会的手抜きというものを検索させていただきまして、別のホームページがございましたので、そちらの解説の方を参照させていただきまして、自分なりの言葉でまとめて、動画内では出なかったものを追加させていただいたということになりますね。」（CP4, IQ1-1）

Behaviors commonly observed across the three groups

Most participants revisited the lecture video to fill gaps in their notes from Stage1-2. They recorded details they had previously missed or misunderstood. Some copied slide content directly, while others summarized it in their own words. Several participants stated that they aimed to make their notes readable during Stage2-1. They focused on the lecture's story, structured their notes with bullet points, and marked important

content using bold fonts or underlining. These practices were consistent across all groups.

> *"Basically, I wrote the content, which is a bit changed from the slides. I sometimes added spoken information not present on the slides to enhance the clarity of my note."*(CP2, IQ1-1)

> 「基本的にはスライドに書いてあることを多少アレンジして、時にはスライドに書いてない口頭部分の説明のところを分かりやすくするために加えた箇所もあります。」（CP2, IQ1-1）

> *"I organized my summary note by following the lecture's narrative from start to finish, ensuring it would be easy to review later."*(QP6, IQ1-1)

> 「話していたことを順を追って、頭から最後まで書くことになって、後から見た時に見返しやすいとか頭の中でも整理がしやすいかなというふうに思って書いてました。」（QP6, IQ1-1）

> *"I paid attention to making a summary note highlighting key points at a glance when reviewing using bold fonts and underlining."*(EP5, IQ1-1)

> 「自分でもすぐパッて要点が分かるようにですね、太文字にしたり、線を引いたりとかして、すぐ見やすいようにっていうのを心がけてノートを作っていきました。」（EP5, IQ1-1）

Behaviors specific to participants in the explaining group

In addition to the common behaviors, participants in the explaining group exhibited three distinct patterns when handling generated explanations. These included (1) leaving the explanations unchanged, (2) modifying and placing them in appropriate positions based on their judgment of necessity, and (3) completely deleting explanations, either after reviewing them or without reading them. The behaviors varied not

only between participants but also within individual participants, depending on the questions.

"I left (the generated explanation) because its content closely matched the lecture material."(EP9, IQ1-2)

「これ（生成された説明）自体は講義で全くほぼほぼ似たようなことを言ってるなっていう感じの言葉だったので、そのまままた使わせてもらおうと思ってこれをまとめたということです。」（EP9, IQ1-2）

"I completely erased the generated questions because the assistant continuously generated explanations with identical content."(EP7, IQ1-2)

「これも多分同じ内容がずっと立て続けに出ていて、なんか同じ内容だとパッと見たタイミングで思ったので、全部削除した記憶ありますね、この辺は。」（EP7, IQ1-2）

"I copied and pasted the explanation after formatting because it clearly described the social cutting experiment by Williams et al."(EP8, IQ1-2)

「実験の内容がはっきり文章に書かれていたのでこれはコピぺして体裁を整えて使って、すごく使いやすかったです。」（EP8, IQ1-2）

Behaviors specific to participants in the questioning group

I did not instruct participants to write answers to generated questions. Consequently, some participants attempted to answer the questions by re-watching the lecture video or consulting their notes and memory, while others chose not to write answers. Some of them started to summarize notes again after they read the questions.

"Because I had already written content that included the answer to this question, I skipped it."(QP10, IQ1-2)

「それも多分スルーしちゃったのかな。なんか社会的手抜きについて防ぐに
はっていうのでもう書いたからいいかなって思って。」（QP10, IQ1-2）

"After reading the question, I revised my notes by erasing some parts, adding content while watching the YouTube video, and reviewing my notes."(QP1, IQ1-2)

「もう 1 回ちょっと YouTube を見て、自分が周囲に書いてる文章を見直し
て、少し最初に書いてた文章を消して、付け加えて文章がまとまるように
書き直した。」（QP1, IQ1-2）

"I initially found this question difficult. I summarized the difference between social facilitation and social inhibition using my notes…. I wrote this section as an answer to the question."(QP9, IQ1-2)

「最初見たときは難しそうな質問だと思ったんですが、社会的促進と社会的
抑制の違いと、どうやったらそういういわゆる優勢な反応というのが出や
すくなるかというのを、下のメモからまとめればいいのかと思ってまとめ
ました。（中略）この質問に答える形の書き方ですね。」（QP9, IQ1-2）

5.4.2 Participants' positive reactions to the explanation-oriented assistant

Positive impact of the explanation-oriented assistant on learning

The explanation-oriented assistant positively influenced participants' learning in several ways. As expected, the assistant highlighted points that participants had overlooked in their summary notes and provided detailed and appropriate explanations. It offered specific definitions of terms and additional details about the lecture content. Several participants stated that the assistant was a helpful partner in their learning

process. By using the AI assistant to write summary notes, participants were also able to verify their understanding of the lecture material.

"I appreciated that the assistant generated passages I missed in my summary note."(EP5, IQ1-5)

「自分の見落としてた点だったりとかっていうのを、AIが加筆修正で文章として出してくれるのはすごいありがたいなっていうふうに思いました。」（EP5, IQ1-5）

"I was grateful for having a system other than myself to check and review my notes. This is because there was no possibility that my understanding was incorrect (after validation by the AI assistant)."(EP2, IQ1-5)

「自分以外にもう1つ何か見返すとか確認できるようなシステムがあるっていう意味ではすごいありがたかったです。その意味が間違っているかもしれないとか、理解が間違っているかもしれないっていう可能性は少なくともないかなというふうに感じたので。」（EP2, IQ1-5）

In the second interview, some participants noted that the assistant's concrete examples and detailed explanations were helpful during the post-test. Two participants mentioned that they should have utilized the assistant more effectively, such as retaining the generated explanations in their notes instead of deleting them immediately.

"The post-test included questions about examples and the experiment's content. ... (Generated) explanations (by the assistant) made it easier to review, and I was able to recall them during the test."(EP1, IQ2-2)

「実際に試験に具体例とか実験の内容を出してくるじゃないですか、（中略）その（AIが生成した）説明があったからそのノートを見返すときに見やすくて、実際に試験を解いているときにも、これあったよねっていうのは反映されますね、やっぱり。」（EP1, IQ2-2）

"From my vague memory, the AI assistant wrote the experimental conditions. I deleted them immediately, thinking they were unnecessary. But, I should have kept them in my note as instructed by the assistant."(EP8, IQ2-3)

「あと AI さんがですね、たぶんちょっとおぼろげながらなんですけど、実験の条件を書いてくれてたと思うんですよね。で、私はそれ必要ないと思ってちゃちゃっと消しちゃったんですけど、もうちょっと AI の言うことを聞いてノートに残しておけばよかったなとちょっと思いました。」（EP8, IQ2-3）

Psychological influence by the explanation-oriented assistant

Several participants perceived the AI note-taking assistant as a reliable and supportive presence. They expressed feelings of comfort and trust in the assistant. A participant mentioned trusting the generated explanations because they aligned with his memory of the lecture content immediately after watching the video. Other participants felt the assistant was another person. They stated that the assistant's focus differed from a human's perspective, making it feel like viewing someone else's notes.

"I felt (the assistant) was comforting and reliable because it made searching for information easier."

「手軽にこう調べれるなっていう、なんていうか、安心感っていうか頼りにしたいなっていう気持ちはあった。」（EP3, IQ1-3）

"I think there is quite a difference between listening to a lecture and taking notes on your own versus doing so in a group with others attending the same lecture. It was like looking at another person's notes."(EP10, IQ1-5)

「違いだと、自分一人で書く場所、自分一人で講義を聞いて書く場所と、他人が同じ講義を受けて書く場所っていうのは結構違うと思うんですけど、人のノートを見てる気分でした。」（EP10, IQ1-5）

Convenience of the explanation-oriented assistant

Several participants highlighted the time-saving advantages of the explanation-oriented assistant. They do not have to copy and paste words manually when writing a note because the AI assistant generates them automatically. Participants noted that summarizing by reducing the generated content was quicker than creating notes from scratch.

> *"I usually copy content at length and summarize it when taking notes. ... Thanks to the AI assistant, the copying process was completed automatically. ... My working hours have become very short, and it is very efficient."*(EP9, IQ2-2)

> 「僕もノートとかだと結構たらたら書き写して、たらたら書いたものを自分で結構要約するみたいな感じになってるんですけど、（中略）、AIのおかげで書き写しがもうされてしまってるので、（中略）、作業という時間ではすごく短くなるので、すごく効率はいいのかなとは感じました。」（EP9, IQ2-2）

> *"It (summarizing a note by reducing explanations generated by the AI) is time-saving compared to typing words to make notes from scratch. This time-saving aspect was a great benefit for me."*(EP4, IQ1-5)

> 「そっち（AIシステムがポンと入れてくれたやつを削ってもうちょっとスマートにしてまとめること）の方が、最初からまとめる言葉を打っていくよりは、入った言葉をまとめていくのがすごくやりやすくて時間の短縮にもなるな。時間の短縮っていうのはすごく大きいなと思いました。」（EP4, IQ1-5）

5.4.3 Participants' negative reactions to the explanation-oriented assistant

Distractiveness of the explanation-oriented assistant

Several participants were disappointed with the explanation-oriented assistants. They found it distracting as they frequently had to erase generated explanations. They felt that the assistant was redundant, as it generated explanations that repeated content already written or previously generated. Although some participants initially appreciated the assistant's convenience and helpfulness, they developed negative opinions after encountering repetitive explanations.

"The note-taking assistant frequently generated explanations with content that had already been written (in my summary note). I felt I did not need such behaviors."(EP8, IQ1-3)

「（まとめノートに）書かれていることを繰り返していること（説明を生成すること）がちょいちょいあったのでそれはいらないなって思いました。」（EP8, IQ1-3）

"At first, I thought the assistant was helpful as it summarized my summary note as a memo. However, as I had almost finished making the note in the later stages, it repeatedly generated similar explanations. Moreover, it wrote full sentences while I used bullet points for clarity. To be honest, I found the assistant a bit inconvenient."(EP5, IQ1-3)

「作り始めの時とかは、自分がメモ書きで書いたものをまとめてくれてたのでありがたいなって思ったんですけど、後半になってもうまとめ作業が終わるにつれて、同じことの繰り返しを、自分がわかりやすいように箇条書きで書いているのに文章で書かれて、正直なところちょっと邪魔だなって思いました。」（EP5, IQ1-3）

Negative opinions on the content and format of the generated explanations

Some participants became skeptical about the accuracy of the generated explanations. They began verifying the content of the explanations by cross-referencing it with the lecture video. To distinguish AI-generated explanations from their notes when reviewing, they separated them.

> *"I felt slightly uncertain about the assistant's accuracy since the generated explanations were not content that I wrote referring to something."*(EP10, IQ1-5)

> 「アシスタントを使ってて思ったこととしては、アシスタントが書いたことは自分が何かを見て書いたものではないので、本当に合っているのか不安になるっていうのはちょっとありました。」（EP10, IQ1-5）

> *"I left the top and bottom brackets to indicate that the AI created this part, not me."*(EP1, IQ1-3)

> 「上と下の括弧はわずかに残してますよね。そこは AI によって（作られたもので）、自分の分じゃねえよっていうのは、自分で見て分かるようにしてますね。」（EP1, IQ1-3）

Additionally, some participants criticized the length of the generated text, noting that it was time-consuming to read and evaluate within a limited timeframe. A participant mentioned that the text format was difficult to understand.

> *"Although I appreciated the assistant in the early stages of writing my summary note, I found it time-consuming to judge whether lengthy explanations should be kept or discarded, as time was limited."*(EP5, IQ1-3)

> 「今回まとめノートを作成にするにあたって時間制限があったので、長い文章で生成されたものを置かれると、読んで扱えそうだなっていう取捨選択をしなきゃいけない時間がちょっともったいないなあって、特に後半です

ね。前半はもうありがたいって思ったんですけど、と思いました。」（EP5、
IQ1-3）

"The long texts generated by the AI were not easy for everyone to understand during the review."(EP6, IQ1-3)

「AI が作ってくるんで、だらーっと文になってくるんで、見返した時に誰が
見ても分かりやすいっていう感じではないんですよね。」（EP6, IQ1-3）

Relationship between the generated explanations and the sense of learning

I asked questions about the influence of the summary note and the explanation-oriented assistant after participants finished solving the post-test as listed in Section 4.3.4. Several participants left negative comments to the assistant, reflecting on their notes and test results. They highlighted the differences between the assistant's explanations and their notes, stating that the AI-generated content felt objective and detached, making it harder to remember and understand. Two participants likened it to reading a textbook.

"(The AI-generated explanations) had no connection to my memory (of the lecture). I felt like I was reading a textbook, as the explanations' style differed from my summary note. This made the process of understanding less smooth."(EP2, IQ2-2)

「やっぱり（AI の生成した文章が）自分で書いた印象というか記憶と全然繋
がらないというのと、それこそ文章の感じとかがやっぱり自分のメモとか
自分が書いたものとは違うのでひたすら教科書を読まされているような感
じ、理解とか納得まではそこまでスムーズではなかったのかなと思います。」
（EP2, IQ2-2）

"(The AI-generated passages) felt objective because I did not write them."(EP7, IQ2-2)

「（AIアシスタントが生成した部分は、）自分で作った分じゃないから、客観
的な印象しかなくて。」（EP7, IQ2-2）

Most participants copied and pasted generated explanations, as noted in Section
5.4.1. However, the method of copying and pasting influenced their sense of compre-
hension. Explanations that participants merely copied and pasted without modifica-
tion were more complicated to recall compared to those they actively reviewed and
positioned appropriately in their notes.

*"The parts I wrote myself were easier to remember. In contrast, the sections
generated by the AI and copied without modification felt like something I might
have heard before."*(EP2, IQ2-1)

「やっぱり自分で書いたところとかは思い出しやすかったですし、AIアシス
タントでダーッと書いた文章をコピペしたところは、例えば聞いたかな、ぐ
らいの感じでした。」（EP2, IQ2-1）

*"The parts I extracted and positioned appropriately were easy to recall. How-
ever, the explanations I just copied and pasted had little effect on my ability to
remember the content."*(EP2, IQ2-2)

「自分で切り張りしたりとかした部分はやっぱり思い出しますが、そのまま
まるっと貼った文に関してはそこまで思い出すとか記憶につながるかとい
うと、そうでもないかなという感じでした。」（EP2, IQ2-2）

Inaccuracies and missing information in the summary notes

The AI note-taking assistant also affected the scope of the participants' reviews. In
the second interview, several participants mentioned discovering missing or incorrectly
summarized parts in their notes after the post-test and chose incorrect options in the
test. Other participants noted that they had focused too much on minor details while

neglecting more critical points. Additionally, the assistant lacked the ability to identify inaccuracies in the learners' notes.

> *"Regarding social psychology, I wondered, 'Was this part really mentioned in the lecture?' I realized I had forgotten this part."*(EP7, IQ2-1)

> 「その社会心理学とはっていうところが、なんかこんなことあったっけなっていうような感じで、結構抜けていたなって感じでした。」（EP7, IQ2-1）

> *"I felt that the assistant lacked the functionality to point out parts where I had an incorrect understanding."*(EP1, IQ2-1)

> 「自分が間違って考えてることを指摘してくれるのが（アシスタントには）何もないっていうのはちょっと感じましたね。」（EP1, IQ2-1）

5.4.4 Participants' positive reactions to the question-oriented assistant

Positive impact of the question-oriented assistant on learning

I observed that the question-oriented assistant helped participants identify missing parts in their summary notes, as expected. However, only one participant explicitly reported this effect.

> *"Not all of it, but the generated questions made me realize areas where I needed to add more details to my summary note."*(QP4, IQ1-3)

> 「全部が全部ではないですけど、（質問は）気付きになる。ここもうちょっと書いといた方がいいのかなとかっていうところの気付きになりました。」（QP4, IQ1-3）

The Questions generated by the assistant also supported participants' understanding and memory retention. Several participants stated that answering these questions could deepen their understanding of the lecture's content by watching the video and reviewing their notes and memory. This process helped them understand the content more accurately and recall it effectively. In the second interview at Stage2-3, some participants mentioned that they could easily recall the content related to the questions during the review phase in Stage2-1, which helped them answer the post-test questions.

"(The Generated questions) were one of the factors that deepened my understanding. All red characters (questions) prompted me to transform sections with just words into complete sentences."(QP1, IQ1-3)

「（生成された質問は）自分的には掘り下げれる一要因になったかなと思ってます。それぞれの赤字の箇所で、まあ最初自分的になんか単語とかだけ書いてあったんですけど、そういった赤字のサポートがあった上でそれを文章化できる要因、きっかけになったかなといったのが全般的、全ての赤字に対してそう思いました。」（QP1, IQ1-3）

"Because the assignment or question came soon from the AI assistant, ..., I found it impressive that my knowledge retention improved through reviewing (my summary note), thinking, and answering."(QP2, IQ1-5)

「今回は AI アシスタントによって、タイムリーで課題あるいは質問が来るので、（中略）再度ノートを振り返り、見ることであったりだとか自分の中で思考する、考える、答えを出すっていう過程において、その知識の定着がより深まったっていうところが非常に印象深かったところですね。」（QP2, IQ1-5）

"Sections where the AI asked questions and I summarized by myself were easier to understand than the memos I had written below. I felt that (these sections) were easier to understand."(QP7, IQ2-2)

「AI が聞いてくれて自分が実際まとめた部分の方が、その下の自分が書いてるメモよりもやっぱり分かりやすい気がしましたね。改めて頭に入ってきやすいっていう感じはありました。」（QP7, IQ2-2）

Some participants viewed the questions as opportunities to enhance their skills in creating summary notes and writing outputs. One participant noted that the questions revealed difficulty articulating answers, even when they thought their notes were complete.

"I found (the note-taking assistant) very convenient. Questions like 'What is this?' forced me to think and output answers, which I realized was beneficial because I sometimes could not write an output when asked again, even if I felt like I had finished summarizing."(QP7, IQ1-3)

「これ（ライティングアシスタント）なんかすごい僕便利だなと思いましたね。やっぱり自分でまとめた気になってても、改めて問われるとなんかうまくアウトプットできなかったりっていうことがあるので、強制的にこれ何ですかって突っ込んでくれた方がアウトプットさせられるので、自分のためになるのかなっていう気はしましたね。」（QP7, IQ1-3）

"The questions encouraged me to think profoundly and summarize the content in my mind."(QP3, IQ2-2)

「（アシスタントが生成した質問は）頭の中で文章をこう考えて、さらに自分の中でまとめる、まとめられるっていう影響があったと思います。」（QP3, IQ2-2）

The assistant also helped participants improve their notes. By answering the questions, participants created notes that were clearer, more detailed, and easier for others to understand.

"Answering questions that asked for objective opinions or clarification and were generated automatically made my summary note clearer and more understandable."(QP1, IQ1-5)

「自動生成があると、客観的な意見とか客観的な問いかけとかわかりやすさとか、そういうのを問いかけてくれるので、結構それを返すと相手に、第三者に見やすいものができるのかなっていうのが思ったことです。」（QP1, IQ1-5）

"The assistant's support improved my note summarization."(QP7, IQ1-5)

「アシスタントしてくれた方がより良いまとめになるのかなと思いました。」（QP7, IQ1-5）

Psychological influence by the question-oriented assistant

Some participants reported psychological benefits from using the question-oriented assistant. One participant stated that the assistant motivated them to write better summary notes, while another felt the assistant provided a supportive presence.

"The experience (of using the AI assistant) motivated me to use it again for taking a summary note because I gained an achievement that I could write a better note by using it. The great merit is lowering the effort and barriers to writing notes because I can write better by simply answering the assistant's questions."(QP7, IQ1-5)

「まずこれ（AI アシスタントを使うこと）を 1 回経験しとくことで、AI にアシスタントしてもらったらまとめがより良くっていうような実績が今でき

たので、次にまとめた時にちょっと一発試しに AI 使ってやるかっていうま
ず動機付けになるのと、向こうから問いかけてくれるんで、それに対して答
えるだけでそれなりに仕上がっていくので、ハードルの低さ、よっしゃまと
めるぞ一人でやるぞっていうのよりは、スタートダッシュも違うし、やる時
の手間も下がるっていうのがすごい利点かなと思いますね。」（QP7, IQ1-5）

"The assistant felt like a guide, highlighting points to focus on for the post-test. It provided valuable support in creating near-perfect notes."(QP4, IQ2-2)

「要は先週やった時に、AI アシスタントが問いかけをしてくれて、来週のた
めにここをやっとくといいよみたいなイメージとにかくメモを完璧にする、
（中略）、サポートをしてくれたっていう位置づけです。」（QP4, IQ2-2）

5.4.5 Participants' negative reactions to the question-oriented assistant

Low quality of the generated questions

The low quality of generated questions is one of the problems. Participants reported having a negative impression of the note-taking assistant because it often generated questions that had already appeared once, and several questions seemed to be unimportant and unrelated to the lecture. They also stated that the questions were complex and advanced.

"The number of automatically generated questions increased toward the end (of Stage1-3). Because (the assistant) generated similar, unrelated, and unimportant questions, I felt slightly confused and panicked due to the 30-minute time constraint."(QP3, IQ1-3)

「（まとめノートを書く時間の）最後の方は自動生成の方も増えて、似たよう
な質問とか、関係ないであろう、あまり重要でないと思うものとかが出て

きたので、ちょっと混乱したというか。30 分という制約もあったんで、そこでちょっと焦ったっていうのもありますね。」（QP3, IQ1-3）

"I found the questions requiring passage writing generated by the AI assistant difficult."(QP3, IQ1-3)

「AI のアシスタントを使うことで、文章問題、自分で文章を書いて回答する問題があり得るんだなっていうので、ちょっとより難しさを感じました。」（QP3, IQ1-3）

"While (the assistant) helped deepen my understanding of the lecture, it did not directly help me perform well on the post-test. Because of the AI-generated advanced questions, I could deepen my understanding (of the lecture). However, I do not think the assistant directly helped me get a high score (on the test) as the test lacked descriptive questions."(QP5, IQ2-2)

「学習には役に立った気がしているんですが、結果的にテストにはあまりいらなかったなという感覚があって。AI が出してくれたのが発展問題につながるような質問が多かったので、理解はあれのおかげで深まったかもなとは思うんですけど。実際のテストが記述問題とかもないので、そんなに直接点が取れる作用っていう印象は受けてない。」（QP5, IQ2-2）

Distractiveness of the question-oriented assistant

Several participants found the assistant distracting due to two primary factors: obstructiveness and psychological burden. They perceived it as obstructive because it generated questions abruptly, disrupting participants' focus on writing notes, watching the video, or thinking. The assistant cut their concentration because it directly wrote questions on their notes with noticeable red characters.

"This is entirely my opinion, but I found (the assistant) slightly inconvenient. It was not pleasant because (it) interrupted me while I was trying to write what I wanted."(QP8, IQ1-3)

「これは完全に私の意見というか感じ方なんですけど、(ライティングアシスタントは)ちょっと邪魔かなって思っちゃいました。自分の考えというか、書きたいことがあって書いてる中で、横から横槍入れられてるような感じ方をしてしまって、あんまり嬉しいものではなかったかもしれないです。」(QP8, IQ1-3)

"I don't perceive (the note-taking assistant) as having contributed. The reason for this perception is that it seemed inconvenient and hindered my thinking process last week. I wouldn't describe the assistant as user-friendly."(QP6, IQ2-2)

「(ライティングアシスタントは)特に貢献したという認識は持っていないですね。理由としましては、先週もあまり私、この AI アシスタントがどちらかというと邪魔だなというか、考察を抑えられる、そのような印象がありまして、あまり使い勝手がいいなというふうに思っていないですね。」(QP6, IQ2-2)

Participants also experienced a psychological burden from the assistant. The dual pressure of completing their summary notes within the 30-minute time limit in Stage1-3 and feeling compelled to answer the generated questions created stress. At the same time, when they found several generated questions, they felt the assistant was rushing them to answer the questions and write notes. These two factors gave participants a sense of distractiveness, and they felt the assistant disrupted their learning pace, understanding, and writing notes.

"This time, I felt a psychological burden because I had to take my note under time pressure within a 30-minute limit."(QP8, IQ1-5)

「今回は 30 分っていう時間がある中で、時間に追われながらやらなきゃい
けなかったので、ちょっと心理的な負担というか。」（QP8, IQ1-5）

"When I scrolled down while I was summarizing at my own pace, I encoun-
tered several pending questions, which made me feel rushed to address them
immediately."(QP9, IQ1-3)

「自分のペースでまとめているときにふと下とかにスクロールしたりすると、
質問が複数、2 個とかかたまってしまったりすると、すぐに対応しなきゃいけ
ないなとか思いました。」（QP9, IQ1-3）

Inaccuracies and missing information in the summary notes

A participant noted that the assistant failed to generate questions about key lecture
topics. This omission led to gaps in participants' understanding of these areas.

"Because the AI assistant did not generate questions regarding Zajonc's drive
theory and the social cutting experiment by Williams et al., I think it contributed
to my unclear understanding of these topics."(QP2, IQ2-2)

「具体的にそのなんか、ザイアンスとか、あとウィリアムズですね。そのあ
たりがちょっと具体的に AI の方で質問もなかったもので、その理解がちょ
っと曖昧だったなっていうふうには思いますね。」（QP2, IQ2-2）

5.4.6 Improvements for note-taking assistants in summary note-taking

To identify areas for improvement in the note-taking assistants, I asked participants
from both the explaining and questioning groups the interview question IQ1-5 (as
described in Section 4.3.4) during the first interview.

The explanation-oriented assistant

Participants provided several suggestions for improving the explanation-oriented assistant. Regarding the content and format of explanations, one participant proposed adding sources for the generated explanations to enhance credibility. Others suggested that the assistant present explanations in bullet points instead of long text, making it easier to identify key points. Additional requests included automatic organization of note layouts, autonomous note generation, and a feature to generate explanations corresponding to a section where a user clicked in the note.

> *"Since I prefer notes in bullet points that focus on key points, I would find explanations formatted as bullet points more helpful than lengthy passages."*(EP5, IQ1-5)

> 「私の場合はなんですけど、文章で出されるよりも要点を絞って箇条書きで書いておきたいタイプなので、文章で出されるよりも箇条書きで要点をまとめて出してもらった方が嬉しいなって思いますね。」（EP5, IQ1-5）

> *"While the AI is intelligent, it occasionally makes mistakes. … I would like a feature that adds sources to the generated explanations to improve reliability."*(EP8, IQ1-5)

> 「AI って賢いですけどちょいちょい間違ったりするので、その元となる情報、（中略）、どこから取ってきたのかっていうのが分かると、その AI が書いた文章の信頼性が増してそのまま使えるなって私の場合は思うので、そういう機能がいいなと思いました。」（EP8, IQ1-5）

The question-oriented assistant

Similar suggestions raised for the explanation-oriented assistant, such as automatic layout organization and autonomous generation of explanations, passages, and notes,

were also mentioned by participants in the questioning group. A unique request from this group was for the AI to correct learners' answers to generated questions. Participants also suggested features for notifications when generating questions and pointing out simple mistakes.

"I wrote several answers, and it would be helpful if the AI could provide feedback, such as asking 'What does this mean?' and offering corrections to my answers."(QP2, IQ1-5)

「答えを私の方でいくつか書きましたけども、これに対する添削機能とかもあったら、これはどういう意味ですかとか、ここはという形でさらに添削というか、AI の方でご助言いただけるようなシステムがあったら助かるかなというふうに思いました。」（QP2, IQ1-5）

Chapter 6

Discussion

6.1 Discussion of the qualitative result

Except for the change in character count from the beginning to the end of the note-taking process, there is no significant difference in the qualitative result. Therefore, it is difficult to state a conclusion for whether using a writing assistant in summary-note taking is effective or not and whether the question-oriented format or the explanation-oriented format is better.

I expected that the explaining group's total number of characters in the completed summary notes and its change would be more significant than the other two groups because participants in the explaining group could copy and paste generated sentences. However, the statistical result was partially different from my expectations. Regarding the total characters, this result stems from the fact that some people found the explanations unnecessary and deleted them. In contrast, regarding the change in character count, several participants left generated explanations in their notes without examining their content closely. The cost of writing characters was relatively lower than other groups' participants, and they wrote more sentences.
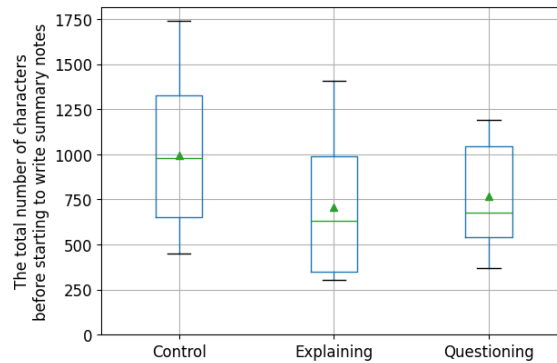
Figure. 6.1 The box plot of the total number of characters before starting to write a summary (at the end of Stage 1-2)notes across the control, explaining, and questioning groups. A green line in the box is a median, and a green triangle is an average.

I found a statistical difference between the explaining group and the control group in the change in character, not the total number of characters. Scarcity of the total number of characters at the end of Stage1-2 in the explaining group could be one of the factors of the statistical difference as shown in the Figure 6.1. It is possible that participants in the explaining group and the questioning group thought that they did not have to take notes seriously during Stage1-2 due to the existence of the writing assistants.

Regarding the quality of notes, I could not find a significant difference. From Figure5.2a, all three groups got high scores (the maximum score is 24 points). This implies that participants improved their notes to a plateau level across all groups. As the reason for this phenomenon, I found that note-taking time, including Stage1-2 and Stage1-3, was slightly longer for making summary notes.

In the research by Danry et al. about the AI assistant supporting people's judging based on logical thinking, it reported that the accuracy of judgment was highest in the order of the questioning assistant group, the explaining assistant group, and no assistant group, and there are significant differences between the three groups [13]. However, in this research, there is no significant difference in the scores on the test across the three groups. One possible reason for this is that short-term memory

strongly affected the score because they took the test after just 5 minutes of review at Stage 2- 1. The fact that there is no statistical difference in the quantity and quality of notes and all participants reviewed almost the same quantity and quality summary notes could support this reason. In a previous study where participants watched a short lecture video, wrote notes using a system, and took a test immediately afterward, no statistically significant differences were also observed between the groups [15].

6.2 Human-AI relationships in summary note-taking

Past studies warned about the harmfulness of over-use or over-reliance on AI technology [4, 9]. Kobiella et al. reported that the use of ChatGPT caused problems such as a decline of ownership, a decrease in sense of challenges, and unremarkable achievements to a part of knowledge workers [33]. I also observed humans' dependency on AI and its adverse effects from participant behaviors during note-taking and the interview script. In both the explaining and questioning groups, several participants could not understand and remember parts that the AI assistants did not mention. This result is similar to the findings that AI-generated images limited the number of ideas, the range of variety, and the originality in a visual ideation task [50]. Several participants in the explaining group unquestioningly believed in generated explanations and left them on their notes. This phenomenon was also observed, especially in participants with low cognitive reflection in the research by Danry et al. [13]. Some participants in the questioning group stated that they wanted the assistant to automatically generate explanations or notes.

These behaviors represent people's harmful dependency on AI. Past studies showed AI assistants with biased models influenced writes' output [24, 44], and I could observe the strength of AI's influence on people's writing and cognition also in this research. From the observation of participants, not questioning the truth or validity of AI's explanations is the risk that AI spreads misinformation, as warned in a previous work [9].

Additionally, these passive attitudes, such as only copying and pasting text without thinking, are not preferable for learning. In contrast, a questioning style does not have these problems and is one of the solutions to the challenge: human dependency on AI.

6.3   Design implication

A major concern participants reported was that the assistant distracted them, mentioned in Section 5.4.3 and Section 5.4.5. Many participants stated that the assistant, who inserted once-generated explanations or questions that were already written in notes, was not only not helpful but also interfered because it cut their concentration on writing notes. The assistant must continue to generate new explanations or questions for users so as not to frustrate them and to improve their motivation. Comprehensiveness is also an essential feature for designing writing assistants for summary note-taking. Participants in both the explaining and questioning groups mentioned that the explanations and questions' bias toward specific parts of the lecture limited their learning scopes. Although it is unavoidable to tolerate dependency on AI, attributing human nature to some extent, the AI assistant at least should not cause harm to learners.

The interface is one of the critical factors when designing assistants. The assistant directly generated explanations and questions on participants' notes, frustrating several participants. This is because the assistant hindered them from focusing on writing notes and organizing their notes neatly in their own way.

Displaying explanations and questions on the side of a page, like the comment function of Google Documents, is effective. Additionally, several participants overlooked a generation of explanations and questions. Therefore, we should consider displaying notifications in a way that is not intrusive but still noticeable at the time of generation.

I found several points that need improvement from the two interviews regarding the explanation-oriented assistant. Because long and complicated explanations de-

mand participants' high cognitive cost [15], Participants preferred explanations with short sentences, bullet points, and simple language. Participants tended to copy and paste generated explanations without deeper thinking and reported that this was a meaningless activity for learning. In another opinion, some participants wanted the source of generated explanations because they were suspicious of AI's accuracy. Therefore, if we adopt the explanation format in summary note-taking assistant, we have to add features to prevent unquestioningly trusting AI's explanation and ensure learners remain active in their learning process.

Regarding the question-oriented assistant, we have to adjust the questions' difficulty and labor to answer and the number of questions. Questions that are too difficult and give learners high cognitive cost to answer are unsuitable for sustaining learners' motivation. There is also an opinion that the assistant was threatening. Thus, it is not desirable to generate questions mechanically one after another. The writing assistant needs to adjust the learners' speed and abilities.

People naturally tend to take the easy way out. However, learning needs cognitive cost. Both human nature and the burden of learning are in the trade-off relationship. Also, in summary-note taking, we must balance convenience and dependency and the learning effect through others' perspectives and enhance motivation through AI writing assistants. Imposing tasks with appropriate cognitive cost on learners is effective [10, 21, 53] and I believe this also applies to the writing assistant for making summary notes. It will also be effective to think of a situation forcing learners to use the writing assistant or giving a reward and heighten the learners' understanding of the effectiveness of summary note-taking and utilizing assistants [55].

6.4 Limitations

I could not find significant differences in the quantitative indexes across the three groups. I cannot conclude that the writing assistant does not positively or negatively

influence learning productivity. There is a possibility that the settings of the user study or defects of the assistant, such as generating similar explanations or questions, are one of the causes for the qualitative result. There are improvement points about the user study, such as that the video lecture was for high school students and was easy for adult participants, the length of the time of note-taking was long, the effect of reviewing notes just before the test had a strong impact, and the number of participants was limited. The writing assistant also has improvement points, such as generating similar explanations or questions and the unnatural characteristic of generating at regular intervals.

I should measure more results such as abilities such as memorizing and other subjects, as well as the long-term effects and cognitive loads. The test only asked selecting questions and measured participants' knowledge and understanding of the lecture content. However, there are other abilities that learners should acquire. We should assess these abilities at the test and so on based on a framework for learning such as Bloom's Taxonomy [1, 2, 26, 34]. Learning is not only for art subjects; there may be different results if writing assistants are used for scientific subjects such as mathematics and biology. Moreover, learners usually need to spend a lot of time mastering one subject and succeeding in examinations for admissions and qualifications. It is essential to measure the long-term effects of change in learning output. In this research, I mainly focused on the output of learning and learning outcomes. However, sustaining the motivation of learners and cognitive loads, which is one of the factors declining motivation, is also essential. We should quantify motivations and cognitive burden using summary note-taking assistants such as NASA-TLX [22].

6.5   Future work

In the future, I want to conduct the user study for measuring indexes described in Section 6.4 using writing assistants for summary note-taking with notifications based

on the items mentioned in Section 6.3. The ideal image of the relationship between human learners and the AI writing assistant is collaboration. Human learners should not rely too much on the assistant and gain merits. I hope the positive cycle of summary note-taking with assistants that people feel the merits, such as completing high-quality notes, getting good scores on the tests, and gaining a sense of accomplishment in learning by using the assistant, and then they spontaneously utilize them.

Chapter 7

Conclusion

I developed a note-taking system with writing assistants for summary note-taking, which generates explanations or questions based on the learners' notes and the lecture content. I conducted the between-subjects user study with three groups (the explaining group, the questioning group, and the control group). The user study mimicked a lecture: participants watched the same video lecture, took summary notes using the system, reviewed them, and took the test to assess the learning outcome one week after writing summary notes.

Almost all qualitative results, such as the number of characters, the quality of notes, and the learning outcome, did not show significant differences across the three groups. However, from the interview scripts of the participants, I found the positive and negative features of the writing assistants and the format of explaining and questioning. They noticed the missing points in their notes and their understanding of the lecture due to explanations and questions by the assistants, as I expected. In contrast, the assistant gave several participants an impression of distractiveness and narrowed down the learners' scope of learning. I observed people's dependency on AI as a negative side but found its positive effects on learning.

Finally, I proposed the design implication of writing assistants for summary note-taking based on the analysis of the interview. The ideal design of writing assistants is leveraging the usefulness of AI while controlling the negative aspects, such as dependency in summary note writing.

Publications

Publication during the Master's program

- 香取浩紀, 山本恒輔, 佐藤安理紗ジエンジエラ, 矢谷浩司. アルゴリズムとデータ構造に関するコードデータのアノテーションとその分析. DICOMO2023, 2023 年 7 月.

Planned publication during the Master's program

- 香取浩紀, 楊期蘭, 矢谷浩司. 授業後の要約ノート作成支援アシスタントの設計と評価. 情報処理学会全国大会. 2025 年 3 月.

- 澤野令, 香取浩紀, 矢谷浩司. 文章読解中における類義語学習を目的とした質問または説明ベースのマイクロタスクの効果検証. 情報処理学会全国大会. 2025 年 3 月.

References

[1] Lorin W Anderson and David R Krathwohl. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition.* Addison Wesley Longman, Inc.

[2] Benjamin S. Bloom. 1956. *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain.* Addison Wesley Longman, Inc.

[3] Janice M Bonner and William G Holliday. 2006. How college science students engage in note-taking strategies. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching* 43, 8 (2006), 786–818.

[4] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (April 2021), 21 pages. https://doi.org/10.1145/3449287

[5] John F Carter and Nicholas H Van Matre. 1975. Note taking versus note having. *Journal of Educational Psychology* 67, 6 (1975), 900.

[6] Pin-Hwa Chen. 2013. The effects of college students'in-class and after-class lecture note-taking on academic performance. *The Asia-Pacific Education Researcher* 22 (2013), 173–180.

[7] Pin-Hwa Chen. 2021. In-class and after-class lecture note-taking strategies. *Active Learning in Higher Education* 22, 3 (2021), 245–260.

[8] Pin-Hwa Chen, Timothy Teo, and Mingming Zhou. 2017. Effects of guided notes on enhancing college students' lecture note-taking quality and learning performance. *Current Psychology* 36 (2017), 719–732.

[9] Valerie Chen, Q. Vera Liao, Jennifer Wortman Vaughan, and Gagan Bansal. 2023. Understanding the Role of Human Intuition on Reliance in Human-AI Decision-Making with Explanations. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 370 (Oct. 2023), 32 pages. https://doi.org/10.1145/3610219

[10] Andy Cockburn, Per Ola Kristensson, Jason Alexander, and Shumin Zhai. 2007. Hard lessons: effort-inducing interfaces benefit spatial learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07).* Association for Computing Machinery, New York, NY, USA, 1571‑1580. https://doi.org/10.1145/1240624.1240863

[11] Salomé Cojean and Manon Grand. 2024. Note-taking by university students on paper or a computer: Strategies during initial note-taking and revision. *British Journal of Educational Psychology* 94, 2 (2024), 557–570.

[12] Alison Crowe, Clarissa Dirks, and Mary Pat Wenderoth. 2008. Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE— Life Sciences Education* 7, 4 (2008), 368–381.

[13] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. 2023. Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 352, 13 pages. https://doi.org/10.1145/3544548.3580672

[14] Gilles O Einstein, Joy Morris, and Susan Smith. 1985. Note-taking, individual differences, and memory for lecture information. *Journal of Educational psychology* 77, 5 (1985), 522.

[15] Jingchao Fang, Yanhao Wang, Chi-Lan Yang, Ching Liu, and Hao-Chuan Wang. 2022. Understanding the Effects of Structured Note-taking Systems for Video-based Learners in Individual and Social Learning Contexts. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP, Article 21 (jan 2022), 21 pages. https://doi.org/10.1145/3492840

[16] Jingchao Fang, Yanhao Wang, Chi-Lan Yang, and Hao-Chuan Wang. 2021. Note-CoStruct: Powering Online Learners with Socially Scaffolded Note Taking and Sharing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 223, 5 pages. https://doi.org/10.1145/3411763.3451694

[17] Mik Fanguy, Matthew Baldwin, Evgeniia Shmeleva, Kyungmee Lee, and Jamie Costley. 2023. How collaboration influences the effect of note-taking on writing performance and recall of contents. *Interactive Learning Environments* 31, 7 (2023), 4057–4071.

[18] Judith L Fisher and Mary B Harris. 1973. Effect of note taking and review on recall. *Journal of Educational Psychology* 65, 3 (1973), 321.

[19] Abraham E Flanigan, Kenneth A Kiewra, Junrong Lu, and Dzhovid Dzhuraev. 2023. Computer versus longhand note taking: Influence of revision. *Instructional Science* 51, 2 (2023), 251–284.

[20] Abraham E Flanigan and Scott Titsworth. 2020. The impact of digital distraction on lecture note taking and student learning. *Instructional Science* 48, 5 (2020), 495–524.

[21] Tovi Grossman, Pierre Dragicevic, and Ravin Balakrishnan. 2007. Strategies for accelerating on-line learning of hotkeys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1591‑1600. https://doi.org/10.1145/1240624.1240865

[22] SG Hart. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload/Elsevier* (1988).

[23] Karin K Hess, Ben S Jones, Dennis Carlock, and John R Walkup. 2009. Cognitive Rigor: Blending the Strengths of Bloom's Taxonomy and Webb's Depth of Knowledge to Enhance Classroom-Level Processes. *Online Submission* (2009).

[24] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 111, 15 pages. https://doi.org/10.1145/3544548.3581196

[25] Renée S Jansen, Daniel Lakens, and Wijnand A IJsselsteijn. 2017. An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review* 22 (2017), 223–233.

[26] Karl O. Jones, Janice Harland, Juliet M.V. Reid, and Rebecca Bartlett. 2009. Relationship between examination questions and bloom's taxonomy. In *2009 39th IEEE Frontiers in Education Conference*. 1–6. https://doi.org/10.1109/FIE.2009.5350598

[27] Matthew Kam, Jingtao Wang, Alastair Iles, Eric Tse, Jane Chiu, Daniel Glaser, Orna Tarshish, and John Canny. 2005. Livenotes: a system for cooperative and augmented note-taking in lectures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Association for Computing Machinery, New York, NY, USA, 531‑540. https://doi.org/10.1145/1054972.1055046

[28] Anam Ahmad Khan, Sadia Nawaz, Joshua Newn, Ryan M. Kelly, Jason M. Lodge, James Bailey, and Eduardo Velloso. 2022. To Type or to Speak? The Effect of Input Modality on Text Understanding during Note-Taking. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 164, 15 pages. https://doi.org/10.1145/3491102.3501974

[29] Kenneth A Kiewra. 1985. Investigating notetaking and review: A depth of processing alternative. *Educational psychologist* 20, 1 (1985), 23–32.

[30] Kenneth A Kiewra. 1989. A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review* 1 (1989), 147–172.

[31] Kenneth A Kiewra, Nelson F DuBois, David Christian, Anne McShane, Michelle Meyerhoffer, and David Roskelley. 1991. Note-taking functions and techniques. *Journal of educational psychology* 83, 2 (1991), 240.

[32] Keiichi Kobayashi. 2006. Combined Effects of Note-Taking/-Reviewing on Learning and the Enhancement through Interventions: A meta-analytic review. *Educational Psychology* 26, 3 (2006), 459–477. https://doi.org/10.1080/01443410500342070 arXiv:https://doi.org/10.1080/01443410500342070

[33] Charlotte Kobiella, Yarhy Said Flores López, Franz Waltenberger, Fiona Draxler, and Albrecht Schmidt. 2024. "If the Machine Is As Good As Me, Then What Use Am I?" ‑How the Use of ChatGPT Changes Young Professionals' Perception of Productivity and Accomplishment. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1018, 16 pages. https://doi.org/10.1145/3613904.3641964

[34] David R Krathwohl. 2002. A revision of Bloom's taxonomy: An overview. *Theory into practice* 41, 4 (2002), 212–218.

[35] Yen-Fen Lee, Gwo-Jen Hwang, and Pei-Ying Chen. 2022. Impacts of an AI-based cha bot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. *Educational technology research and development* 70, 5 (2022), 1843–1865.

[36] Ching (Jean) Liu, Chi-Lan Yang, Joseph Jay Williams, and Hao-Chuan Wang. 2019. NoteStruct: Scaffolding Note-taking while Learning from Online Videos. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1‑6. https://doi.org/10.1145/3290607.3312878

[37] Julia M. Markel, Steven G. Opferman, James A. Landay, and Chris Piech. 2023. GPTeach: Interactive TA Training with GPT-based Students. In *Proceedings of the Tenth ACM Conference on Learning @ Scale (L@S '23)*. Association for Computing Machinery, New York, NY, USA, 226‑236. https://doi.org/10.1145/3573051.3593393

[38] Cuong Nguyen and Feng Liu. 2016. Gaze-based Notetaking for Learning from Lecture Videos. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2093‑2097. https://doi.org/10.1145/2858036.2858137

[39] Srishti Palani, Zijian Ding, Austin Nguyen, Andrew Chuang, Stephen MacNeil, and Steven P. Dow. 2021. CoNotate: Suggesting Queries Based on Notes Promotes Knowledge Discovery. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 726, 14 pages. https://doi.org/10.1145/3411764.3445618

[40] Richard J Peper and Richard E Mayer. 1986. Generative effects of note-taking during science lectures. *Journal of Educational psychology* 78, 1 (1986), 34.

[41] Stephen T Peverly, Vivek Ramaswamy, Cindy Brown, James Sumowski, Moona Alidoost, and Joanna Garner. 2007. What predicts skill in lecture note taking? *Journal of Educational Psychology* 99, 1 (2007), 167.

[42] Mary J Pickard et al. 2007. The new Bloom's taxonomy: An overview for family and consumer sciences. *Journal of Family and Consumer Sciences Education* 25, 1 (2007).

[43] Annie Piolat, Thierry Olive, and Ronald T Kellogg. 2005. Cognitive effort during note taking. *Applied cognitive psychology* 19, 3 (2005), 291–312.

[44] Ritika Poddar, Rashmi Sinha, Mor Naaman, and Maurice Jakesch. 2023. AI Writing Assistants Influence Topic Choice in Self-Presentation. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 29, 6 pages. https://doi.org/10.1145/3544549.3585893

[45] Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L. Murnane, Emma Brunskill, and James A. Landay. 2019. QuizBot: A Dialogue-based Adaptive Learning System for Factual Knowledge. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1‐13. https://doi.org/10.1145/3290605.3300587

[46] Issa I Salame and Ashley Thompson. 2020. Students' Views on Strategic Note-Taking and Its Impact on Performance, Achievement, and Learning. *International Journal of Instruction* 13, 2 (2020), 1–16.

[47] Beth Simon, Krista Davis, William G. Griswold, Michael Kelly, and Roshni Malani. 2008. Noteblogging: taking note taking public. *SIGCSE Bull.* 40, 1 (mar 2008), 417‐421. https://doi.org/10.1145/1352322.1352278

[48] Ranjitha Jaddigadde Srinivasa, Samuel Dodson, Kyoungwon Seo, Dongwook Yoon, and Sidney Fels. 2021. NoteLink: A Point-and-Shoot Linking Interface between Students' Handwritten Notebooks and Instructional Videos. In *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 140–149.

[49] Jan van Aalst, John Kamimura, and Carol KK Chan. 2017. Exploring collaborative aspects of knowledge building through collaborative summary notes. In *Computer Supported Collaborative Learning 2005*. Routledge, 697–706.

[50] Samangi Wadinambiarachchi, Ryan M. Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 380, 18 pages. https://doi.org/10.1145/3613904.3642919

[51] Andrea Weiland and Steven J Kingsbury. 1979. Immediate and delayed recall of lecture material as a function of note taking. *The Journal of Educational Research* 72, 4 (1979), 228–230.

[52] Amber E. Witherby and Sarah K. Tauber. 2019. The Current Status of Students' Note-Taking: Why and How Do Students Take Notes? *Journal of Applied Research in Memory and Cognition* 8, 2 (2019), 139–153. https://doi.org/10.1016/j.jarmac.2019.04.002

[53] Koji Yatani, Zefan Sramek, and Chi-Lan Yang. 2024. AI as Extraherics: Fostering Higher-order Thinking Skills in Human-AI Interaction. (2024). arXiv:cs.HC/2409.09218 https://arxiv.org/abs/2409.09218

[54] 犬塚美輪 and 三浦麻子. 2019. 復習のためのノート作成は内容理解を向上させるか. 日本教育心理学会総会発表論文集 61 (2019), 292. https://doi.org/10.20587/pamjaep.61.0_292

[55] 犬塚美輪, 三浦麻子, and 小川洋和. 2022. 大学生が作成する事後ノートの質とテスト成績の関連. 教育心理学研究 70, 1 (2022), 35–47. https://doi.org/10.5926/jjep.70.35