인지과학박사 학위논문

A Time for Better Discussions: Focus on Self-study and Discussion Combination

자율학습과 토론 조합을 통한 학습 효과 향상 방안

2023 년 8 월

서울대학교 대학원

협동과정 인지과학전공

임 재 서

A Time for Better Discussions: Focus on Self-study and Discussion Combination

자율학습과 토론 조합을 통한 학습 효과 향상 방안

지도교수 박 주 용

이 논문을 인지과학박사 학위논문으로 제출함

2023 년 7 월

서울대학교 대학원

협동과정 인지과학전공

임 재 서

임재서의 인지과학박사 학위논문을 인준함

2023 년 6 월

위 원 장 Sowon Hahn (인)

부위원장 장 병 탁 (인)

위 원 이 희 승 (인)

위 원 정 수 정 (인)

위 원 박 주 용 (인)

Abstract

A Time for Better Discussions: Focus on Self-study and Discussion Combination

Jaeseo Lim
Interdisciplinary Program in Cognitive Science
The Graduate School
Seoul National University

Not enough discussions are taking place in Korean classrooms. Despite numerous studies showing their educational effectiveness, traditional lectures dominate class time from elementary school to university. However, since there is abundant evidence showing the limitations of lectures, changes are needed to reduce lectures and utilize discussions in classes. In this context, this thesis explores how to utilize discussions in class. In particular, a series of studies were conducted to examine what preparations lead to effective discussions and to generalize the results.

In Chapter 1, three groups of undergraduate students were compared: lecture and review, lecture and discussion, and self-study and discussion. Law material was used to minimize the influence of prior knowledge. The results from the three experiments showed that the discussion group scored significantly higher than the review group. Furthermore, the self-study and discussion group performed better than the lecture and discussion group. Analyses of discussion content revealed that discussions of the self-study group were more constructive and productive than those of the lecture group. These results seem to be ob-

tained because the difficulties and questions during the self-study were resolved in the follow-up discussion.

In Chapter 2, three experiments were conducted with different academic domains and age groups to generalize the benefits of combining self-study and discussion. First, in two experiments undergraduate students studied science-related topics. The result showed that the self-study and discussion group scored significantly higher than the lecture and discussion group, which was superior to the lecture and review group. The combination of self-study and discussion was also superior to the other combinations in terms of gains in learning. In order to further generalize the findings to high school students, the law materials as in Chapter 1 were used. The result showed that high school students in the self-study and discussion group outperformed those in lecture and discussion group in both tests taken immediately after learning or one week later. In short, the self-study and discussion group showed the highest performance compared to other groups, regardless of age or academic subject.

In Chapter 3, in order to examine whether instructor intervention is needed in the discussion, discussion formats were manipulated at two levels; student-led and teacher-led. Each was combined with lecture or self-study as an individual preparation before the discussion. The results showed that the highest scores were observed in the self-study and student-led discussion condition, followed by those in the self-study and teacher-led discussion condition, those in the lecture and teacher-led discussion condition, and those in the lecture and student-led discussion, in descending order.

Analyses of discussion content revealed that discussions of the self-study group were more constructive and productive than the lecture group. Analyses of the observed levels of cognitive-behavioural engagement during the discussions showed that compared to the lecture group, the self-study group was more

engaged in the discussion, and this engagement remained high in the student-led discussions but gradually decreased in the teacher-led discussions. On the other hand, the lecture groups were relatively more engaged in teacher-led discussions, but they were less engaged in student-led discussions. These results provide a rationale for encouraging more student-led discussions after self-study, with a few exceptions.

This thesis has two major contributions to learning science. One is that the self-study and discussion combination enhances learning compared to other combinations. This finding is highly practical in that this combination can be directly applied to classes as an effective alternative to lectures. Another contribution is that it sets a new direction for learning science research. Although my focus was on the self-study and discussions, there is room to explore the effects of other combinations, such as the combination of writing and discussion, or even the triple combination of self-study, writing, and discussion. I hope that the follow-up studies will shed a new light on better learning and education.

Keywords: Discussion, Self-study, Combination, Individual Preparation, Lecture

Student Number: 2019-39236

Contents

List of Tables

List of Figures

Introduction

It is well known that the more students actively participate in learning, the greater the learning effect (Bonwell & Eison, 1991; Chi & Wylie, 2014). However, traditional lectures, where an instructor stands in front of a class, delivers a talk, and shows slides while students listen and take notes, still dominate in the classroom (e.g., Stains et al., 2018). The problem-based learning (PBL) is used in some medical schools, as an instructional method that can replace traditional lectures. Research has shown that PBL, where students form teams to build the necessary knowledge while solving authentic problems, is superior to traditional lectures (Strobel & Barneveld 2009); Schmidt et al., 2010). However, PBL is limited in that it requires a great deal of preparation and sustained support for it to be successful (Hung, 2011).

Park (2017) proposed a writing and discussion-oriented class management method as an alternative to lectures or PBL. In this method, students read assigned material before class and write a one-page essay on a given topic related to the material, which they upload on a web-based system. Students are also asked to post questions or express difficulties they encountered while reading the material or writing the essay. Then they are randomly assigned 3-6 essays written by their peers and are asked to grade the essays using a provided rubric. Thus, before coming to class, students read the assigned material, write an essay, and grade the essays of their peers, which significantly reduces the need for lectures. The saved class time can be allocated to discussions to address the questions posted by the students. After the class, the comments and grades from other students are sent to the author of the essay, who then rates the helpfulness of the comments provided. All these activities are conducted anonymously, managed by a computerized system, and all results are recorded as data. This method allows students to prepare for more in-depth learning, while also letting the instructor assess each student's readiness.

Then, is this new instructional method really effective? Because the activities involved in this method are so diverse, this thesis intends to limit the scope of the question to self-study and discussion. These two activities take place at the individual and group level, respectively, and by combining them, the strengths of each can be integrated. As Murphy et al. (2009) pointed out, discussion alone does not lead to good results; for the discussion to be effective, the preparatory work is needed. Among possible preparatory activities, self-study was chosen because it is the most basic individualistic learning activity. I defined self-study as individual or solitary learning to understand or master the learning material. It can include various activities like reading and re-reading, rehearsing, self-explaining, taking notes, or drawing concept maps etc.

The goal of this thesis is to determine whether the combination of self-study and discussion leads to better learning outcomes compared to other possible combinations of learning activities. It consists of three chapters: i) investigating the learning effect of discussion and the advantages of the self-study and discussion combination, ii) examining the generalizability of the self-study and discussion combination upon learning, and iii) comparing the effectiveness of student-led and teacher-led discussions. In addition to test scores, the quality of discussion was compared through analysis of conversation content.

Chapter 1

1.1 Self-study Enhances the Learning Effect of Discussions

1.1.1 Introduction

Academic learning in higher education requires diverse activities such as reading texts, listening to lectures, holding discussions with others, and solving related problems and tasks. These learning activities are complex, and each has its research agenda, such as processing models and underlying mechanisms (see Cho et al., 2015; Dunlosky & Rawson, 2019; Mayer & Alexander, 2011; Moje et al., 2020 for a general overview). For example, reading texts and listening to lectures provide students with organized knowledge; discussions enable them to share their understanding or to generate alternative interpretations of knowledge; and solving related problems is used to assess students' knowledge and understanding. Thus, except for classes that require direct performance, such as experiments or practices, it is important to balance and align these activities based on the learning material's content and the student's characteristics.

Surprisingly, however, there are relatively few studies on the effect of the combination of these activities on learning outcomes. The purpose of combining various learning activities is to enhance the learning effect.

A seminal learning sciences study found that some sequences of activities are better than other sequences and that transfer of learning is the litmus test in comparing the effectiveness of different sequences of learning (Schwartz &

Bransford, 1998). In their Experiment 3, Schwartz and Bransford divided participants into three groups. The first group was told to summarize the given material before listening to a lecture. The second group was told to analyze the material twice, and the last group was told to analyze the material before listening to a lecture. Afterward, all three groups were given a true-false recognition test and a transfer task. They found that although there was no difference in the recognition test performance among the three groups, in the transfer task, the group which analyzed the material before listening to a lecture performed significantly better than the other two groups.

Other studies have also found that unique combinations of learning activities promote student learning, such as the combination of an invention activity and the availability of a worked example (Schwartz & Martin, 2004), studying alone beforehand and then collaborating, (Lam & Muldner, 2017; Tsovaltzi et al., 2015) and generative preparation before collaboration (Lam & Kapur, 2018; Mende et al., 2021). In addition, students who were taught after the problem-solving phase demonstrated significantly greater conceptual understanding and ability to transfer to novel problems than those who were taught first and then solved problems (Kapur, 2014).

We investigated a two-part instructional sequence. In the first part of the sequence, we considered self-study. Self-study is a good candidate to replace traditional lectures. In the present study, we defined self-study as individual or solitary learning to understand or master the learning material which is selected by capable others such as teachers or instructors. Self-study can include activities such as reading and re-reading, rehearsing, self-explaining, taking notes, or drawing a concept map etc. Still, it does not have direct instruction, such as lectures.

Finding replacements for the traditional didactic lecture is important because the learning outcome of lectures is relatively low, despite their popularity (e.g., Deslauriers et al., 2019; Stains et al., 2018). Moreover, Bonawitz et al. (2011) showed that direct instruction could limit spontaneous exploration and discovery. Thus, while lectures are helpful but not crucial for learning (e.g., Bonawitz et al., 2011; Poh et al., 2010; Schmidt et al., 2010), self-study can provide an opportunity for learners to consolidate learning better or integrate new knowledge (e.g., Fiorella & Mayer, 2016). Self-study is also meaningful because it involves agency yet serves as a basis for cooperative learning, such as discussions with others (e.g., Fiorella & Mayer, 2016; Lam & Kapur, 2017; Lam & Muldner, 2017; Mende et al., 2021). Students often find it difficult to learn cooperatively (Andrews & Rapp, 2015). This difficulty can be alleviated through prior self-study because self-study activates prior knowledge, exposes knowledge gaps, and facilitates engagement.

The second part of our instructional sequence is discussions. Discussions are a common activity in the classroom, and there is a great deal of evidence that they are beneficial for learning (e.g., Murphy et al., 2009, 2018; Soter et al., 2008). Murphy et al. (2009), for example, performed a meta-analysis of 42 papers published between 1964 and 2003 and found that high-level comprehension, critical thinking, reasoning and argumentation of text improved with discussions among elementary and junior high school.

The attractiveness of discussion is that successful discussions can be one of the most effective learning activities. Chi and colleagues (Chi, 2009; Chi & Wylie, 2014; Menekse et al., 2013) have emphasized interactive learning activities such as discussions. They proposed the Interactive-Constructive-Active-Passive (ICAP) framework and found that academic achievement was lowest at P, then increased at A, C, and I in ascending order.

Of course, productive discussions do not take place naturally in the classroom. As McNamara et al. (1996) suggested, students should possess a basic understanding of the learning material. Murphy et al. (2016) also argued that an effective discussion requires the participants to be informed of relevant content beforehand and to engage in significant interactions during the discussion. We postulated that self-study, rather than lectures, can serve as the basis for productive discussions. To verify this assumption, we first checked whether self-study combined with discussion showed better learning outcomes than when a lecture was combined with discussion. Then we tried to see if the difference was due to the content of the discussion. Analysis of discussion content was carried out by Chi et al. (2017), which will be described later.

In order to examine the effect of discussion on learning, we also compared the combination of watching a lecture and discussion with that of watching a lecture and review. Reviewing is one of the most common activities used by students after a lecture and before a test. Even though reviewing may seem similar to self-study, we assumed that it is more likely a passive activity because participants were likely to be already affected by the lecture, similar to the kindergarteners who watched demonstrations in Bonawitz et al. (2011)'s study. However, there is no reason to hold discussions if student-led discussions are not more effective than students' reviewing. Thus, we wanted to check whether student-led discussion leads to better learning than reviewing.

To compare learning outcomes across different combinations, assessments incorporating tests, essay writing, complex problem solving, or interpersonal skills can be used (e.g., Schmidt et al., 2009). In this study, a test of learning content was used, including transfer and memory problems. Thus, the test questions were made up of a mix up of verbatim, paraphrased, and transfer items.

Our hypotheses are as follows:

- Hypothesis 1: Self-study combined with discussion will produce better learning outcomes measured by a mix of verbatim, paraphrased, and transfer items than watching a video lecture combined with discussion.
- Hypothesis 2: As the second part of the instructional sequence, discussion groups will score higher than the review group.
- Hypothesis 3: One mechanism of enhanced learning will be greater activity and productivity, as measured by coding the types of student utterances during the discussion, which is preceded by self-study.

1.1.2 Experiment 1

In Experiment 1, participants were divided into three groups: reviewing after watching a video lecture (Lecture Review: LR), discussion after watching a video lecture (Lecture Discussion: LD), and discussion after self-study (Self-Study Discussion: SD). The final test scores of the three groups were compared. Across all three experiments, we will use a value of $p < .05$ to judge significance. We predicted that the combination of self-study and discussion would produce learning outcomes significantly higher than watching a lecture and discussion (Hypothesis 1). To examine the learning effect of discussions, we also compared the LR group and the LD group (Hypothesis 2). Thus, we predicted that learning performance would be higher in ascending order of LR, LD, and SD.

Method

Participants and design

Undergraduate students at a selective university in Seoul participated in this experiment. All experiments were approved by the university's Institutional Review Board (IRB, No. 2004/003-017). The majors of undergraduate students included psychology, economics, physics, computer engineering, etc., and there was no significant difference in the number of students by major. Ninety students signed up, but two left during the experiment, thus leaving 37 men and 51 women. All the participants were Korean. The participants were randomly assigned to three different groups: LR (n = 30), LD (n = 29, G, number of discussion groups = 9), and SD (n = 29, G = 9). Three or four students formed a discussion group. There were no significant differences in age among the groups: LR group ($M_{age} = 20.47$, $SD_{age} = 1.94$), LD group ($M_{age} = 20.31$, $SD_{age} = 2.05$), and SD group ($M_{age} = 19.69$, $SD_{age} = 1.53$).

Materials and procedure

Background knowledge (Beyer, 1987; Miyake & Norman, 1979) can have an influence on test scores. Therefore, we used topics that covered the criminal procedure code, accusation and charge for our learning material (see, **Appendix A**). The topic was chosen because it was less likely to be affected by background knowledge since the courses on law are not available as an undergraduate course in South Korea. Even so, to make sure that the participants did not have significant background knowledge, they were given a background knowledge survey. The survey used a seven-point Likert scale, 1 (having no knowledge) to 7 (having expert knowledge) over the six topics, including the criminal procedure code, accusation and charge (two topics relevant to this study), and the genome project, civil law, the legalization of same-sex marriage and the Special Act on Sexual Violence (four topics irrelevant to this study).

After the background survey was completed, the participants were randomly assigned to three different groups. Different groups of students conducted learning in different locations, and there was one experimenter in each setting. All groups received the same written instructional material after the survey. The written instructional material was seven pages long and contained information covered in the video lecture. However, it was not a transcription of the video clip. The instructional video clip used in the experiment corresponded to lecture-style monologue sessions. The video clip was a lecture on law available on the internet. The video clip was 18 minutes long.

The LR group watched the 18-minute video lecture with the provided written instructional material and subsequently studied on their own for 18 minutes with only the written material. During review, students were not able to ask the instructor questions and were told to prepare for a test based on the video lecture and learning materials. Participants could take notes or underline content in the written material while taking a lecture, reviewing, self-studying on their own or discussing. The LD group watched the video lecture with the instructional material and subsequently discussed for 18 minutes in groups of three or four. While watching the video clip, the LR and the LD groups could not manipulate it such as stopping or rewinding. Lastly, the SD group studied the written material by themselves and subsequently discussed for 18 minutes in groups of three or four. Groups that had discussions were able to participate in discussions with the written materials.

To minimize the experimenter's interference during the discussion among the participants, the experimenter did not intervene while the students were learning (e.g., during review, self-study, and discussion). At the beginning of the discussion, the experimenter mentioned the end time of the discussion and informed that a test would be conducted after the discussion. The experimenter did not comment on the direction or the format of the discussion and just stated to freely exchange what they have learned. Participants were not allowed to ask questions to the experimenter. In other words, the experimenter did not give further instructions and did not engage in any intervention during the discussions.

After their learning activities, all three groups took a 20-minute test on the given material. The test questions were divided into three types as shown in **Appendix B**. (1) Verbatim items consisted of multiple choice or short answer questions. These questions asked straightforward information from the lecture or the written material. (*Ten possible points for ten questions*). An example of the item is: “Who does not have the right to file a complaint?”, followed by answer choices of 1) Legal representatives 2) Property manager 3) Guardian 4) A representative of the corporation which is the victim. Each item was worth 1 point.

(2) Paraphrased items. These were items that tested knowledge of concepts from the given material. Students had to explain concepts from the lecture or the written material. (*A total of twenty-two points for six questions*). Unlike verbatim items, paraphrased items required explanations or descriptions and partial credit could be given. An example of this item is “Explain who the entitled person with the right to file a complaint is.”

(3) Transfer items. These items required students to utilize their understanding of the whole content and apply it in new situations. (*A total of eighteen points for four questions*). Like paraphrased items, transfer items required explanations and partial credit was available. However, unlike paraphrased items which required reiteration of the presented material, transfer items sought to assess how much transfer of learning took place using legal case scenarios. An example of this item is as follows. “On May 1st, 2016 the father (F) of a famous under-aged celebrity (V), accused person (D) of infringement on privacy claiming that D had hacked V’s computer. However, F’s parental rights on V were terminated on January 1st, 2017. D was charged with infringement on privacy and was convicted guilty at the first trial. Eventually, D made an ap-

peal claiming that F has no right to file a complaint because his parental rights were terminated. Thus, D claims that the accusation should be dropped. Will the Court of Appeals accept D's claim? Students could receive partial points in cases of paraphrased and transfer items because several key concepts and expressions were involved.

In order to increase grading objectivity, 30 % of the total answer sheets were randomly selected and graded by two raters. We confirmed that there was no rater discrepancy for verbatim items. For paraphrased and transfer items, the raters' agreement measured by Intra-class correlation [ICC (3, k)] was .90 and .90 respectively for the following analyses. Thus, the remaining answer sheets were graded by the first author. We also set the criterion for outliers at above or below two-standard deviation from the mean, and there was none. Thus, significant experimental effect could not be due to outliers.

All materials such as the video lecture, the written instructional material, and the test questions were previously reviewed by legal experts and modified when necessary. In particular, experts examined whether the test items corresponded to the three types we classified, and whether the duration of the time compared to the time spent in learning and the score for each item were appropriate.

Results

The results of the background knowledge survey showed that most of the participants had almost no knowledge on the given content. The students had low self-reported knowledge of the two relevant topics.

The mean and standard deviation of the test scores are presented in **Table 1.1**. Analysis of variance (ANOVA) was conducted to examine the differences between the groups for the different types of test items according to our hy-

Table 1.1
Means (standard deviations) of planned comparison for Experiment 1

	LR ($n = 30$)	LD ($n = 29$)	SD ($n = 29$)	F	p	η_p^2	Summary
Total Score (50 points)	33.48 (*4.58*)	34.62 (*3.72*)	38.41 (*4.74*)	10.01	<.001	0.19	LR ≈ LD < SD
Verbatim (10 points)	6.40 (*1.61*)	6.35 (*1.29*)	7.35 (*1.20*)	4.82	.010	0.10	LR ≈ LD < SD
Paraphrased (22 points)	17.93 (*2.24*)	19.00 (*2.30*)	18.93 (*2.69*)	1.79	.173	0.04	LR ≈ LD ≈ SD
Transfer (18 points)	8.47 (*2.76*)	10.31 (*2.66*)	12.21 (*2.76*)	13.86	<.001	0.25	LR < LD < SD

Note. Mean scores for the total and the three different types of items as a function of different learning activities. LR = Lecture Review group; LD = Lecture Discussion group; SD = Self-study Discussion group.

potheses.

Significant differences were found among the groups in the means of the verbatim and transfer items, and total means. There was no significant difference in the means of the paraphrased items.

Subsequent planned comparisons revealed that the total mean of the SD group was significantly higher than that of the LD group, $t(85) = 3.26$, $p = .012$, $d = 0.88$, and that of the LR group, $t(85) = 4.29$, $p < .001$, $d = 1.05$. However, there was no difference between the LR and the LD groups, $t(85) = 1.00$, $p = .320$, $d = 0.26$.

For the verbatim items, the SD group scored significantly higher than the LD group, $t(85) = 2.76$, $p = .007$, $d = 0.64$, and the LR group, $t(85) = 2.63$, $p = .010$, $d = 0.46$. However, there was no significant difference between the LR and the LD groups.

For the paraphrased items, there was no significant difference between the three groups; SD, LD, and LR. For the transfer items, the SD group scored significantly higher than the LD group, $t(85) = 2.65$, $p = .010$, $d = 0.69$, and the LR group, $t(85) = 5.27$, $p < .001$, $d = 1.32$. In addition, the LD group scored significantly higher than the LR group, $t(85) = 2.60$, $p = .012$, $d = 0.65$.

Although there was no significant difference among the groups for the paraphrased items, possibly due to the ceiling effect (out of 22 points the means were over 18 points), the results support our predictions. In line with Hypothesis 1, the combination of self-study and discussion improved learning outcomes more than that of watching a lecture and discussion. In addition, in line with our second prediction (Hypothesis 2) that stated that discussions will promote greater learning than reviewing, the LD and SD group scored much higher than the LR group in the transfer items in the final test.

1.1.3 Experiment 2

Experiment 1 showed that combining self-study and discussions can lead to effective learning. However, a counterargument regarding the second hypothesis could be raised. One could argue that the reason why the SD group scored higher than the LD group was simply that self-study was more effective than watching a lecture. In other words, it is possible that the significant difference in scores between the SD and LD groups was due to differences in learning efficacy before holding a discussion. In order to rule out the possibility that the difference in learning outcome occurred simply because of the different preliminary activities themselves, the final test was administered immediately after the learning activity, i.e., self-study (S) and watching a lecture (L). The lack of significant difference between the S and the L groups would support our hypothesis that there is an interaction between the preliminary activity and the subsequent discussion.

Method

Participants, materials, and procedure

Fifty-two undergraduate students who did not participate in Experiment 1 participated in the experiment. Of the fifty-two students who signed up, two did not finish the experiment, leaving 38 men and 12 women. There were no significant differences in age among the groups. L group (M_{age} = 20.40, SD_{age} = 1.63), S group (M_{age} = 20.84, SD_{age} = 1.97). Participants were randomly assigned to one of the two groups: L (Lecture, n = 25), and S (Self-study, n = 25). Procedures and instructional materials were the same as Experiment 1, except that the discussion was replaced by a survey. The survey was about image evaluation, irrelevant to the learning content. Specifically, the L group watched the video lecture for 18 minutes with the written instructional material and then completed the survey for 18 minutes. Participants of the S group studied the learning material by themselves for 18 minutes and subsequently completed the survey for 18 minutes.

Results

Similar to Experiment 1, participants showed little relevant knowledge and there was no significant difference in the mean between the two groups. Among the participants, there were no outliers.

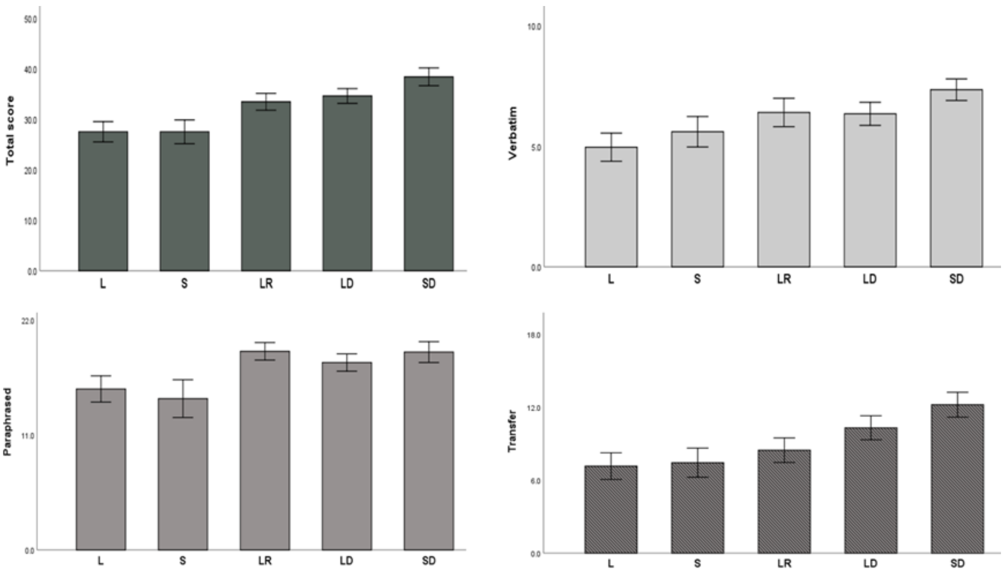
The test scores from Experiment 2 are provided in **Figure 1.1**, along with the results from Experiment 1. As expected, there was no difference between the two groups in all types of items.

The total means, verbatim items, paraphrased items, and transfer items between Experiments 1 and 2 (i.e., five groups) were significantly different, $F(4, 133) = 25.82$, $p < .001$, partial-eta squared = 0.44; $F(4, 133) = 10.61$, $p < .001$, partial-eta squared = 0.24; $F(4, 133) = 12.64$, $p < .001$, partial-eta squared = 0.28; $F(4, 133) = 16.03$, $p < .001$, partial-eta squared = 0.33, respectively.

We carried out a post hoc analysis using the Tukey test to find out any difference between two experiments. The total means of the L group ($M = 27.52$, $SD = 5.03$) and the S group ($M = 27.52$, $SD = 5.89$) were significantly lower than those of the three groups of Experiment 1, $ps < .001$. The L group ($M = 4.96$, $SD = 1.46$) scored significantly lower than other three groups in verbatim items, $ps < .01$, and the S group ($M = 5.60$, $SD = 1.58$) scored only lower than the SD group, $p < .001$. The L group ($M = 15.40$, $SD = 3.14$) and the S group ($M = 14.48$, $SD = 4.52$) scored significantly lower than the other three groups in the paraphrased items, $ps < .05$. For transfer items, the L

Figure 1.1

Comparisons of the final scores of Experiments 1 and 2 as a function of different learning activities



Note. L = Lecture group; S = Self-study group; LR = Lecture Review group; LD = Lecture Discussion group; SD = Self-study Discussion group. All types of items consist of verbatim, paraphrased, and transfer items. Error bars are ± 2SE. The top of the scale reflects the possible highest score on that scale.

group ($M = 7.16$, $SD = 2.73$) and the S group ($M = 7.44$, $SD = 3.00$) scored significantly lower than the two discussion groups, $ps < .01$, but showed no significant difference from the LR group. These results suggest that discussions enhance scores for paraphrased and transfer items.

The results of Experiment 2 suggest that the different outcomes from the LD and the SD groups in Experiment 1 were not because of the direct effect of the difference in preliminary activity, but because of the interaction between the preliminary activity and the discussion.

1.1.4 Experiment 3

The goals of Experiment 3 were twofold. One was to replicate the results of Experiment 1 to find out the difference between the SD and the LD groups. The other was to analyze the content of the discussions of the two groups. We hypothesized that the discussions of the SD group would be more active and productive than those of the LD group (Hypothesis 3). We then explored the relationship between specific factors in the discussion content and students' learning.

Method

Participants, materials, and procedure

Another seventy-four students who did not participate in Experiment 1 or 1.2 participated in this experiment (32 men and 42 women). There were no significant differences in age among the groups: LD group ($M_{age} = 20.30$, $SD_{age} = 1.47$, $G = 10$), and SD group ($M_{age} = 20.54$, $SD_{age} = 2.30$, $G = 10$). There were no outliers among the participants. Procedures were the same as in Experiment 1, except for the following two changes. One was that the only SD and LD groups were included in the experiment. The other change was that all the discussions were recorded. As in Experiment 1, at the beginning of the discussion, the experimenter notified the end time of the discussion and informed that a test would be conducted after the discussion. Participants were not allowed to ask questions to the experimenter.

Coding of discussion in small groups

As mentioned above, the present study followed Chi et al. (2017) in analyzing the content of the discussions. They categorized meaningful content of the discussions into three major types: substantive comments, interaction episodes and co-constructive turns. Substantive comments are defined as meaningful contributions in ongoing discussions, such as a relevant response to the concept taught or learned on one's own to other group members. Interaction episodes refer to coherent conversations among members that include at least one substantive comment. A co-constructive turn is defined as a conversation that includes "a change in a speaker that contains substantive contributions from at least two speakers". We predicted that the discussion within the SD group would contain more meaningful content than the discussion within the LD group. A detailed explanation is given below in the method section.

In the detailed coding scheme, each transcript was segmented according to speaker turns. A turn, by definition, is speech by a single speaker (Traum & Heeman, 1996). When successive conversational turns among group members are about the same topic, they form an episode. An episode is a multi-turn conversation on the same topic and line of thought that includes at least one substantive comment. The example of topics of the lecture and instructional material were shown in **Appendix A**. When a group discussed topic A and then topic B, then topic A again, they were coded as three different episodes. A turn within an episode is judged to be a substantive comment when it refers to questions and conversations related to concepts of the lecture or the written instructional material. Only the relevance of the comments to concepts was considered, regardless of the accuracy of the comments. For example, the following comments and questions were judged to be substantive: "What exactly is an offense subject to complaint? Does it mean punishment is possible if there are charges?" (Related to the concept of offense subject to complaint), "Then if the relative is the suspect, how do you file a complaint? I actually don't understand this part either, but if the relative is the suspect, what does it mean to independently charge the relatives? What does it mean by the independent victim?" (Related to the legal representative concept of victim). In contrast, comments irrelevant to the topic such as "Yes", "Umm" or "Maybe" were judged as non-substantive comments.

When two or more of the group members consecutively make substantive comments within an episode, it is called an interaction episode. Thus, in an interaction episode, at least two different members of the group needs to make at least one substantive comment as shown in the interaction episode 1 of **Appendix C**.

A co-constructive turn is a special case of an interaction episode, in that there is "a change in the speaker that contains substantive contributions from both speakers" (Chi et al., 2017). An example can be seen in the interaction episode 2 of **Appendix C**. Student 4 asked a question regarding the offense subject to complaint, which is a substantive comment. Student 5 then made a substantive comment answering the question. Then Student 6 made another substantive comment that explained the importance of the relationship in the offenses subject to complaints. These comments form a sequential turn of substantive comments, i.e., a co-constructive turn.

To summarize, productive discussions are characterized by more substantive

Table 1.2

Means (standard deviations) of planned comparison for Experiment 3

	LD ($n = 37$)	SD ($n = 37$)	t	p	d	Summary
Total Score (50 points)	31.57 (*6.38*)	36.19 (*4.64*)	3.57	.001	0.83	LD $<$ SD
Verbatim (10 points)	6.11 (*1.56*)	7.27 (*1.15*)	3.65	$<$.001	0.85	LD $<$ SD
Paraphrased (22 points)	17.43 (*2.81*)	17.57 (*3.31*)	0.19	.851	0.05	LD $\approx$ SD
Transfer (18 points)	8.05 (*1.84*)	11.19 (*2.49*)	4.16	$<$.001	0.97	LD $<$ SD

Note. Mean scores for the total and the three different types of items as a function of different learning activities. LD = Lecture Discussion group; SD = Self-study Discussion group.

comments and interaction episodes in conversations among students, and more substantive comments and co-constructive turns in interaction episodes.

Results

Similar to Experiment 1 and 2, most of the participants had almost no knowledge on the given content, and there was no significant difference between the two groups. The average and standard deviation of the results of final test from Experiment 3 are provided in **Table 1.2**. The results showed that the SD group performed better than the LD group in total score, verbatim, and transfer items. Thus, the results of Experiment 1 were successfully replicated for different participants.

Analysis of discussions

As described in the methods section, we adhered to Chi and her colleagues (Chi et al., 2008, 2017) in analyzing the content of the discussions. Twenty percent of the discussion videos (two from the SD group and two from the LD group) were randomly selected and transcribed and coded by two raters: the first author and a graduate student. The raters were blind to conditions. They agreed on 306 phrases uttered by the students within the discussion groups (96% of the total). The rater agreement measured by Kappa was over .90 for the following analyses. Thus, the remaining 80% of the videos were transcribed by the first author.

Results from the coded data are presented in **Table 1.3**. The SD group had significantly higher frequency of substantive comments compared to the LD group (37.80 vs. 15.20, $p = .020$, $d = 1.18$). As for the interaction episode, there was no significant difference between the two groups (3.80 vs. 5.50. $p = .205$, $d = 0.59$). However, the number of substantive comments per interaction

Table 1.3

Mean frequencies (standard deviations) of members' engagement behaviors per discussion group for the LD and SD groups

	Discussion				
Behavior	*After Lecture*		*After Self-study*	*p*	*d*
Conversations					
Frequency of substantive comments per discussion group	15.20 (*10.34*)	$<$	37.80 (*25.13*)	.022	1.18
Frequency of interaction episodes per discussion group	3.80 (*3.01*)	$\approx$	5.50 (*2.76*)	.205	0.59
Substantive comments per interaction episode	2.43 (*1.78*)	$<$	6.85 (*3.47*)	.003	1.60
Co-constructive turns per interaction episode	1.79 (*2.30*)	$<$	4.45 (*2.53*)	.025	1.10

episode was significantly greater in the SD group than in the LD group (6.85 vs. 2.43, $p = .003$, $d = 1.60$). Finally, the number of co-constructive turns per interaction episode from the SD group significantly exceeded that from the LD group (4.45 vs. 1.79, $p = .025$, $d = 1.10$).

Summarizing the results above, the groups that studied by themselves before discussion made more substantive comments, and also received higher scores on transfer items. These results suggest that substantive comments may have a direct impact on the performance of transfer items. To examine whether substantive comments had direct impact on the outcome of transfer items, an exploratory regression analysis was performed. The in-group members' average score of transfer items was set as the dependent variable and the number of substantive comments as explanatory variable. The results showed that the effect of substantive comments on transfer items was significant, $F(1, 18) = 5.38$, $p = .032$, $R^2 = .23$. The results are presented in **Table 1.4**.

The findings support our Hypothesis 3 that postulated that the discussions of the SD group would be more active and productive than those of the LD group. This was evidenced by the SD group generating more substantive comments and engaging in a greater number of co-constructive turns than the LD group. Moreover, the amount of substantive comments affected the final test scores positively.

Table 1.4
The result of explanatory regression analysis for Experiment 3

Effect	Estimate	*SE*	95% CI		*t*	*p*
			LL	UL		
Intercept	8.06	0.83	6.32	9.81	9.70	<.001
Substantive comment	0.06	0.02	0.01	0.11	2.32	.032

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

1.1.5 Discussion

In this study, we examined a two-part instructional sequence of self-study and discussions (SD) in higher education. In order to compare the efficacy of this combination with the most widely used format in typical college classes, we included the Lecture-Review (LR) combination. We also compared the SD combination with the Lecture-Discussion (LD) combination to determine which preparatory activity enhances the learning effect of discussions. The research hypotheses were: i) the learning outcome would be SD > LD > LR (Hypothesis 1 and Hypothesis 2), ii) the discussions of the SD group would be more active and productive than those of the LD group (Hypothesis 3).

To test our hypotheses, we conducted three experiments. In Experiment 1, we compared the test scores of three conditions: LR, LD, and SD. Apart from the fact that there was no significant difference among the groups for paraphrased items, the results supported our predictions. In line with Hypothesis 1, self-study and discussion improved learning outcomes more than watching a lecture and discussion. In addition, in line with Hypothesis 2, which states that discussions will promote greater learning than reviewing, the LD group scored much higher than the LR group in the transfer items.

In Experiment 2, we examined the possibility whether preliminary activities alone regardless of subsequent discussions, could lead to different outcomes. A lecture group and a self-study group took the final test without holding a discussion, and there was no difference in the test scores between the two groups. Thus, we verified that the different outcomes from the LD and the SD groups in Experiment 1 resulted from the interaction between the preliminary activity and the discussion. The result was especially meaningful in that the scores for transfer items showed a significant difference.

In Experiment 3, we examined the content of discussions between SD and the LD groups. Consistent with Hypothesis 3, the results showed that the SD group made more constructive and productive comments than the LD group. Although the mean number of interaction episodes was not different between the two discussion groups, this may be due to the small number of interaction episodes. Therefore, to examine the difference in a follow-up study, it seems necessary to increase the number of key concepts in the learning material or to perform a finer-grained analysis of the interaction quality.

Considering the results of previous studies by Chi et al. (2009, 2014), it is not surprising that the discussion groups outperformed the review group and that discussions are a very effective learning activity. Nevertheless, we were able to demonstrate for the first time that when self-study is combined with discussion, it leads to more productive discussion than when a lecture is combined with discussion.

So, why would self-study lead to a more productive discussion than watching a lecture? One speculation is as follows: compared to listening to a lecture, self-study can provide students with relatively more time for reflection. During self-study, students can concentrate on parts they want to focus on while skipping through others. In a lecture, on the other hand, information is delivered at a constant rate. As a result, the students in the self-study group are more likely to mark parts they did not understand. The marked parts may not be helpful for immediate learning, but they can be drawn out during discussions. Thus, the subsequent discussion would be more productive and lead to better learning outcomes.

Since the present research is at a preliminary stage and our explanation is provided as post hoc and hypothetical, its validity needs to be confirmed through follow-up studies. One limitation of this study is that the participants were students enrolled at a highly selective university in South Korea. Therefore, the study needs to be expanded to cover a variety of populations. In particular, it would be desirable to explore whether this method is effective even for elementary, middle and high school students.

Another limitation is that only one subject (i.e., law) matter was used. Although using obscure topics helps to ensure the internal validity of our experiments, it lowers the external validity in typical educational contexts in which students have diverse prior background knowledge. Thus, it would be worthwhile to expand experiment to different subjects, including science, technology, engineering, and mathematics (STEM) courses.

In addition, it is worth pointing out that we utilized student-led discussions. We know that teacher-led discussions improve students' understanding (e.g., Almasi & Garas-York, 2009; Sweller et al., 2007). The benefits come from the direct scaffolding during the discussion and the summative feedback afterward (Murphy et al., 2016; Sweller et al., 2007). However, many researchers argue that when teachers let go of their control and authority and act as facilitators, students' questions, co-construction, and learning increase (Schmidt et al., 2007; Soter et al., 2008; Wei & Murphy, 2017). Thus, future research should investigate the difference between the two discussion formats in their processes and effectiveness.

Nevertheless, the results of this study can contribute to learning sciences by drawing attention to the topic of self-study and the combination of learning activities. Self-study is essential when we acquire knowledge, solidify what we have learned from others, or reflectively examine what we know. Thus, it is important to understand self-study to better enhance students' learning. Despite its importance, however, not many studies on self-study have been conducted so far.

There is also a need for more research that examines combinations of two or more learning activities. In addition to the combination of self-study and discussions covered in this study, we can explore the learning effects of combinations such as self-study, discussion, and writing; self-study, writing, and discussion; problem-solving and instruction; or problem- solving and self-study.

Furthermore, a practical contribution of this study is that it suggests specific class activities that can replace traditional lectures. Despite the fact that lectures are not conducive to student learning compared to students' (inter) active learning, they are still the most widely used form of instruction in the classroom (e.g., Deslauriers et al., 2019; Freeman, et al., 2014; Stains et al., 2018). In this context, the combination of self-study and discussion is a promising activity that can replace traditional lectures. Thus, the present study can help educational stakeholders in higher education who may be looking for effective replacements for traditional lectures. It would be desirable to provide students with more opportunities to study on their own and to track the process and results by individuals or groups (e.g., Lam & Kapur, 2017; Mende et al., 2021; Schmidt et al., 2007). The tracking process involves asking the students to voice the difficulties or questions that they faced during self-study. The instructor can use this information to help students not only to digest the material, but also further their investigation of the topic. In addition, self-study can go beyond the scope of the present study to include more self-directed and open efforts, such as initial exploration of a topic, idea generation, and self-directed search of information to address unknowns.

Therefore, after more research and evidence that support their effectiveness, a significant portion of class time can be devoted to student-led activities such as self-study and discussion, leading to better learning outcomes. We hope that future follow-up studies on student-led learning activities and strategies will shed new light on the learning sciences and lead to more effective education.

Chapter 2

2.1 Self-study and Discussion Promote Student Learning in Science Education

2.1.1 Introduction

The impact of science and technology on our lives is greater than ever, and thus efforts are being made to teach them better (e.g., Taber & Akpan, 2017). Part of this includes an exploration of how science education differs from education in other fields. Compared to humanities or social sciences, natural science requires conceptual understanding (Konicek-Moran, & Keeley, 2015), proportional and hypothetico-deductive reasoning (Moore & Rubbo, 2012), and extensive mathematical knowledge. Thus, some researchers argue that science teaching methods differ from social sciences. For example, Hutner and Sampson (2015), referring to both A Framework for K-12 Science Education (NRC, 2012) and The Next Generation Science Standards (NRC, 2013), emphasize the importance of focusing on students' activities in the science classroom. They urge science teachers to create a need for learning, to make students' thinking visible, to engage students in an activity before delving into the learning content, and to involve students both in science practice and in negotiating meaning. However, these recommendations, except for engaging in actual scientific activity, are relevant not only to science education but also to social science education, to varying degrees. For now, suffice it to say that there are different views on science education.

Then are the learning processes different for the social sciences and the natural sciences? Even within the natural sciences, are the learning processes different between biology and physics? The present study was conducted to find answers to these questions. Our strategy is to explore whether the learning method discovered by Lim and Park (2023) is effective in science education.

A recent study by Lim and Park (2023) examined two-part instructional sequences composed of individual preparations and discussions. As for individual preparation activities, they contrasted traditional lectures and self-study. Self-study, in their research, was referred to individual or solitary learning to understand or master the learning material, such as reading and re-reading, rehearsing, self-explaining, taking notes, or drawing a concept map, etc. They chose an unfamiliar topic to learn for undergraduate students, specifically criminal procedure code. They found that combining self-study and discussion improved learning outcomes more than combining watching a lecture and discussion. They also found that the learning outcome of these two combinations was better than the combination of watching a lecture and reviewing. The results are important in that the combination of self-study and discussion can be an alternative to traditional lectures.

However, their study is preliminary and their finding needs to be replicated and extended to other topics in biology, chemistry, physics, etc. In acquisition of scientific knowledge, neither self-study nor discussion is widely used. Thus, whether the combination of self-study and discussion is advantageous in learning is an empirical issue to be examined.

Preparation for future discussion

Before examining the effect of self-study and discussion combination in the science domain, it is necessary to revisit the effectiveness of individual preparation before discussion on student learning first. The notion of preparation for subsequent learning is not new. There was an active search for ways to improve student learning by combining preparations and then other activities such as debate or discussions (e.g., Lam & Kapur, 2018; Lam & Muldner, 2017; Lim et al., 2019, 2022; Smith et al., 2009; Tsovaltzi et al., 2015). This concept generally stems from preparation for future learning (PFL, Schwartz & Bransford, 1998; Schwartz & Martin, 2004). PFL refers to a paradigm in which the sequence of instructional activities includes two main stages. The "preparation" stage, which consists of providing the information directly (Glogger-Frey et al., 2015), or through activating prior knowledge (Bielaczyc & Kapur, 2010; Bielaczyc et al., 2013) and the "learning" stage where the student receives other explicit instruction afterwards. Research on PFL has shown that students who receive explicit instruction after engaging in these preparation activities learn better than students who engage in the traditional sequence of acquiring formal knowledge from lectures and later applying it to a task (e.g., Schwartz et al., 2011; Schwartz & Martin, 2004; Kapur, 2008, 2014; Van Amelsvoort et al., 2007; Van Boxtel et al., 2000).

Since our interest is to compare the impact of individual preparations on student learning from subsequent discussion, our study is loosely based on the PFL paradigm. In our case, we designed the subsequent learning phase in the form of peer discussion, rather than explicit and direct instruction. Thus, we widened the lens of the PFL paradigm to discussions and named our framework, *Preparation for future discussion.*

There is evidence that individual preparation, along with information, can influence subsequent discussions. For example, Van Boxtel and colleagues set up two conditions studying concepts in physics where students were either involved in more generative preparation (i.e., creating a concept map) or less generative preparation (i.e., descriptive poster) before collaboration. The results showed that preparation overall supported proposals, confirmations, and criticisms during discussions. However, there were no significant differences in conceptual knowledge outcomes across the two conditions.

In addition, preparation influences the process of argumentation positively (Van Amelsvoort et al., 2007). Van Amelsvoort and colleagues compared diagrams provided by an experimenter before collaborative argumentation with diagrams created by students themselves. The result showed that students who generated their own diagrams demonstrated higher levels of argumentation and deeper learning in subsequent discussions.

As described earlier, Lim and Park (2023) provided empirical evidence on the benefits of preparatory activities before discussions; therefore, the current study extended their findings and focused on the advantage of combining self-study and discussion compared to other common and possible learning combinations in science education.

Overview of the present research

In this study, we adapted the pipelines of the experiments by Lim and Park (2023). We first examined the combination effect of different individual preparation activities and discussion on learning: lecture and discussion versus. self-study and discussion. We also checked the effect of discussion itself on learning since reviewing is the common activity after class or before exams. Furthermore, the present study added an exploratory group that proceeded the learning phase with additional instructional materials since students who are provided with more contents are predicted to be more successful (cf. Glogger et al., 2012; Winne & Perry, 2000).

In Experiment 1, we addressed the question of whether the natural science fields are truly different from the social science domains. Students with a variety of majors (not limited to science) were assigned to five conditions: discussion after watching a video lecture (LD), discussion after self-study (SD), reviewing after watching a video lecture (LR), reviewing after self-study (SR), and reviewing after self-study with additional instructional materials (SRA). Students were then introduced to introductory biology and test scores were compared.

Since combinations of different learning activities can affect transfer as well as memory tests (as shown by Chi & Wylie, 2014; Schwartz & Bransford, 1998), the test consisted of items that measure shallow learning (corresponding to verbatim), and deep learning (near-transfer). Unlike in the social sciences domain, we predicted that the combination of self-study and discussion will produce learning outcomes significantly higher than those of watching a lecture and discussion, which is superior to reviewing. Thus, we predicted that the order of learning performance would be LR, SR, SRA, LD, and SD, in ascending order.

2.1.2 Experiment 1

Method

Participants

All experiments were reviewed and approved by the university's Institutional Review Board (IRB, No. 2112/003-001). Undergraduate students at a university in Seoul participated in this experiment for course credits (ages 19 – 24 years). A total of 137 Korean students signed up for this experiment (including 54 women). Students were from diverse backgrounds (e.g., psychology, education, chemistry, and mechanical engineering) and there was no significant difference in the number of students between majors. There were no outliers among the participants. Students were randomly assigned to the five conditions: LR ($n = 27$), SR ($n = 30$), SRA ($n = 18$), LD ($n = 30$), and SD ($n = 32$).

Instructional video

The instructional video used in the experiment's lecture conditions were of monologue-style. The instructional video was not edited, and so the students saw the lecture exactly as it was originally filmed. We used a video lecture from a course in introductory biology, specifically on covalent bonds, non-covalent bonds, the octet rule, etc. The length of the video was 20 minutes and was taught by a male instructor. To keep the sessions as natural as possible, the instructor covered information in the written materials teaching when and how to use the key high-level concepts. In the two lecture conditions, students individually watched the video lecture on their personal screens. While watching the lecture, students were not allowed to stop or rewind the video, and manipulate it any other way.

Written materials

The written instructional material was six-page long and contained information covered in the lecture. However, it was not a transcription of the video. Two domain experts reviewed the instructional video and written material to ensure that both addressed the same key concepts (see **Appendix D**). All the students were provided with the same written material at the start of the experiment, except that the SRA group received an eight-page long material. The number of key concepts covered in the material for the SRA group were the same, but additional examples were added.

Test items

Test items are shown in **Appendix D**. The test questions were made up of items testing deep knowledge. We defined deep knowledge as the ability to analyze, evaluate, and create. Students are required to possess factual definitions and descriptions of the concepts while also demonstrating the ability to apply or transfer them. Thus, the final test was consisted of ten questions worth a total of 33 points, and it required students to transfer knowledge (i.e. accompanying explanations of novel situations that were not present in the instructional materials).

Although there was an answer sheet, since the questions required explaining the concepts learned, there was room to get the answer right even if it was not written in the exact words. Thus, to exclude the subjectivity of scoring we calculated intra-class correlations (ICC) between the three raters. Forty-two answer sheets were randomly selected and scored by the first author and two raters using an exemplary answer sheet (30% of the final test answers). ICC (3, k) was high by .90 ($p < .001$) and thus, the remaining test responses were graded by the first author.

Design and procedure

Students were briefed by the researcher, who explained experimental participation would be anonymous and would not affect course grades. After students signed the consent form, they were randomly assigned to five conditions and participated in separate locations accordingly. (1) The LR group individually watched the 20-minute video lecture with the provided written instructional material. After a five-minute break, students studied on their own using the written material for 15 minutes. (2) The SR group self-studied the written material for 20 minutes. After a five-minute break, they studied the same material for 15 minutes. (3) The SRA group had the same learning process as the SR group but received the longer learning material with additional explanations and examples. (4) The LD group had a 15-minute discussion in groups of three or four after individually watching a video lecture along with the written material for 20 minutes, then took a five-minute break. (5) The SD group, after self-studying the written material by themselves for 20 minutes, took a five-minute break, and then discussed for 15 minutes in groups of three or four ($\sim$ 40 min).

The researcher did not intervene while the students were learning (e.g., during review, self-study, and discussion). In other words, the experimenter did not give further instructions or engaged in any intervention during the experiment process. After the learning phase, the researcher administered the test with a time limit of 20 minutes. Students who were finished before 20 minutes passed were asked to go over their answers until the time was over ($\sim$ 60 min).

Result and discussion

Learning outcome

The mean and standard deviation of the test scores are presented in **Table 2.1**. A series of analysis of variance (ANOVAs) was conducted to examine the differences between the test scores of the groups according to our assumptions.

Significant differences were found among the groups in total means, $F(4, 132) = 23.73$, $p < .001$, partial eta squared = 0.42. Subsequent comparisons revealed that the total mean of the SD group was significantly higher than that of the other four groups, $ps < .05$. The LD group received significant higher scores than those of the LR group, $p < .001$, those of the SR group, $p = .001$, and those of the SRA group, $p < .001$. However, there was no difference among the LR, the SR, and the SRA groups, $p = 1.000$.

Consistent with the previous findings (Lim & Park, 2023), in the biology

Table 2.1
Means (standard deviations) for each condition in Experiment 1

	LR ($n = 27$)	SR ($n = 30$)	SRA ($n = 18$)	LD ($n = 30$)	SD ($n = 32$)
Total score (33 points)	17.15 (*3.96*)	17.67 (*3.40*)	16.72 (*4.11*)	21.23 (*2.92*)	24.03 (*3.09*)

Note. LR = Lecture Review group; SR = Self-study Review group; SRA = Self-study Review with additional instructional material group; LD = Lecture Discussion group; and SD = Self-study Discussion group.

domain, the two discussion groups scored significantly higher than the other review groups. Interestingly, the SRA group, which was given more material, also scored lower than the discussion group. Thus, we can infer that the higher scores do not come from the lack of learning material. Among the discussion groups, the SD group score significantly higher than the LD group. In the end, our results show the importance of engaging in a discussion and the effectiveness of self-study and discussion combined.

Nevertheless, counterarguments may be raised. First, the participants of Experiment 1 only learned concepts from the biology field while their background or majors were unrelated. Second, introductory biology may not require a lot of mathematical knowledge. Therefore, we extended our experiment to science and engineering students using topics in physics.

2.1.3 Experiment 2

Experiment 2 addressed the question of whether there really is a difference between topics that require mathematical knowledge and others. Thus, Experiment 2 had two differences from Experiment 1, besides the fact that we used a material in physics instead of biology. One difference was that only discussion groups, the SD and the LD groups were compared. The other was that we performed a pretest to check whether there are significant differences in prior knowledge among the participants, and the posttest to analyze differences in learning gains. We also analyzed the content of the discussions to explain why the self-study group achieve higher scores than the lecture group.

Method

Participants

Second- and third-year undergraduate students who did not participate in Experiment 1 were recruited ($N = 73$; 51 men and 22 women). Students were all majoring in science and engineering (e.g., physics, life science and chemical engineering) and there was no significant difference in the number of students by major. All participants were Korean. There were no outliers among the participants. Students were randomly divided into each group: the LD group ($n = 37$) and the SD group ($n = 36$).

Materials

The study used the following materials related to the Photoelectric effect and the Compton effect: (1) an eight-page instructional written text, (2) a 30-min instructional video lecture, and (3) a pretest and posttest.

Instructional video and written materials. The 30-minute video lecture was of monologue-style. We used one video lecture on physics, specifically dealing with Photoelectric effect and Compton effect. A male instructor covered information in the written instructional materials explaining how to apply and use the key concepts. For example, the instructor would solve a sample problem and apply the formula learned to explain it. While watching the video, students could not manipulate it such as stopping or rewinding.

The written instructional material is presented in **Appendix E**. The eight-page written text was designed to provide a general overview of the photoelectric effect and the Compton effect to help prepare students for the discussion afterwards. The material was not a transcription of the video lecture. All the students were provided with the written material at the start of the experiment.

Pre-test and posttests. We first created a test with 20 questions to divide it into two isomorphic tests of similar difficulty. The internal consistency reliability of the test was estimated by calculating the correlation of scores between the odd and even number items and then applying the Spearman–Brown prophecy formula. The internal consistency reliability was .82.

The two isomorphic tests were used as the pretest and immediate posttests after learning. Both tests consisted of (1) four multiple-choice questions with five options assessing verbatim and shallow knowledge, (2) three multiple-choice questions also with five options and three descriptive type questions assessing transfer of learning and deep knowledge. The multiple-choice questions included an "I have no idea" option. For instance, "Select the statement that is not correct about Photo-emission". The transfer questions included items that required comprehension of the Photoelectric effect and the Compton effect but without explicitly mentioning these concepts. For example, a question on the Compton effect asked "X-rays of 10 [pm] wavelength is scattered at 45 degrees from the target. Find the maximum kinetic energy of the reflected electron" (see **Appendix E** for more examples).

All tests were scored using the same method, blind to condition. ICC was used to check reliability among raters, where three raters scored twenty percent of the data, randomly selected from each of the two tests (pretest and posttest).

ICC (3, k) was high both for the pretest was .94, $p < .001$, and for the posttest was .91, $p < .001$, thus one rater scored the remaining answer responses.

Design and procedure

Procedures were the same as in Experiment 1, except for the following two changes. One was that only the SD and the LD groups were included. The other change was that all the discussions in each condition were audio-recorded. Students were randomly assigned to two groups. The two groups studied in different locations. (1) In the LD group, after watching a video lecture with written materials for 30 minutes, students took a break for 5 minutes, and then in a group of three to four engaged in a discussion for 15 minutes. (2) The SD group studied by themselves for 30 minutes and then took a 5-minute break, followed by a 15-minute discussion in a group of three to four students ($\sim$ 50 min). The discussion proceeded as a student-led discussion, and the researcher did not provide any direction. Finally, both groups took a 20-minute test ($\sim$ 70 min).

Result and discussion

Learning outcome

Shallow learning. We first checked prior shallow knowledge using the pretest and found no significant differences across the two conditions, $F(1, 71) = 1.49$, $p = .227$, partial eta-squared = 0.02. We used an analysis of covariance (ANCOVA) with the pretest score as a covariate to measure gains from pre to post learning. ANCOVA with the pretest score as the covariate and the posttest score as the dependent variable revealed a significant difference between the two groups, $F(1, 69) = 25.13$, $p < .001$, partial eta-squared = 0.27. Planned comparison showed that the posttest scores of the SD group on shallow knowledge items were significantly higher than those of the LD group, $t(71) = 5.05$, $p < .001$,

$d = 1.19$.

Deep learning. The pretest and posttest results are summarized in **Table 2.2**, based on that of 73 students who completed all learning activities and transfer posttest items. We used ANCOVA to analyze the differences in outcome across conditions. The level of deep learning shown in the pretest scores between the two groups was not significant, $F(1, 71) = 0.05$, $p = .821$, partial eta-squared $= 0.00$. However, as with previous results regarding the shallow items, there was a significant difference in deep learning at posttest, $F(1, 69) = 8.73$, $p < .05$, partial eta-squared $= 0.11$. We found that the posttest score of the SD group was significantly higher than that of the LD group, $t(71) = 3.55$, $p = .001$, $d = 0.83$.

Thus, results of Experiment 1 (also study by Lim & Park, 2023) were successfully replicated for different participants and in physics, which require mathematical knowledge. Students in both discussion groups had the same constructive opportunities, but their interactions were not the same, even though the instructor did not lead the interactions. What caused this difference between the two discussion conditions?

As Kapur (2014) has shown, even in mathematics, where conceptual knowledge and procedural knowledge are deemed important, students achieve better outcomes than those who experience direct instruction, particularly when they first experience the materials independently. Thus, students may first encounter productive failure through self-study, followed by sharing their questions and difficulties in the subsequent discussion.

In addition, we need to check and interpret this difference in scores between the two discussion conditions in order to examine how student preparation can affect learning in general, as well as interaction and substantive contributions during discussions in particular (e.g., Chi et al., 2017; Lim & Park, 2023; Muld-

Table 2.2

Means and standard deviations of each condition at pretest, posttest, and gain in Experiment 2

Variable	Discussion after lecture (LD)		Discussion after self-study (SD)	
	M	*SD*	M	*SD*
Shallow (8 points)				
Pre-test	3.43	1.95	2.91	1.65
Post-test (adjusted[a])	4.97 (4.96)	1.86	6.83 (6.85)	1.21
Gain[b]	1.54		3.92	
Deep (22 points)				
Pre-test	8.24	3.30	9.44	2.95
Post-test (adjusted[a])	14.08 (14.29)	3.34	16.92 (16.71)	3.48
Gain[b]	5.84		7.48	

[a]The values in parentheses report the posttest scores adjusted by the pretest covariate.

[b]Gain = unadjusted posttest – pretest.

ner et al., 2014). Thus, we analyzed students' substantive contributions during the discussion to explore how results from Experiment 1 and Experiment 2 can be interpreted. We then correlated substantive contributions with students' learning outcome.

Students' conversations within small groups

In this set of analyses, we explored whether there were differences in the way the students interacted with each other and with the instructional material. We analyzed their conversation data to capture students' engagement behavior. The present study referred to Chi et al. (2017) when analyzing the verbal content of the discussion. The audio-recorded discussions were first transcribed for the analyses, then segmented into statements and coded. To identify students' substantive contributions, we segmented their verbal utterances at a phrasal level according to Chi et al. (2008). We then adapted the following definition of a substantive comment: meaningful contribution to an ongoing activity such as problem solving. Given these definitions, segments related to topics learned by students after self-studying the learning material and watching lectures were considered substantive. However, simply reading or repeating statements in the material or video, and mentioning off-topic comments or meta-talk such as "umm" or "yeah" were not classified as substantive. We wanted to analyze ideas and thoughts that were related to the learning topic (the Photoelectric effect and the Compton effect) but those that were not explicitly from the instructional material. In this context, conversational starters related to the material were also considered substantive comments. The following consecutive student utterance (from script no.7) are examples of substantive comments:

[1] Then, what did you think about electromagnetic waves?

[2] Is this part related to the electromagnetic wave theory?
[3] (reading a text) The text shows that when the intensity of light decreases, it takes time to reach a large vibration, or amplitude.

Segment [1] is related to the material on hand and is a starting point of a new conversation topic, which is considered substantive. Segment [2] is also considered a substantive comment because it is a phrase that connects to and specifies what was said in segment [1]. However, segment [3] is not considered substantive as the student is simply reading what is written in the material.

Students' conversations were further analyzed to explore how interactive students were within the groups. The transcript was first segmented into episodes. An episode is a multiturn conversation on the same topic and concepts within the instructional material. Each episode was then considered an interaction episode if each student in a group provided at least one substantive comment. **Appendix F** includes an example of an interaction episode. Student 1 provided substantive comments in lines 1 and 4, while Student 2 and Student 3 also provided substantive comments in line 2 and 3 in response. Overall, Episode 3 on the Compton effect was identified as an interaction episode.

A turn is usually defined as a change in speakers (e.g., Traum & Heeman, 1996). Chi et al. (2017) defined a co-constructive turn as a change in speakers that contains substantive contributions from both speakers. Here, we defined a co-constructive turn as a change that contains substantive comments from two or more students in a group. For example, in **Appendix F**, Episode 7, on the topic of Photoelectric effect, Student 6 contributed a substantive comment at the start of Turn 1, and Student 4 contributed a substantive comment afterward. That is, these two turns in which two students provided substantive comments are identified as one co-constructive turn. However, immediately afterward, Student 5 did not provide a substantive comment. Thus, Episode 7 was considered to have one co-constructive turn in total. We predicted that the SD group would have more meaningful utterances than the LD group in their discussions.

Analyses of discussions

One of the researchers and two graduate students coded 30% of the randomly selected transcripts to analyze the dialogues. Interrater reliability was appropriate (ICC (3, k) = .92, $p < .001$), thus the remaining transcripts were coded by one of the researchers. **Table 2.3** shows the mean values and standard deviations for the number of substantive comments, interaction episodes and co-constructive turns in the two conditions. We also examined how many utterances the students gave in each group and the proportion of substantive comments within the utterances.

The SD group spoke more utterances than did the LD group (58.36 vs, 40.83). The SD group also generated more substantive comments than did the LD group (49.45 vs. 24.17). Interestingly, the SD group had a higher proportion of substantive contribution than the LD group (79.45% vs. 55.88%). There was no significant difference between the two groups in the number of interaction episodes (3.73 vs. 3.33). This may possibly be due to a lack of a total number of concepts in the instructional material. There was also a lack of overlap in the frequencies of interaction episodes for the students in the LD and SD groups. This result suggests that a further analysis of interactions is necessary to identify significant differences between the two discussion groups.

Accordingly, we analyzed whether an interaction episode contained more substantive comments. The result showed that when the members in the SD group discussed, there were more back-and-forth substantive contribution (also

Table 2.3
Means (standard deviations) of discussion conditions for students' conversations

	Discussion group				
Conversations	*Lecture*		*Self-study*	*p*	*d*
Total utterances during discussions	40.83 (*11.18*)	=	58.36 (*26.78*)	.050	0.85
Substantive comments per discussion group	25.17 (*11.66*)	<	49.45 (*27.15*)	.008	1.21
Substantive comments of total utterances (percentage)	61.65%	<	84.73%	.002	1.48
Interaction episodes per discussion group	3.33 (*1.78*)	=	3.73 (*2.20*)	.640	0.20
Substantive comments per interaction episode	7.97 (*3.39*)	<	14.60 (*7.97*)	.015	1.08
Co-constructive turns per interaction episode	11.17 (*5.95*)	<	18.91 (*7.05*)	.009	1.19

co-constructive turns) within each episode than those in the LD group. Thus, discussion from the SD group was richer in that an episode contained more content-relevant information than one from the LD group. The result showed that the frequency of substantive comments per interaction episode was significantly higher in the SD group than in the LD group (14.60 vs. 7.97). The frequency of co-constructive turns per interaction episode of the SD group were also significantly higher that of the LD group (11.17 vs. 18.91).

As predicted, the SD group was beyond active, also constructive and productive during the discussion compared to the LD group. Specifically, the SD group shared significantly more substantive comments, not only because they spoke more utterances but also more substantial contributions within each utterance. Although the difference in the mean number of interaction episodes were not significant between two groups, this may be due to a lack of concepts in the instructional material. Thus, we tried to measure the richness of the interaction during discussions in other granular ways of analysis. Accordingly, from the significant difference in the number of co-constructive turns, we can infer as more substantive comments are shared between the students within an interactive episode, the interaction becomes more constructive.

Therefore, given that the generation of substantive contributions has been shown to be associated with learning (Chi et al., 2008), these results provide a possible explanation for the differences in student learning outcome. That is, in Experiment 2, students in the SD group generated more substantive contributions than did students in the LD group, and this pattern corresponds to their difference in learning, where the SD group scored higher than the LD group. Similarly, in Experiment 1 there were also significant differences in learning outcomes between the LD and SD groups, which suggests these results could be derived from a reliable difference in students' substantive contributions during discussion between the two conditions. The next analysis attempts to go through this possibility step by step. We now present exploratory analyses on our sample of audio-recorded dialogues to shed light on the results.

From substantive comments to learning outcomes

The above results suggest that substantive comments may have a direct impact on student learning outcome (in particular transfer). To substantiate our interpretation that substantive comments had direct impact on the outcome of transfer items, we ran an exploratory regression analysis. We set in-group members' average scores for transfer items as the dependent variable, the number of substantive comments, individual preparation activity and substantive $\times$ individual preparation activity as the explanatory variables. Since individual preparation (self-studying or taking a lecture) is a categorical variable, we transformed it into two binary dummy variables (0, 1), where 0 corresponds to taking a lecture and 1 corresponds to self-studying. The results are presented in **Table 2.4**.

The overall model was significant, $F(3, 19) = 11.43$, $p < .001$ ($R^2 = .64$). In detail, the number of substantive comments and individual preparation were not significant, $p = .193$, $p = .103$, respectively. More importantly our main focus, which is the number of substantive comments $\times$ individual preparation, was significant, $p = .012$. These results imply how substantive contributions might have influenced students' learning outcome in the LD and SD groups. In line with the findings from previous research (Muldner et al., 2014), our results also show that the interaction term is modest yet significant, indicating substantive comments have a stronger positive relationship with student learning outcomes in the context of self-studying as compared to taking a lecture.

Table 2.4
Results of the linear regression analyses

Predictor	**Unstandardized coefficient**	**Standardized coefficient**	t	p
Experiment 2				
Substantive × Individual preparation	0.22	0.75	2.79	.012
Constant	25.73	-	14.38	<.001
Substantive comment	-0.06	-0.32	-1.77	.193
Individual preparation	-12.55	-1.50	-4.82	.103

2.1.4 General discussion

The present study proposed the *preparation for future discussion* paradigm by testing our predictions on the importance of preliminary activities before discussion, and focused on the self-study. This study extended the findings of Lim and Park (2023) by focusing on the learning effects of self-study and discussion combined in the context of science education.

In Experiment 1, undergraduate students were introduced to topics in introductory biology. We compared students' posttest scores of five different groups. As a result, the SD groups had the highest scores, followed by the LD group and other review groups in descending order. There was no significant difference between the review groups, regardless of the amount of content in the learning material.

In Experiment 2 we recruited science and engineering students to study concepts in physics that required mathematical knowledge. Consistent with the results of Experiment 1, the self-study group scored higher than the lecture group. The combination of self-study and discussion was also superior to the other combinations in terms of learning gains. Since we predicted differences in the overt behaviors shown by students when discussing with their peers after self-studying and taking a lecture, we analyzed the dialogue of their discussions.

The results indicated that the SD group made more active and constructive comments than the LD group. Thus, we confirmed that combining self-study and discussion does not simply produce an additive effect, but rather suggests an interaction between the activities.

Therefore, we can now answer questions we posed earlier. We confirmed the self-study and discussion combination enhances student learning, regardless of learning domain (whether it is natural or social sciences) and student background. Even within natural sciences, self-study improved the learning effects of discussions in topics that require conceptual and mathematical knowledge. This seems to be because students clarified their questions and difficulties regarding the contents through self-directed or self-paced learning, and resolved them by actively interacting in the discussion afterwards. In addition, through self-study, students might have first explored the materials relating learning contents and then shared their questions during the subsequent discussion.

Limitations and implications

Further studies are needed to generalize our findings. One limitation is that our research was not conducted during courses in a regular semester. For example, it would be more impactful to compare how students' learning changes over several classes or weeks compared to their initial state. Another obvious limitation is that only college students participated. It would be helpful to generalize our finding and test this design for middle school or high school students.

The current study suggests that when students engage in self-study individually before participating in discussions, they tend to have more active and constructive discussions, which eventually leads to increased learning outcomes. We first want to emphasize that both the SD and the LD groups showed higher overall learning outcomes than the review groups. Considering the ICAP framework (Chi & Wylie, 2014; Lam & Muldner, 2017; Muldner et al., 2014), it may be easily inferred why the discussion groups outperformed the review group. Nonetheless, we insist again that educators and teachers should put more effort in securing time for discussion among students on what they are learning.

In order to expedite change in this direction, it is necessary to determine and help students in areas where they find difficult to study on their own. It would be desirable to provide students with more opportunities to study on their own and to track the process and results by individuals or groups. As shown by research on self-directed learning, students during self-study can have chances to select, combine, and coordinate cognitive strategies in their own way, and in a more effective way. Students during self-study devote much attention to direct and monitor their own learning and regulatory processes. Through the attainment of metacognitive knowledge in one's own learning, students can improve their ability to transfer knowledge and strategies from one context to new situations.

This kind of endeavor would take much time and effort, but the results of our study would provide valuable information in today's fast changing world where lifelong learners are required. In order to secure class time to support the growth of such learners, it is necessary to reduce lecture time. In this vein, it is worth examining attempts made by the Minerva School to limit the percentage of lectures to 25% of their class time (Kosslyn & Nelson, 2017). Instead of filling up class time with lectures as usual, we need to investigate topics or domains which require lectures as well as ways to reduce them.

More importantly, as Schwartz et al. (2007) demonstrated, preparation activity can reveal the hidden value of educational experiences. our findings have provided evidence that the combination of self-study and discussion is a useful design for improving learning outcomes in science classes compared to traditional worked designs. Ultimately, within a classroom that values students' dialogic interactions, we hope students to take responsibility for co-constructing their understandings together, a process Mercer (2000) termed "inter-thinking".

2.2 Self-study before Discussion Enhances Learning for High School Students

2.2.1 Introduction

The Advanced Placement (AP) program has been instrumental in bridging the gap between K-12 and higher education. Beginning in 1952, prep schools have collaborated with universities to introduce the radical concept of offering high school seniors the opportunity to study college-level material and take achievement exams (Kolluri, 2018). AP has played an important role in improving the quality of education in schools by setting academic achievement standards instead of solely administering tests. This precedent suggests that, despite the differences in structure and philosophy between K-12 and higher education, it is possible to prepare high school students with college-level knowledge and skills.

However, some scholars can claim that "students of different age groups have different modes of learning, and their learning could be improved by matching one's instruction with that appropriate learning mode", even though there is no credible evidence (e.g., Riener & Willingham, 2010). We rather seek to ask a question, "is there a need to distinguish instructional strategies by age groups?"

To answer this question, we extend the core features of the design studied by Lim and Park (2023) to high school students. In three experiments conducted in the higher education context, the performances of the review after lecture group, the discussion after lecture group, and the discussion after self-study group were compared. The results showed that the discussion after lecture group outperformed the review after lecture group. Particularly, the scores for the self-study and discussion group was significantly higher than the lecture and discussion group of memorization and transfer. Thus, the results indicate that self-study can significantly enhance the learning effect of discussions.

Thus, the present study compares two distinct instructional combinations in a context different from those previously compared (e.g., Lim & Park, 2023). Previous research on the combination of learning activities has focused on higher education and simple immediate learning measures. To address these shortcomings, we investigated the relative effectiveness of combining self-study and discussion compared with the combination of lecture and discussion in a 12th grade classroom. Additionally, we examined acquisition and retention over a longer time period of a week, in contrast to the study conducted by Lim and Park, which only measured immediate learning. Given the readiness of students for higher education, the adoption of a “long-term perspective” in our present study seems valuable for investigation.

Preparatory activity for discussion

The present study focuses on combining discussions with preparatory activities. In line with the idea that discussions require preparation to activate prior knowledge about the content as well as to facilitate interaction among students (e.g., Murphy et al., 2016), previous research have found that preparatory activities promote students learning through discussions, such as *think-pair-share*, where students individually think about the answer, then discuss their idea with a peer (e.g., McTighe & Lyman, 1988), preparation for future collaboration (Lam & Kapur, 2018; Lam & Muldner, 2017), and generative preparatory task before discussions (Mende et al., 2021; Van Amelsvoort et al., 2007).

Preparatory activities before the discussion are also efficient. Most approaches shown to improve learning through discussions include training teachers or students specific techniques such as scaffolding, guiding through scripts or prompts (e.g., Fischer et al., 2013), forming structured groups for argumentation (e.g., Asterhan & Schwarz, 2009; Kuhn & Crowell, 2011; Reznitskaya et al., 2012), and designing tasks to elicit substantive discussion (e.g., Engle & Conant, 2002; Kapur & Bielaczyc, 2012). However, these approaches that scaffold students interaction throughout the discussion, can limit the range and diversity of the interaction with too much structure and guidance, which in turn may lead to less learning (Cohen, 1994; Kapur, 2016). Moreover, these strategies may benefit students in the short-term, but students often stop using the skills after time (Webb et al., 2006). Therefore, it is more efficient to support student learning and reduce teacher burden by using a preparatory activity before the discussions.

In the present study, we compare lectures and self-study as preparatory activities. Since lectures are convenient and an efficient way to deliver accumulated knowledge, they are the most used instructional method across all disciplines (e.g., Knight & Wood, 2005; Schmidt et al., 2009; Stains et al., 2018). On the other hand, self-study is also a widely used individual activity but is relatively under-researched (e.g., Schmidt et al., 2010). In this study, the self-study is defined as intentional learning where learners are involved in their learning process, establishing their learning goals and understanding, monitoring learning strategies and implementing evaluation before the subsequent discussion. Self-study offers students an opportunity to distinguish what they know from what they still need to learn, and to refine their questions for discussion. In a practical setting, Lim and Park's experiment (2023, Experiment 2) showed that there was no significant difference in immediate learning achievement between lecture itself and self-study itself. However, self-study elicited more constructive and productive content during the discussion when compared to lectures.

Therefore, the findings broadly claim that the relative effectiveness of self-study and discussion need to be conditionalized and extended in particular to other educational contexts.

2.2.2 The present study

In the present study reported here, we replicated the findings of Lim and Park (2023). Unlike past experiments that only recruited undergraduate students, we took the classroom discussions to high school to extend and generalize their findings. The first condition required three or four students to discuss the instructional material after watching a lecture individually. In the second condition, students self-studied the instructional material individually and then discussed it in groups of three or four.

Although Korean high school students rarely experience discussions in classrooms, we predicted that students in the self-study condition would outperform those in the lecture condition because they would have a more dynamic and productive discussions. Thus, we proposed the intentional self-study hypothesis, which refers to cognitive processes aimed at learning. Compared to lectures, self-study promotes more active and constructive engagement with learning materials and peers during discussions, resulting in both short- and long-term learning outcomes.

Method

Participants

The university's Institutional Review Board (IRB, No. 2211/004-002) reviewed the present study. Seventy-three grade-12 students at a high school in South Korea participated in this experiment. There were no outliers among the participants. Students were randomly assigned to two conditions: discussion after lecture (n = 35, the number of groups = 10) and discussion after self-study (n = 38, the number of groups = 11). Because of the nature of classroom research, it can be argued that the number of participants tends to be a relatively small number, which can lead to statistical issues. However, our sample size was considered sufficient, based on the power analysis conducted with the G*Power system (Faul et al., 2008), and observed effect size from previous research (cf. Chi et al., 2008).

Learning materials

As far as the materials are concerned, we found those from the prior study (Lim & Park, 2023) to be appropriate and used them. The materials for the current study were based on materials from Experiment 3 of Lim and Park (2023) but with some minor refinements. For instance, the tests used were divided into 6 pretest questions (corresponding to paraphrased items in Lim & Park) and 14 posttest questions (corresponding to verbatim and transfer items). The instructional materials included a video lecture, a written instructional text, and test items.

Video lecture. The video lecture available on the internet focused on a law-related topic. Specifically, it provided an explanation of the concepts of criminal procedure code and accusation. Criminal procedure code and accusation can be challenging to comprehend, particularly for K-12 students who often have limited knowledge of the details within the Act or only have a superficial understanding of related terms like “charge.”

In the video, a teacher covered all of the concepts related to our learning contents, such as difference between accusation and complaint, accusation capacity, offense subject to complaint, etc. needed to answer the questions in a monologue style. The test questions were based on real-life scenarios and required retention of concepts learned in order to solve them. The teacher thus explained the concepts by citing actual cases.

Written instructional text. The written instructional text spanned seven-pages and contained the information discussed in the video lecture. It was not a transcript of the video, but rather an excerpt from a criminal procedure code textbook. The learning materials were created by one domain expert and then reviewed by another expert to ensure that all materials covered the same concepts (refer to **Appendix G** for more details).

Test and scoring. We made two types of tests. One was a screening test that measured how much students knew about the learning content prior to the learning phase. Students were asked to answer questions on similar concepts in the learning materials in an open-ended format. A maximum of 22 points was possible for a total of 6 questions.

The other test was a posttest (including immediate and long-term) that measured how well students learned the content during the learning phase. The posttest was made up of two levels. First, retention items consisted of five multiple-choice questions with four options and five short answer questions assessing conceptual knowledge on criminal procedure and accusation (*a total of 10 points*).

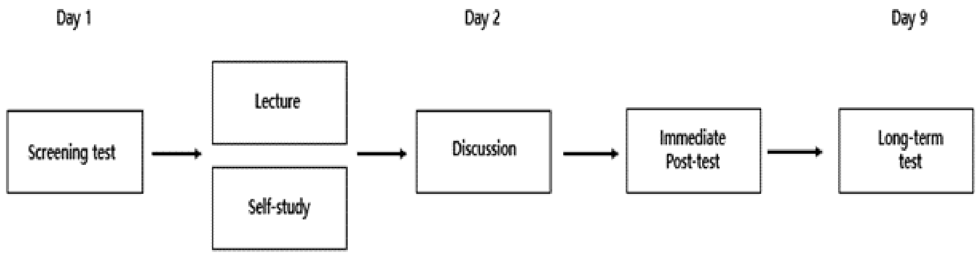
In addition, the near-transfer items consisted of four questions and were worth 18 points. The items were designed to describe the right verdict for each of the actual cases using the concepts students learned. The four scenarios for each of the item consisted of some description of the learning content and the situation of a legal representative. Some scenarios showed both a legal representative of an under-aged victim and the victim's complaint against their parents. For instance, students were instructed to solve questions such as "The under-aged victim (V) accused the offender (D) of contempt, and then withdrew his accusation on July 26th, 2017 ... Will the Court of Appeals accept D's claim? Explain your answer."

There were exemplar answers written by the experts to the questions. However, students can use their own language when answering near-transfer items within the posttest. Although students may not be able to use the exact terminology, points were awarded if the meaning was consistent. Eleven posttest answer sheets per condition were randomly selected and scored by one author and two other raters, using the exemplar answer (30% of the immediate posttest answer responses). The raters' agreement measured by ICC was .91, $p < .001$, thus rest of the answer sheets, were graded by the author.

Procedure

Figure 2.1 shows experimental procedures.

Figure 2.1
Description of the study design and procedure



Day 1: the researcher explained the experimental procedure to the students who consented the experiment. Students then completed the 15-min screening test.

Day 2: two researcher facilitated the learning phase. Students were informed that they would first spend some of the class time working alone and then work with two peers to hold a discussion. Students were randomly divided into lecture first or self-study first group and the two conditions worked in separate classrooms. At the beginning of the learning phase, students were given the same written instructional materials to study. In the lecture condition, students individually watched an 18-minute video lecture on their personal screens. However, they were not allowed to manipulate the lecture, such as stopping or rewinding it. For the self-study condition, students were instructed to study on their own for 18 minutes instead of watching a lecture. Following the individual learning phase, students formed groups of three and engaged in a 15-minute discussion. As the discussion involved only the students, the researcher did not provide additional guidance or make any announcements regarding how the discussion should be conducted. During the discussion, students had access to the written materials provided at the beginning of experiment. After the discussion, students were asked to complete a 20-min posttest. Students who finished early had the option to review their answer sheets and leave the classroom.

Day 9: the researcher administered the long-term test for 20 minutes, and provided students additional time if more time was needed. The long-term measures serve as a type of dynamic assessment (Bransford & Schwartz, 1999; Haywood & Tzuriel, 2002), measuring students' ability to transfer their learning to authentic instructional situations.

Result and discussion

Immediate learning

Immediate retention. Analysis of variance (ANOVA) was performed on the students' date from the immediate screening test and posttest. ANOVA performed on the screening test data did not show significant differences between conditions, $F(1, 71) = 0.55$, $p = .460$, partial eta squared = 0.01. There was no significant difference between the two conditions in their prior knowledge on criminal procedure code and accusation.

Table 2.5
Means and standard deviations of two discussion conditions in immediate and long-term learning

	Immediate retention		Immediate near-transfer		Long-term retention		Long-term near-transfer	
Condition	M	*SD*	M	*SD*	M	*SD*	M	*SD*
Discussion after lecture	3.63	*1.40*	6.40	*3.04*	2.54	*1.27*	4.74	*2.77*
Discussion after self-study	4.32	*1.38*	9.40	*2.75*	3.97	*1.44*	8.82	*2.58*

We performed a priori orthogonal contrasts to test our prediction. The discussion after self-study group received higher scores than the discussion after lecture group in retention scores, $t(71) = 2.12$, $p = .038$, Cohen's $d = 0.47$. Thus, it seems that students who had small group discussions after self-study learned more than those who had discussions after watching lectures according to the immediate retention data (see **Table 2.5**).

Immediate near-transfer. In line with the result of retention scores, scores for discussion after self-study condition were significantly higher than those for discussion after lecture condition, $t(71) = 4.42$, $p < .001$, $d = 1.03$. Therefore, the combination of self-study and discussion improved the immediate learning outcomes more than that of watching a lecture and discussion.

Long-term learning

Long-term retention. We performed a priori contrasts to test our prediction. The results show that in the long-term retention, students in self-study and discussion condition had higher scores than those in lecture and discussion condition, $t(71) = 4.49$, $p < .001$, $d = 1.05$.

Moreover, students in the lecture and discussion group showed a significant drop in scores over time, $t(68) = 3.41$, $p < .001$, $d = 0.82$, while students in the self-study and discussion group revealed no difference over time, $t(74) = 1.06$, $p = .294$, $d = 0.25$.

Long-term near-transfer. We again conducted a priori contrasts to determine differences between groups. The test revealed that students in the self-study and discussion condition scored significantly higher than those in the lecture and discussion condition, $t(71) = 6.51$, $p < .001$, $d = 1.52$.

Interestingly, we confirmed similar results to those regarding scores of the retention test. For the lecture and discussion group, students' scores dropped over time, $t(68) = 2.38$, $p = .020$, $d = 0.57$, while, for the self-study and discussion group, there were no significant difference, $t(74) = .95$, $p = .346$, $d = 0.22$.

We then compared our results with those from Lim and Park (2023).

Reconciling the results from high school students and undergraduates

In the present study, we investigated whether previous results from prior research could be generalized to a younger population. We thus adapted the materials and methodology of Experiment 3 for Lim and Park (2023), and applied it to high school students and added long-term learning measures. In both studies, the same instructor explained the same concepts and problems in a monologue lecture, using the same multimedia presentations. By contrasting these two studies we wanted to provide robust evidence why student individually studying on their own and then discussing is an effective instructional design.

A comparison of results is shown in **Table 2.6**. We performed a priori orthogonal contrast to compare four groups. Collapsing across age differences, the results from both studies show that the students in the self-study and discussion group learned significantly more than the those in the lecture and discussion group in both immediate retention and transfer tests, $ps < .05$. Collapsing across conditions, undergraduate students had significantly higher scores than high school students in immediate retention and transfer test, $ps < .01$.

Specifically, in both the self-study and discussion group, and the lecture and discussion group, undergraduate students scored higher than high school students in the immediate retention test, $ps < .05$. However, there were no significant difference in scores between undergraduate students in the lecture

Table 2.6
Descriptive statistics for two studies

	Discussion after lecture (LD)		Discussion after self-study (SD)	
	High school	Undergraduate	High school	Undergraduate
Variable	($n = 35$)	($n = 37$)	($n = 38$)	($n = 37$)
Immediate retention				
(corresponding to *Verbatim*)	3.63 (*1.40*)	6.11 (*1.56*)	4.32 (*1.38*)	7.27 (*1.15*)
Immediate near-transfer				
(corresponding to *Transfer*)	6.40 (*3.04*)	8.05 (*1.84*)	9.40 (*2.75*)	11.19 (*2.49*)

Note. Means (and *standard deviations*) is shown in this table. Maximum scores for immediate retention are 10 points and for transfer are 18 points.

and discussion group, and high school students in the self-study and discussion group, $p = .088$. We then analyzed students' dialogue during the discussion to shed light on why students in the self-study group are superior to those in the lecture group in terms of learning. Before explaining the result of the analysis, we provide details on its method.

Analysis of students' dialogue during discussion

In addition to our intentional self-study learning hypothesis, we analyzed students' interaction during discussions to explore potential reasons why the two discussion conditions differ. We thus examined the students' dialogue in which they interacted with peers and the instructional material. To do this, we adapted the methodology used in prior studies (Chi et al., 2008, 2017; Muldner et al., 2014). We first transcribed audio-recorded discussion videos, and subsequently segmented the transcriptions into statements. These segments were then coded by three independent raters.

- Substantive comment: a substantive comment, as referred by Chi et al. (2017), pertains to how students engage constructively in their learning process. When a student state an idea related to the concept studied or taught, the idea was coded as a substantive comment, regardless of whether it was correct or not. However, students were required to state their ideas or thought beyond what was presented in the lecture or written materials, not just repeat it.

  The means and standard deviations of the dialogue analyses are presented in **Table 2.7**. The interrater reliability for this coding was found to be adequate, with an ICC (3, k) value of .90, $p < .001$ (30% of the randomly selected transcripts). First, there was no significant difference between the two groups in overall amount of utterances, $t(19) = 1.79$, $p = .089$, $d =$

Table 2.7

Mean values (standard deviations) for two discussion groups of total students' dialogues

	Condition				
Dialogue	*Lecture Discussion*		*Self-study Discussion*	*p*	*d*
Total utterances during discussions	34.00 (*18.06*)	=	52.27 (*27.22*)	.089	0.79
Substantive comments per discussion group	23.50 (*11.04*)	<	48.36 (*23.22*)	.006	1.21
Substantive comments out of total utterances	72.56%	<	94.15%	<.001	1.37
Interaction episodes per discussion group	3.80 (*1.55*)	=	4.36 (*1.86*)	.462	0.33
Substantive comments per interaction episode	6.44 (*2.03*)	<	12.78 (*6.55*)	.002	1.31
Co-constructive turns per interaction episode	2.46 (*0.85*)	<	6.16 (*6.07*)	.019	1.64

0.79. However, the proportion of substantive comments to total utterances was significantly different, ${}^2(1, N = 19) = 228.64$, $p < .001$, $d = 2.05$. The self-study and discussion group generated more substantive comments than the lecture and discussion group, $t(19) = 3.08$, $p = .006$, $d = 1.37$. Furthermore, each discussion member in the self-study and discussion group generated a greater number of substantive comments (24.53 per person) than each student in the lecture and discussion group (13.62), and this difference was significant, $t(71) = 4.84$, $p < .001$, $d = 1.15$.

- Interaction episode: segments can be combined into an episode, corresponding to a consecutive conversation (Chi et al., 2008). Thus, an episode contains at least one substantive comment from either member of the discussion group. An episode was further coded as an interaction episode if two or more students in the group each provided at least one substantive comment. Detailed examples are shown in **Appendix H**. ICC (3, k) was .97, $p < .001$ (30% of the randomly selected transcripts). However, the difference in the mean number of interaction episodes did not show a significant difference between the two groups, $t(19) = 0.75$, $p = .462$, $d = 0.33$. Accordingly, we further analyzed that the more interactive a group is, the more substantive comments it generates in an interaction episode. The result indicated that students in the self-study and discussion group generated more substantive comments per interaction episode than those in the lecture and discussion group, $\chi^2(1, N = 19) = 10.07$, $p = .002$, $d = 1.31$.

- Co-constructive turn: we used a more sophisticated construct, not just limiting the definition to a change in speaker (i.e., a *turn*). A prior study, which analyzed dialogues from dyads, referred to a co-constructive turn as a change in speaker that contains substantive contributions from both speakers. Since the purpose of our study is to analyze students' dialogue during discussion, we defined a co-constructive turn as the continuation of substantive comments from two or more speakers. Per interaction episode, students in the self-study and discussion group had a significantly higher number of co-constructive turns than students in the lecture and discussion group,, $\chi^2(1, N = 19) = 5.52$, $p = .019$, $d = 1.64$.

To summarize, in this study, students in the self-study and discussion group generated more substantive contributions than did students in the lecture and discussion group, and this pattern corresponds to their learning, in that the self-study and discussion group had higher scores than the lecture and discussion group. Therefore, we now need to provide an explanation for the differences in student learning through designing a possible model.

Discussion to learning outcomes

To shed light on our interpretation that substantive comments had positive impact on the learning outcome, we designed a mediation model. A simple mediation analysis was conducted using ordinary least square path analysis to examine whether the effect of discussion conditions (i.e., lecture and discussion or self-study and discussion) on learning outcome (i.e., sum of immediate learning measures) was mediated by substantive comments. The mediation analysis was conducted using the PROCESS macro for SPSS (Model 4) (Hayes, 2017). A percentile bootstrap estimation with 5,000 replications was used to generate a 95% confidence interval (CI). The effects were considered significant if the interval between the lower limit and upper limit of the CI did not include zero. Since discussion conditions are categorical variables, we transformed them into

Table 2.8

Parameter estimates for the simple mediation model

Model	Estimate	*SE*	*p*	95%CI (lower)	95%CI (upper)
Model without mediator					
Intercept	6.15	1.34	$<.001$	3.34	8.97
Discussion condition → Learning outcome (*c*)	3.85	0.84	$<.001$	2.09	5.60
$R^2(y, x)$	.53				
Model with mediator					
Intercept	6.25	1.03	$<.001$	4.09	8.40
Discussion condition → Substantive comment (*a*)	24.86	8.07	.006	7.96	41.76
Substantive comment → Learning outcome (*b*)	0.07	0.02	.001	0.03	0.11
Discussion condition → Learning outcome (*c'*)	2.12	0.78	.015	0.47	3.76
Indirect effects (a × b)	1.73	0.78		0.46	3.38
$R^2(m, x)$	.33				
$R^2(y, m, x)$	.74				

two binary dummy variables (0, 1). In this transformation, 0 corresponded to the lecture and discussion condition, while 1 corresponded to the self-study and discussion condition.

Overall, our model significantly accounted for 73.86% of the variance in learning outcome, $F(2, 18) = 25.43$, $p < .001$. This analysis confirmed that there was a significant total effect of discussion condition on learning outcome (c), and this relationship was also significant when the effect of substantive comments was considered (c'). This analysis reveals that the discussion condition is positively related to substantive comments (a) and substantive comments is also positively related to learning outcome (b). See **Table 2.8** for all parameter estimates.

2.2.3 Discussion

An important goal of education is to prepare students for future learning. In line with this belief that underlines education, educators and educational institutions hope that students are prepared to learn in subsequent activities, future classes, and throughout their lives as lifelong learners. However, despite these hopes, higher education often overlooks the goal of preparing students for learning and instead focuses on selecting 'superior' students from secondary schools. In light of this context, it becomes imperative to reconsider our approach to education, placing a greater emphasis on preparing students to be more proactive and interactive participants in the learning process.

The present study investigated the combination of watching a lecture and discussing with peers in contrast with studying on one's own and discussing with peers. Building upon previous findings (Lim & Park, 2023), we examined whether self-study enhances the learning outcomes of high school students during discussions. We went beyond a mere comparison of the immediate effectiveness of these two different instructional combinations. Additionally, we proposed the intentional self-study hypothesis, suggesting that students in the self-study and discussion group would demonstrate superior learning outcomes compared to those in the lecture and discussion group.

Comparing the two discussion conditions, we successfully replicated the findings of Lim and Park (2023). That is, in consistent with the previous finding that undergraduate students in the self-study and discussion group were superior to the lecture and discussion group, we obtained similar results regarding immediate learning measures also with high school students. We further observed a same pattern between immediate learning and long-term learning measures. Both in immediate learning and long-term learning, the discussion after self-study group outperformed the discussion after lecture group. Thus, significant differences between the two conditions in the immediate and long-term data are both consistent with the intentional self-study hypothesis.

The observed results could be due to students' self-regulation ability or their monitoring of their own learning processes. We assume that students were capable of perceiving, to a considerable extent, their level of knowledge and the areas they needed to work on while independently regulating their learning. In other words, self-study would have allowed students not only to recall the instructional material immediately but also to gain a deeper comprehension required for long-term learning.

Moreover, the findings were robust as the performance of high school students in the self-study and discussion group did not significantly differ from that of college students in the lecture and discussion group. Thus, our findings suggest that students may not require daily didactic lectures, indicating the potential to reduce the costs related to recruiting an experienced teachers in schools.

Nonetheless, in order to figure out how to actually design and apply this combination in a real classroom, we need to understand why the combination of self-study and discussion is more effective than the combination of lecture and discussion, even though all the instructional materials (video lecture and written texts) were the same. We thus examined students' dialogues during the discussions after the lecture and after the self-study groups to shed light on this matter. The results showed that the self-study and discussion group generated more constructive interactions (i.e., substantive comments, interaction episodes, and co-constructive turns) than the lecture and discussion group. Interestingly, this difference was not simply due to the fact that the self-study group was active, i.e., they talked a lot, but that they also made more substantive contributions during the discussion. The results showed the benefits of self-study over lecture as a preparation activity for discussion as it helps subsequent discussion to become more active and constructive. These results are also meaningful because it replicates the results of previous literature (e.g., Chi et al., 2017; Lim & Park, 2023; Muldner et al., 2014). Furthermore, this effect is not simply due to an increase in the total number of utterances, but rather to the increase in the proportion of substantive contributions among the total utterances.

Through further analysis, we confirmed the self-study and discussion condition had a stronger positive relationship to posttest achievement compared to the lecture and discussion condition. The mediation effect of substantive comments on posttest scores was also stronger for the combination of self-study and discussion compared to the combination of lecture and discussion.

In K-12 classrooms, teachers still believe that "Tell and Practice" instruction is important because it delivers explanations and solutions invented by experts, and students need opportunities to hear and practice these ideas. However, considering previous findings (Lim & Park, 2023) repeatedly demonstrating that discussion after lectures is superior to traditional instruction such as reviewing after lectures, it is recommended that teachers increase the incorporation of discussion activities in the classroom. Specifically, intentional self-study promotes a state of “readiness” for subsequent discussions in part by encouraging students to generate, compare, and contrast a range of ideas and/or questions. This prepares students to learn more effectively during the subsequent discussion due to the primed contents and heightened awareness. As a result, this combination strategy results in improved performance, especially in terms of transfer.

Our study materials, based on the domain of law, specifically criminal procedure, were appropriate because it is a topic that is relatively unaffected by students’ prior knowledge and is convenient to measure transfer of knowledge. However, follow-up research needs to expand to different topics such as molecular diffusion, a highly misconceived and challenging topic for a student to learn (e.g., Chi et al., 2012). Also, future research can also measure changes in student learning with the content students actually learn in class.

In addition, further examination can be conducted in classrooms targeting younger age groups such as primary and secondary schools. It is time for both K-12 educators and higher education instructors to step out of their comfort zones, break down barriers, and immerse themselves in each other’s domains. Only through this collaboration can both sides truly innovate, similar to their efforts in 1952, in order to shape the education of today.

Chapter 3

3.1 Effects of Manipulating Individual Preparations for Different Classroom Discussion on Student Learning

3.1.1 Introduction

Discussion is an effective instructional strategy in the classroom. Extensive literature on discussions have suggested that classroom discussion can be an effective instructional strategy for student learning such as text comprehension (e.g., Almasi & Garas-York, 2009 for a review; Murphy et al., 2009; Soter et al., 2008). Although these studies showed that discussions can enhance student learning outcomes, their classroom application is limited because it is not easy to convert class content into a debate or discussion format, and training of techniques for either teachers or students is require in these cases (Chin & Osborne, 2010; Kuhn & Crowell, 2011; Reznitskaya et al., 2012). This is true not only in elementary and secondary school but also in college courses (Stains et al., 2018; Yannier et al., 2021). Thus, the question is no longer whether discussion is an effective strategy in the classroom, but is how to utilize and design discussions in classroom.

Simply placing groups of students together in a classroom does not guarantee improved learning (Chi et al., 2017; Craig et al., 2009, Lam & Muldner, 2017; Schwartz & Bransford, 1998). Lam and Muldner (2017) proposed that an appropriate preparatory manipulations inspired by the preparation for future learning to examine the role that individual preparation has on subsequent discussions. The result showed that studying alone first and then collaborating is more effective than collaborating from the beginning. These studies have shown that different learning activities can be combined to promote student learning, as results of Schwartz and Bransford (1998)'s seminal study.

In this context, we adopted the framework of Lam and Muldner (2017) while focusing on classroom discussions followed by different individual preparations, lecture or self-study. Especially, our purpose of research is to investigate the learning effect of the self-study and discussion combination. Thus, the present two-part study focused on different individual methods to prepare for classroom discussion to design and investigate which combination of preparatory activity with classroom discussion most impacted student learning.

Individual preparation for discussion

Individual preparation before subsequent collaborative learning enhances student learning (Mende et al., 2021). Mende and colleagues investigated whether individual preparation affects retrieval, inferencing, and referencing differently on subsequent collaborative learning. They predicted that generative individual preparation tasks and provision of awareness support can improve the advantages and ameliorate the disadvantages of individual preparation before collaboration. The results indicated that individual preparation per se might not automatically enhance learners' interactive activities, but that the application of information obtained during the preparation may depend on the task type (generative or non-generative), and task complexity. Thus, Mende et al. suggested that teacher can use students' individual preparation time in terms of generating new conclusions, connections, or ideas which could be discussed and further developed during subsequent collaboration.

In line with the idea that students need individual preparation for collaborative learning, the preparation for future collaboration has focused on whether and how students should cognitively prepare themselves prior to collaboration (Lam & Kapur, 2018; Lam & Muldner, 2017). In Lam and Kapur's study, for example, the generative preparation and non-generative preparation before collaboration were compared. Students in the generative condition produced more ideas during preparation compared to the non-generative condition. In transfer task, the generative condition performed better during the preparation phase and subsequently performed better on the collaboration phase. In other words, before collaboration, students who have a generative preparatory activity learn more in a subsequent collaboration. In collaboration phase, students collaborated with peers through discussions. Student discussed their knowledge, and their only access to the canonical forms of the concepts can come via the discussion. Lam and Kapur thus addressed in their paper that student learning was enhanced when students prepare by generating knowledge and then discuss in a future task. Therefore, it seems to be a reasonable instructional strategy for student to make individual preparation before classroom discussions.

Similarly, productive and effective discussions in classroom cannot arise naturally. Students need a basic understanding of what they are learning and access to newly acquired knowledge. Although students should be primed, not just possessing the knowledge (e.g., McNamara et al.,1996; Murphy et al., 2016), individual preparations for discussions has often been rarely applied (e.g., Engelmann et al., 2009) or was often not subject to systematic investigation. Therefore, we examined different individual preparation and discussions where students work together with peers to share their internal thoughts aloud on what they know or have learned to demonstrate learning or continue to learn during discussion has shown to promote student learning.

Manipulating individual preparation

Research has shown that without proper learning preparation with clear objectives, simply placing students in the classroom will not guarantee an effective collaborative learning outcome (Gadgil & Nokes-Malach, 2012; Lam & Muldner, 2017). How then can instructors design instruction to prepare students for discussion that leads to effective learning? What is the best combination of instructional strategies and sequence?

In this study, we compare lectures and self-study. Comparing lecture and self-study is educationally meaningful. Researchers have posited that conventional lectures frequently do not result in deep learning (Barrows & Tamblyn, 1980; Boud & Feletti, 1998; Freeman et al., 2014; Graffam, 2007), fail to promote students' high level of comprehension (Bligh, 1998), and cannot guarantee students' performance (Carpenter et al., 2013; Lim et al., 2019) Although lectures give satisfaction to both instructors and the students, there are many studies that show that the learning outcome of lectures is not very high (e.g., Deslauriers et al., 2019; Stains et al., 2018). However, since lectures are still the mainstay in various disciplines (Schmidt et al., 2015; Stains et al., 2018; Steele, 1997; Steele & Aronson, 1995), it may not be a good way to eliminate lectures at once or to reduce them unconditionally.

Self-study is a representative candidate to replace the lecture in that it involves agency but also serves as a basis for cooperative learning such as discussions with others (e.g., Fiorella & Mayer, 2016; Lam & Kapur, 2017; Lam & Muldner, 2018; Mende et al., 2021). While reading a learning material on their own, students can engage in a constructive cognitive activity such as self-explaining (e.g., Bisra et al., 2018; Chi & Wylie, 2014). For instance, Lim and Park (2023) found that self-study enhances the learning effect of discussions, compared to lectures. They compared three combination groups: lecture and review, lecture and discussion, and self-study. The results showed that discussion groups outperformed review groups, and self-study groups scored higher than lecture groups. Through dialogue analysis, they also found that self-study made students more productive in discussions than lectures.

However, in Lim and Park's study, the discussions were only student-led discussions. There may be a counterargument that the teacher's role in the discussion is important since teachers can participate in discussions depending on their instructional strategies. To address the question of whether instructors are necessary in discussions, we further investigated the impact of types of classroom discussion on student learning.

Formats of classroom discussion

Discussion studies in the classroom have mixed results because there are many ways to conduct a classroom discussion. Research has found that student-led small group discussions (or peer discussions) are effective in learning (Almasi, 1995; Lim et al., 2019, 2022; Murphy et al., 2018; Schmidt et al., 2007). For example, Murphy et al. (2018) conducted year-long study of Quality Talk, a teacher-facilitated, small-group discussion approach to enhance students' comprehension. After analysis of discussions, they argued that small-group discourse is an effective instruction foster individual student learning outcomes.

On the other hand, some studies have found that teacher-led discussions improve students' understanding and serve as a scaffolding for students' understanding (e.g., Almasi & Garas-York, 2009; Sweller et al., 2007). In particular, since teachers should give direct and explicit scaffolding during the discussion and summative feedback after the discussion, there are advantage in teacher-led discussions (Murphy et al., 2016; Sweller et al., 2007). That is, the role of teachers implies that the discussion formats can enhance students' outcome according to types of individual preparation. Because, students learn individually and then paying attention to and taking up content-relevant information externalized by teachers during discussions can affect student' learning (referencing; cf. Mende et al., 2021; Teasley, 1997; Weinberger et al., 2007).

In the current study, in line with Lim and Park's findings, the goal was to determine whether interaction of individual self-study and student-led discussion is a more effective instructional design for learning, compared to other combinations. Therefore, we expanded previous literature by reviewing discussion formats and linking them to improvements in students learning gains.

3.1.2 Experiment 1

The current study examined individual preparatory activities before discussions. In particular, we investigated how the use of self-study and subsequent discussions influenced students' learning, as well as replicating the findings of Lim and Park (2023). We set conditions for individual preparation, the lecture condition for students to watch a video on a personal screen, and the self-study condition for students to study the learning materials on their own.

Method

Design and participants

Students were able to voluntarily apply for this experiments for the course credit from psychology class. Undergraduate students from science, technology, engineering, and mathematics (STEM) disciplines participated in this experiment. All student were Asian, and the college was located in a mid-high socioeconomic area of the city. The participants came to laboratory on time for their application. The study was conducted early in the second semester of the academic year.

A total of 91 students were assigned to two experimental conditions at random: a traditional lecture before small group discussion condition (LS group, n = 46) and a self-study and small group discussion condition (SS group, n = 45). The average age of the participants was 19.40 [standard deviation (SD) = 0.97] years.

An ethical statement was supervised and approved via the selective university's Institutional Review (approval No. S-D20210031) in 2019 and 2021.

Materials

Individual preparation. A total of study time and materials in the individual preparation was equivalent across the two groups in that two groups were designed to target the same shallow and transfer knowledge about learning content (see **Appendix I**).

Students in the LS groups listened a monologue-style lecture video with a seven-page long learning material. The lecture was a law-related topic that had nothing to do with the background knowledge of the STEM students. In the SS group, students received an same instructional material, but did not take lectures.

Items for learning activity. Items for the shallow knowledge were items were made up of ten questions that were factual recall and superficial concepts from an instructional material or lecture video. The test used identical True/False questions and multiple-choice questions. Thus, there were no scoring error or subjectivity for the shallow knowledge questions, ICC (3, k) = 1.00, $p < .001$.

The transfer knowledge items consisted of four questions, measuring deep knowledge, required students' novel application, evaluation and explanation.

Detailed samples of the learning materials and test items are shown in **Appendix I**. The transfer knowledge items were worth a total of eighteen points for four questions. Thus, although there was an exemplary answer to these questions, students could not ask for an exact answer.

Accordingly, intra-class correlations (ICC) were calculated. The authors' subjectivity of scoring for transfer items was minimized by randomly selecting 30% final test papers of the total, which were graded by first author and other two independent raters. Each rater scored each of the two conditions with an exemplary answer sheet. The agreement evaluated by ICC $(3, k) = .91$, $p < .001$, by which values greater .80 indicate excellent reliability. As a result, the remaining final test paper was graded by the first author.Samples of scoring rubrics for transfer items are shown in **Appendix J**.

Procedure

Students were briefed that participation would be anonymous and that it would not affect grade of psychology class, and course credits would be awarded for attending this experiment.

Two researchers organized the learning activities. Students were randomly assigned two different groups worked in separate laboratories. In the individual preparation activity phase, for the LS group, students individually spent watching the lecture video. Students in the LS group were unable to manipulate the video. Students were encouraged to attend all lectures with concentration. Individual students in the SS group studied the written learning material on their own. Students in the SS groups received learning materials at the beginning of individual learning. They were given 18 minutes to complete the preparation activity. Students were free to take notes or underline given learning materials, but they were not able to use them on tests. At the end of the individual prepa-

Table 3.1
Descriptive statistics for groups in Experiment 1

Learning outcome	LS ($n = 46$)	SS ($n = 45$)	p	η_p^2
Total score	13.37 (*2.03*)	17.89 (*3.35*)	<.001	0.41
Shallow knowledge	7.65 (*1.20*)	8.67 (*0.95*)	<.001	0.17
Transfer knowledge	5.72 (*1.94*)	9.22 (*3.15*)	<.001	0.33

Note. Data represent the means (*SDs*). Total, shallow, and transfer knowledge items scores are given for both LS and SS groups. LS = lecture and student-led small group discussion; SS = self-study and student-led small group discussion.

ration phase, the students stopped learning and waited for the experimenter's next instruction.

In the discussion phase, students from either the LS group or the SS group were randomly assigned to a group of three or four students who had learned the same individual preparation. The researcher did not intervene during discussions to ensure that the students had student-led small group discussion. Students engaged in the discussion phase, such as exchanging the learned contents with each other or continuing insufficient learning in the individual preparation before discussions with peers. Following the 20-minute discussion, two discussion groups took a 20-minute final test on the same content. Students finished before 20 minutes passed were asked to go over their answer sheets.

Result and discussion

The means and *SD*s of the final exam for total, shallow, and transfer knowledge items are shown in **Table 3.1**. There are no missing data, and outliers.

The SS group displayed a significantly higher total score than that of the LS group, $F(1, 89) = 61.43$, $p < .001$. In specific, the SS group had a significantly greater score for the shallow knowledge than the LS group, $F(1, 89) = 18.28$, $p < .001$. The SS group scored significantly greater in transfer knowledge than the LS group, $F(1, 89) = 42.87$, $p < .001$.

As a result, before the student-led small group discussions, it is effective for students' learning to involve in self-study. In line with the findings of Lim and Park (2023), our results support the fact that self-study boosts student learning than watching lectures before subsequent discussion. It might be also explained that self-study group may get more opportunities for self-generated learning on novel learning materials compare to lectures.

However, using a learning content far from students' background and majors may have lowered experimental reliability. In addition, although a laboratory study can provide valuable data since the controlled settings and variables allow us to understand the details of the obtained results, one can argue that Experiment 1 is not the reliable data for educational studies because of the artificial settings. Accordingly, it is necessary to use learning content related to students' discipline.

3.1.3 Experiment 2

Experiment 2 was run an in situ lesson on covalent bond. We attempted to ensure ecological validity through introductory biology classroom and relevant learning material with students' disciplines such as biology, which is fundamental knowledge for undergraduate STEM discipline.

In addition, taking into account the teacher's scaffolding such as support or intervention for the students that occurs naturally in the classroom, we investigated whether discussion formats (teacher-facilitated small group or student-led small group) worked differently when individual preparation preceded the discussion. We also analyzed the recorded discussions using dialogue analysis.

Method

Design and participants

Independently of the participants in Experiment 1, a total of 113 participants were taken for the undergraduate introductory biology course at selective university. The average age of the participants was 19.96 ($SD = 4.61$) years, with no significant differences in group or gender. The study used a 2 × 2 experimental design examining the factors of types of individual preparation (watching a video lecture or self-studying) and types of discussions (student-led small group or teacher-led small group) thus, with four conditions: a lecture and student-led small group discussion (LS group, $n = 31$), a lecture and teacher-led small group discussion (LT group, $n = 24$), a self-study and student-led small group discussion (SS group, $n = 31$), as well as self-study and teacher-led small group discussion (ST group, $n = 27$). Experiments were conducted in different classrooms for each condition, and one instructor conducted a teacher-led discussion to avoid instructor-specific effect.

Materials

Instructional materials and test items. The 20-minute didactic lecture by an instructor was recorded before experiment. The video lecture was on introductory biology, specifically regarding on covalent bonds, non-covalent bonds, the octet rule, and water for life. The lecture and six-page written instructional materials were equivalent across the four classrooms.

All of the final test items were transfer knowledge items with 11 questions. These were open-ended items, multiple choice items, and essay items. Samples of learning materials and test items are shown in **Appendix K**. The total score of eleven transfer knowledge items was 33 points. Forty-one final test papers were randomly graded by two additional raters, as in Experiment 1 (30% of the total). The agreement evaluated through intraclass correlation was .87 for the analysis [ICC (3, k)]. Thus, the remaining final exam paper was graded by the first author.

Procedure

Experiment 2 was carried out in the second semester of the academic year. Students in the student-led groups (the LS and SS groups) had taken part in at least three small group discussions before this experiment. Thus, by the time the data were collected students were quite familiar with this approach to discussion, with peer-cantered participation in which they spoke freely without the instructor nomination or intervention. Students in the teacher-led groups (the LT and ST groups) had a small group discussion with teacher-facilitated or supported participation. For instance, students bid for turns by raising their hands and waiting to be nominated by the teacher.

In detail, participants were randomly divided to four discussion groups: LS, LT, SS, and ST. The individual preparation phase such as watching a recorded lecture or self-studying lasted 20 minutes. Same as Experiment 1, students in the SS and ST groups received learning materials at the beginning of individual learning. Students can take notes or underline given learning materials, but they were not able to use them on tests.

After the individual preparation, the students stopped learning and were divided to groups of three or four by the instructor. The discussion lasted 20 minutes, and the instructor did not intervene in the discussion to ensure they had an un-structured discussion among participants in the LS and SS groups. On the other hand, students in the LT and ST groups continued the discussion by questioning and answering to the instructor. The instructor moved around the classroom providing students with incidental learning or scaffolded instruction.

Subsequently, the four groups each took a 20-minute test on the same learning material.

Result and discussion

The main effect of the research condition, lecture versus self-study was significant, $F(1, 109) = 42.43$, $p < .001$. As a result, the main effect of the type of discussion condition, teacher-led small group discussions versus. student-led small group discussions displayed no significant difference, $F(1, 109) = 2.56$, $p = .113$. The interaction impact of these two factors was significant in terms of students' learning benefits, $F(1, 109) = 10.27$, $p = .002$.

The means and *SD*s of the final test scores are summarized in **Table 3.2**. Significant differences existed between the four groups in total score, $F(3, 109) = 20.10$, $p < .001$, partial eta-squared = 0.36. The SS group had the highest score for transfer knowledge items and all groups, $ps < .05$, exception of the ST group displayed significant differences, $p = 1.00$. The ST group found no significant difference in comparison with the LT group, $p = .166$, but scored significantly greater than the LS group, $p < .001$. Students in the LT group scored significantly greater than the LS groups, $p < .05$.

Thus, using relevant content and discipline for students, these results successfully replicated Experiment 1. It is difficult to robustly compare the outcomes of the two studies. In summary, the differences (% of the total) of transfer test between the LS and SS groups were analyzed. In Experiment 1, there was a significant difference in the mean value between the LS group and the SS group in transfer, $p = .021$. In Experiment 2, there was a significant difference

Table 3.2
Descriptive statistics for groups in Experiment 2

Learning outcome	Mean	*SD*	*F*	*p*	η_p^2
Condition 1:					
Individual preparation			42.43	<.001	0.28
Lecture ($n = 55$)	18.91	2.98			
Self-study ($n = 58$)	22.95	3.49			
Condition 2:					
Type of discussion			2.56	.113	0.02
Teacher-led ($n = 51$)	20.53	4.44			
Student-led ($n = 62$)	21.53	2.83			
Condition 1 × Condition 2		10.27	<.001	.09	
Groups ($N = 113$)					
LS ($n = 31$)	17.68	2.97			
LT ($n = 24$)	20.50	2.15			
SS ($n = 31$)	23.39	3.80			
ST ($n = 27$)	22.44	3.08			

Note. Data represent values of means and *SD*s of transfer knowledge and are detailed for main effects, interaction effect, and four groups. LS, lecture and student-led discussion; LT, lecture and teacher-led discussion; SS, self-study and student-led discussion; and ST, self-study and teacher-led discussion.

Table 3.3
Summary of Experiment 1 and Experiment 2

Test item	Experiment 1		Experiment 2		ps
	LS ($n = 46$)	SS ($n = 45$)	LS ($n = 31$)	SS ($n = 31$)	
Transfer knowledge	31.78% (10.78%)	51.22% (17.50%)	53.58% (9.00%)	70.88% (11.52%)	<.001

Note. LS, lecture and student-led discussion, SS, self-study and student-led discussion, transfer type item scores are given.

between transfer items in the LS and the SS groups, $p < .001$, which was in agreement with the results of Experiment 1 (**Table 3.3**).

Thus, self-study as an individual preparation for subsequent discussion was shown to improve student learning, regardless of discussion formats. Furthermore, the combination of self-study and student-led discussion improved student learning compared to other combinations. However, it is also significant that the students' performance in the lecture group improved with teacher-led discussions.

To interpret these results, we subsequently examined exploratory discussion analysis on our recorded dialogues in classroom to attempt to clarify the results. We utilized verbal protocols from Chi (1997) and coding strategy from Chi et al. (2017).

Students in student-led discussion

To shed light on our results, we analyzed the content of discussions.

The present research categorized the meaningful content of student-led discussions into one major type: substantive (also constructive) comments (Chi et al., 2017). Coding schemes for discussions are shown in **Appendix L**. We had two sets of data to analyze to capture students' dialogue in discussions. We first transcribed the dialogues of twelve video-recorded groups in two conditions (i.e., the LS and SS groups). And then we segmented in to statements following the analyses of Muldner et al.'s (2014). All dialogues were segmented into unit such as total utterances at the phrase level.

We then checked students' substantive comments to peers. Regardless of whether the comment or answer is correct or not, the substantive comment is based on content in lectures and learning materials, and furthermore, the comment served as the starting point, elaboration, and generation of the preceding comments and questions to actively interact. But, it was not substantive that the student simply answered peers' question and no further interaction took place (e.g., meta-talk and off-talk). Non-substantive comments were either comments irrelevant to the learning content or those such as "I agree with you," or "Oh." Example of excerpt is shown in **Table 3.4**.

Three raters coded 30% of the interacting dialogues and ICC (3, k) was .81, $p < .001$. Remaining dialogues were coded by first author. Students in the self-study group spoke more in student-led discussions than did students in the lecture group (38.90 vs. 30.80, respectively). More importantly, students in the self-study group generated more substantive comments than students in the lecture group (34.00 vs. 24.40, respectively). Furthermore, there was significant difference between the SS and LS groups in the proportion of substantive comments out of total utterances (87.40% vs. 79.23%, respectively).

Therefore, the result shows students who self-studied as an individual preparation before student-led discussion are more active and constructive than those who listened to lecture.

Table 3.4
Excerpts from dialogue transcripts from the student-led discussion groups

Transcripts	Utterances among students	Category
T1 (*p*. 1)	Student 1: Can you explain this? What exactly is the van der Waals force?	Statement (Substantive)
	Student3: van der Waals forces are intermolecular forces.	Statement (Substantaive)
	Student 2: It is a weak force of attraction between electrically neutral molecules that only acts at extremely short distance. For example, oil molecules hold together because of van der Waals forces.	Statement (Substantive)
T5 (*p*. 4)	Student 5: Sodium and cholrine were opposites, right?	Statement (Substantive)
	Student 4: That's right. The number of outermost electron shells is different.	Statement (Substantive)
T7 (*p*. 1)	Student 6: Are we doing this right?	Statement (Not substantive, Meta-talk)
	Student 7: Maybe?	Statement (Not substantive)

Table 3.5
Excerpts from dialogue transcripts from the teacher-led discussion groups

Transcripts	Utterances among students	Category
T2 (*p.* 2)	Student A: Carbon has 6 protons and 6 electrons, with 2 electrons spinning in the first shell and 4 electrons spinning in the second shell. Is this second shell where the electrons that actually participate in the chemical reaction are?	*Question* (Substantive)
	Instructor: Excellent. This second electron shell is called the outermost electron shell.	Elaborative feedback
	Student B: What does it mean exactly?	*Question* (Substantive)
T4 (*p.* 5)	Student C: As the electrons drop off, the elements are released and then filled with electrons to follow the octet rule again forming new bonds between the elements.	Statement (Substantive)
	Student D: By what rule are the electrons bound?	Question (Substantive)
T6 (*p.* 1)	Student E: How much discussion time is left?	Statement (Not substantive, Off-talk to the Instructor)
	Instructor: About fives minute.	Simple answer (Not substantive)

Students in teacher-led discussion

Since students in the lecture group scored significantly higher in the teacher-led discussion than in the student-led discussion, we subsequently analyzed the dialogues to investigate the students' talk with the instructor influence learning in teacher-led discussion. Interacting with the instructor could plausibly occur since student created many opportunities for themselves to actively engage in constructive activities (e.g., Chi et al., 2017). We checked students' deep questions (i.e., substantive questions) because literature has been shown to enhance students' learning (e.g., Craig et al., 2006). In addition, since Narciss (2007) showed that instructor (or tutor)'s elaborated feedback facilitate students' learning, we categorized the meaningful content of teacher-led discussions into three major type: students' substantive comments among peers, substantive questions to instructor, and instructors' elaborative feedback (cf. Chi et al., 2017). We transcribed the video-recorded dialogues of eight LT groups and nine ST groups. First, we segmented an instructor's utterances into statements and identified those that referred to students' questions and answers (correct or not). We then coded these instructor's statements as elaborative feedback if the instructor remarked elaborative explanation beyond a simple answer. We also coded students' utterances among peers. Excerpt are shown in **Table 3.5**.

Three raters coded 30% of the dialogues and ICC (3, k) was .97, $p < .001$. Thus, remaining dialogues were coded by first author. **Table 3.6** summarizes the results. There was no significant difference between the ST and LT groups (34.22 vs. 35.13, respectively). In detail, there was also no significant difference in students' utterances among students between the ST and LT groups (31.44 vs. 30.50, respectively), also in substantive comments (26.11 vs. 24.88, respectively). There was no significant difference between the ST and LT groups in the proportion of substantive comments among the total utterances (83.05%

vs. 81.57%, respectively).

However, for students' questions to instructor, students in the LT asked more to instructor than students in the ST group (4.63 vs. 2.78, respectively). Also, students in the LT group asked the instructor more deep questions than students in the ST group (4.00 vs. 2.11, respectively). But, there was no difference between the ST and LT groups in the proportion of substantive questions among the total questions (75.90% vs. 86.39%, respectively).

Lastly, instructors gave more elaborated feedback to students in the LT group than students in the ST group (3.88 vs. 2.00, respectively). The results showed that not only was there more interaction between students in group, but also deeper questions were asked to the instructor, and more sophisticated feedback was provided from the instructor.

Exploratory regression analyses

In the student-led discussion, the SS group had significantly higher learning outcomes and number of substantive comments than those of the LS group. On the other hand, in the teacher-led discussion, the differences in scores and substantive comments between the ST group and the LT group were not significant. Thus, we expected that the type of individual preparation may have influenced the number of substantive comments depending on discussion formats. We further examined whether the number of substantive comments had an impact on students' learning. To substantiate our interpretation, we analyzed the relationship of substantive comments among peers and learning outcome depending on type of individual preparation (i.e., lecture or self-study) and discussion format (teacher-led or student-led), by running an exploratory linear regression. We set final test scores as the dependent variable, substantive, type of individual preparation and substantive $\times$ type of individual preparation interaction as the

Table 3.6
Distribution of utterances in student-led and teacher-led groups

Discussion format		Type of individual preparation			p	d
Student-led discussion		Lecture		Self-study		
	Total amount of utterance	30.80 (*6.55*)	<	38.90 (*6.47*)	<.01	1.24
	Substantive comments per group	24.40 (*5.10*)	<	34.00 (*7.01*)	<.001	1.57
	Percentage	79.23%	<	87.40%	<.05	
Teacher-led discussion						
	Total amount of utterance	35.13 (*5.28*)	=	34.22 (*5.02*)	.361	0.17
	Total amount of utterance among student	30.50 (*5.18*)	=	31.44 (*3.68*)	.334	0.26
	Substantive comments per group	24.88 (*5.44*)	=	26.11 (*1.54*)	.164	0.31
	Percentage	81.57%	=	83.05%	.263	
	Total questions to instructor	4.63 (*1.60*)	>	2.78 (*1.92*)	<.05	1.23
	Percentage	86.39%	=	75.90%	.187	
	Elaborative feedback from instructor	3.88 (*1.36*)	>	2.00 (*1.45*)	<.01	1.45

Means (*SD*s) of students' utterances with peers or instructors according to discussion format and type of individual preparation

Table 3.7
Exploratory linear regression for student-led and teacher-led discussions

Predictor	**Unstandardized coefficient**	**Standardized coefficient**	t	p
Student-led discussion				
Substnative × Individual preparation	1.87	5.97	2.88	.013
Constant	6.86	-	1.44	.175
Substantive comment	0.55	0.50	2.91	.012
Individual preparation	-47.39	-5.79	-2.79	.015
Teacher-led discussion				
Substnative × Individual preparation	-0.24	-1.02	-2.19	.043
Constant	1.98	-	0.87	.397
Substantive comment	0.63	1.11	7.04	<.001
Individual preparation	8.11	0.98	2.59	.020

explanatory variables. Since type of individual preparation is a categorical variable, we transformed condition into two binary dummy variables (0, 1), where 0 corresponded to lecture and 1 corresponded to self-study.

For the student-led discussion, the overall model was significant, (adjusted $R^2 = .91$), $p < .001$ (**Table 3.7**). Substantive comments and type of individual preparation were significant, and it indicates that substantive comments among peers and self-study have positive relationship with scores for final test. Overall, substantive comments during student-led discussion has a modest but stronger positive relationship to transfer knowledge to self-study compared lecture. For the teacher-led discussions, the overall model we obtained was also significant, (adjusted $R^2 = .59$), $p = .002$. Likewise, substantive comments, and type of individual preparation were significant, and substantive comments during teacher-led discussion has a stronger positive relationship to transfer knowledge for lecture compared to self-study.

3.1.4 General discussion

We examined whether and how discussions could be designed as an instructional method in the classroom. Experiment 1, which was conducted in the laboratory, compared lecture before student-led discussion (LS) and self-study before student-led discussion (SS). The result showed that the SS group outperformed the LS group. Significant differences in the shallow and transfer knowledge items demonstrated a significant practical implication. In addition, we found that the individual preparation activities made the learning effect of discussion different.

However, we needed to consider the level of prior knowledge. In addition, it was necessary to replicate the results of Experiment 1 in a more natural setting and explore the effect of the teachers support which commonly occur in the classroom. In Experiment 2, we compared four conditions: a lecture and student-led discussion (LS), lecture and teacher-led discussion (LT), self-study and student-led discussion (SS), and self-study and teacher-led discussion (ST). As in the results of Experiment 1, self-study was more advantageous than lectures as an individual preparation of discussions, regardless of discussion formats.

Meanwhile, we noted that there was no significant difference between the LT group and the ST group, but students in the LT group tended to perform lower than students in the ST group. In other words, the instructor's support or intervention seems to have had a positive effect for students who listened to lectures. On the other hand, self-study is self-regulated or self-paced learning, and instructor intervention hurt students' engagement in the discussion.

To substantiate these interpretations, we analyzed dialogues of the two discussion. The results showed that the self-study group made more constructive and substantive comments than the lecture group in the student-led discussions. However, when the instructors participated in the discussion, the students who listened to the lecture were active in interacting with peers, and even actively interacting with the instructor, which improved students' learning. Therefore, our findings indicated that teachers should employ their discourse to scaffold students' thinking according to their instruction or condition (e.g., Hattie & Timperley, 2007). However, one point we wish to emphasize more is that teachers should gradually create classes where students study on their own and have discussions among themselves.

Individual preparation has often been suggested to benefit discussion in the classroom. Yet, so far, the theoretical bases and the empirical evidence have not been systematically addressed in previous literature (e.g., Mende et al., 2021). The present studies addressed these gaps by developing and testing our prediction regarding the claim that individual preparation is important in that it enhances the learning effect of discussions when properly combined. Teachers should gradually reduce their control in the classroom, and then students become more active in the learning process.

Limitations and educational implications of the findings

Our findings indicate that the combination of self-study and student-led discussion enhances student learning. However, as an individual preparation, a dichotomy between self-study and lectures in this study may overlook the fact that lectures can promote students' learning. Our research does not mean that teachers should use self-study instead of lectures right now, but rather that they should gradually adopt a more student-centered approach to teaching. Before that, teachers need to utilize the right design for individual preparatory work and discussion combinations to improve student learning.

The present study targeted STEM students so that it can be pointed out that this study is discipline-specific. This study is still in the early stages of exploring discussions and effective strategies for higher education. Therefore, future studies are required to examine students from other disciplines, ages and other ethnicities.

In addition, we propose that the lecture used in the present study is a microcosm of a class using in higher education in that traditional lectures in that most STEM education still use traditional lectures in classes (e.g., Lombardi et al., 2021; Stains et al., 2018). Accordingly, many studies recommend combining lectures and active learning (Chi, 2009; Chi & Wylie, 2014; Muldner et al., 2014). Our study therefore has pedagogical and practical implications for how to minimize didactic lecture time and make more effective use of lectures in conjunction with or prior to discussions.

3.2 Improving Learning Effects of Student-Led and Teacher-Led Discussion Contingent on Prediscussion Activity

3.2.1 Introduction

Studies have identified engagement as a key factor for students' meaningful learning (e.g., Sinatra et al., 2015; Sun et al., 2022) and have demonstrated its importance in students' positive learning outcomes, academic resilience, and avoidance of dropping out (Bae & DeBusk-Lane, 2018; Bae & Lai, 2020; Chi et al., 2018; Sinatra et al., 2015; Sun et al., 2022; Wang & Eccles, 2013). In this regard, classroom discussions have the potential to enhance students' learning beyond what studying alone would achieve because they provide the opportunity to engage in beneficial behavior, including explaining, questioning, and generating ideas. For example, Lim and Park (2023) compared discussion with review in higher education. After listening to the lecture, students were assigned to discussion or review groups. On comparing final test scores, it was observed that students in the discussion group scored significantly higher on verbatim and transfer questions than students in the review group. Meanwhile, Wu et al. (2013) compared collaborative discussion with teachers' whole class instruction in primary school. Their research showed that collaborative discussion approaches affect students' behavioral and emotional engagement positively.

Although researchers have argued that students' engagement is a key concept of learning effect of group discussions on student learning (Bae & Lai, 2020; Chi et al., 2018; Guthrie et al., 2004; Henrie et al., 2015; Sinatra et al., 2015; Wu et al., 2013), most of the previous research on discussion has focused on outcomes (e.g., Nokes-Malach et al., 2015, Slavin, 2012), remaining unclear whether and how student engagement differs between discussion approaches (Sun et al., 2022). Hence, this study explored classroom discussion as a core learning strategy that facilitates students' engagement in the learning process, enables them to actively interact with their peers, and helps them construct knowledge. It further focused on the degree of cognitive-behavioral engagement that can be detected by moment-by-moment overt behavior during students' discussions.

Simply having a discussion in class cannot guarantee improved student engagement and effective learning (Craig et al., 2009; Gadgil & Nokes-Malach, 2012; Lam & Muldner, 2017). There is also increasing evidence that engagement changes depending on teachers' instructional strategies and classroom settings (Bae & Lai, 2020; Kyndt et al., 2013; Patall et al., 2019; Sun et al., 2022). Accordingly, relatively recent studies have investigated the learning effects of various instructional interventions, including prediscussion activities, such as flipped lectures or self-study (e.g., Lee & Choi, 2019; Lim et al., 2019).

Furthermore, individual preparation for future collaboration suggests that collaboration with peers require combination with prior learning activities (Lam & Kapur, 2018; Lam & Muldner, 2017). Lam and Kapur (2018), for instance, proposed another type of combination in which the same activity was first prepared alone and then worked in pairs during the rest of the time (preparation). College students' performance of this combination was compared with that of having pairs work together from the beginning (non-preparation). The two tasks used were to summarize the content of the learning material, and to generate new ideas or reasoning based on the learning material. The final performance was then compared in a transfer task. They observed that the preparation group scored higher than the non-preparation group but found no differences based on the type of tasks. This shows that studying alone first and then collaborating with others is more effective than collaborating from the beginning. The result also suggests that different learning activities can be combined to subsequent discussions for students' learning.

In the present study, we widened the lens for the classroom study of engagement during discussion. The present research investigated students' learning in classroom discussions, moderated by different combinations of prediscussion activities and discussions. We compared undergraduate students' learning gain as seen in student-centered small group discussions with in teacher-led discussions when students had previously displayed different prediscussion learning activities, lectures or self-study. We analyzed overt cognitive-behavioral engagement during discussions to elucidate the students' learning.

The following sections (i) describe the features of students' engagement during the learning process, (ii) discuss the prediscussion activity, (iii) compare differences in students' learning and engagement between student- and teacher-led discussions, and (iv) provide an overview of the study.

Engagement during the learning process

Engagement has been classified into four dimensions (e.g., Lombardi et al., 2021). First, the social-behavior aspect implies that students participate actively in academic activities with their peers (Linnenbrink-Garcia et al., 2011). Second, the cognitive aspect signifies that students partake in effortful thinking and invest mental energy in the learning process (Fredricks et al., 2004). Third, the emotional aspect denotes that positive feelings are generated during the performance of academic tasks, such as the recognition of the value of the tasks (Furrer & Skinner, 2003). Fourth, the agentic aspect symbolizes the inculcation of autonomy in students, for instance, their self-identification as knowledge constructors, individually or as part of a peer group (Chi & Wylie, 2014; Patall et al., 2019; Reeve & Tseng, 2011). A study by Patall et al. (2019) has, particularly, revealed that high levels of social-behavioral, cognitive, and emotional engagement lead to students' improved academic achievement and deepen their agentic engagement.

The multifaced concept of engagement, however, can be difficult to distinguish among the dimensions, such as between cognitive and behavioral engagement (Eccles, 2016). Thus, we adapted extended conceptualization by Pekrun and Linnenbrink-Garcia (2012). Pekrun and Linnenbrink-Garcia postulated cognitive-behavioral engagement as an integrated form of engagement during learning activities. This includes students' self-regulated learning and (meta) cognitive strategies in the learning process. We also refer to the interactive constructive active passive (ICAP) framework, which proposes that engagement can be detected and differentiated through a detailed analysis of students' overt behavior during learning (Chi & Wylie, 2014). The ICAP framework generally predicts that students who interact constructively with other individuals will learn better than those who learn alone. In other words, interaction can provide opportunities for students to explain their point of view, elicit responses from peers, and incorporate peers' contributions. Given these benefits of interaction, students' learning outcome was lowest in P mode and increased in ascending order in A, C, and I modes.

Student learning and engagement depending on prediscussion activity

A typical combination of instructional methods entails the instruction then problem-solving (I-PS) or problem-solving then instruction (PS-I) structure (e.g., Loibl et al., 2017; Schalk et al., 2018). I-PS signifies that teachers tell before students practice (i.e., T&P). Conversely, in PS-I, students are given problem-solving tasks and then provided with relevant instructions by teachers, for instance, invention (Schwartz & Bransford, 1998) or productive failure (Kapur, 2008).

Some previous studies have indicated the benefits of learning methods that involve active student engagement (e.g., Billings & Hasltead, 2019; Lam & Muldner, 2017). Lam and Muldner (2017) compared four conditions of collaborative learning activity among college students: no preparation-active, no preparation-constructive, preparation-active, and preparation-constructive. The two preparation conditions were separated by different degrees of student engagement by the ICAP framework (Chi, 2009; Chi & Wylie, 2014). The active condition required student preparation in the form of imbibing canonical concepts before applying them to specific tasks. In the constructive condition, students produced ideas in the preparatory phase before comparing their concepts and finally generating ideas collaboratively. The results showed that students who prepared first by working through instructional materials individually and then collaboratively learned more than those who did not prepare. Although the type of preparation did not have a significant effect, Lam and Muldner (2017) suggest that a cognitively engaged individual preparations lead to better learning as it encourages students to collaborate constructively, even in types of tasks that do not induce such engagement.

These studies showed that different learning activities can be combined to subsequent discussions for students' learning. Besides enhancing student learning, students must be provided with time on their own to perform activities that help them process the learning material before participating in the collaboration exercise (cf. Lim et al., 2022, 2023; Tsovaltzi et al., 2015) since students do not often spontaneously engage in interactive collaborative activities such as problem-based collaborative learning or classroom discussions (Jeong &

Hmelo-Silver, 2016; Menekse & Chi, 2019). Therefore, there is an urgent need to provide instructors with adequate preliminary activities for subsequent discussions. Among the numerous activities that can be selected, we focused on self-study and lectures, which are different and widespread learning activities in higher education. Furthermore, self-study and lectures are complementary activities for subsequent discussions (Akçayır & Akçayır, 2018; Schmidt et al., 2010).

Here, we defined self-study as activities one performs by oneself before learning from others. Self-study refers to a wide range of activities including planning, learning, and evaluating learning outcomes on one's own (e.g., Zimmerman & Schunk, 2008). A key aspect of self-study is the students' ability to direct their own learning. They can regulate their learning process as an inherently constructive and self-directed process. Students, in this process, can use their prior knowledge and integrate new knowledge in which self-regulated learning can be successfully acquired, including the use of metacognitive skills such as planning, executing, evaluating, correcting etc (cf. Boekaerts, 1999). Thus, self-study can entail constructive cognitive engagement (Chi, 2009; Chi & Wylie, 2014).

On the other hand, conventional lectures represent widely used instructional methods in higher education (Deslauriers et al., 2018; Stains et al., 2018) described as passive cognitive engagement. However, not only can lectures followed by students' analysis be effective, as reported by Schwartz and Bransford (1998), but Kapur (2008) reported that lectures in the form of solutions to problems can be equally effective. Indeed, some studies suggested that lectures are more effective when combined with complementary learning activities (e.g., Chi & Wylie, 2014). Therefore, in line with previous results that these two activities can lead to significant differences in students' learning outcomes (cf. Chi & Wylie, 2014; Lam & Muldner, 2017), we expected that students would be more engaged in discussions after self-study rather than lectures.

Student learning and engagement in classroom discussions

A vast body of literature exists on the effects of various discussion approaches on student learning and has sufficiently yielded reviews and meta-analyses (e.g., Almasi & Garas-York, 2009; Garas-York & Almasi, 2017; Murphy et al., 2009, 2018). Murphy et al. (2009) quantitatively reviewed research on classroom discussions to determine their effects on comprehension. Their investigation included concepts such as collaborative reasoning (Waggoner et al., 1995), philosophy for children (Sharp, 1995), questioning the author (Beck et al., 1996), book club (Raphael & McMahon, 1994), and so on. This comprehensive meta-analysis of empirical studies examined evidence of how teachers and students talk during discussions affected individual student comprehension as well as critical thinking and reasoning outcomes. Their examination revealed that several discussion approaches strongly increased the amount of student talk and concomitantly reduced teacher talk, followed by a substantial enhancement in text comprehension.

In addition, among the student-centered small group discussion approaches, collaborative reasoning produced greater motivation (i.e., interest) and engagement in children than fully teacher-directed discussions (Wu et al., 2013). Wu and colleagues examined whether children engagement observed during a discussion sustained positive correlations with self-reported student interest and engagement several months later. The major finding is that student-centered discussions produce enhanced motivation, increased discussion value, and greater engagement.

Unlike a teacher-led classroom discussion, where the teacher manages turn-taking and the topic, a student-centered discussion encourages students to take control over speaking turns and the subject matter. Here, students hold interpretive authority or autonomy to evaluate their own or their peers' ideas and reasons that are presented. For example, collaborative reasoning appears to implement the classroom practices that Guthrie and Humenick (2004) have demonstrated to support student engagement. Noteworthy is that collaborative reasoning allows students' autonomy during discussions, entail peer collaboration and enhance students' engagement through stimulation from the fact that students are discussing controversial issues.

Meanwhile, some researchers have identified that teacher-led or -supported discussions (i.e., conventional classroom discussions) are an effective instructional strategy in improving student learning (e.g., Kirch & Siry, 2012; Lombardi et al., 2018). In classroom discussion settings, teachers ask more open-ended questions, employing queries and probes designed to help students understand and think deeply (e.g., Liang & Dole, 2006; Manz & Renga, 2017; Wortham, 2006). Meyer (1993) considered these instructional scaffoldings as temporary teacher supports extended to the student in the zone of proximal development (ZPD, Vygotsky, 1978).

Davin (2013) examined whether teacher-directed class discussions enhance students' engagement in learning. In particular, it was examined how a primary school teacher utilized instructional conversation within an L2 classroom. Instructional conversations were used to co-construct a group ZPD in response to less predictable student errors or inquiries. The flexible mediation offered by teachers enabled more active engagement and increased responsive conversations among students. These teacher-directed frameworks can make classroom discussion more accessible to students, in alignment with Morocco and Hindin's (2002) findings. Hence, Bae et al. (2021) recommended that teachers provide explicit support with students to enhance student engagement in classroom discourse. Simultaneously, they posited that teachers should encourage students' agency and participation in discourse through meaningful leadership positioning and leveraging students' diverse strengths.

To summarize, student-led discussions promote students' engagement and learning gains based on student autonomy, whereas teacher-led discussions encourage students to participate in conversations and work on their strengths under the teacher's control and support. Therefore, student-led discussions, which are based on autonomy and engagement, when combined with self-study, are expected to increase students' performance, whereas teacher-led discussions are rather improved since learning and scaffolded engagement through lectures presupposes teachers' assistance to students.

3.2.2 The present study

The present study was designed to compare undergraduate students' engagement and learning gains attained through student-centered small group discussions against those acquired through teacher-led discussions when self-study or lecture was preceded. Regardless of the discussion types (formats), we predicted that self-study, as a prediscussion activity, would lead to higher engagement and outcomes for students than lectures. We also expected that the engagement and learning outcomes of the lecture group would improve through teacher-led discussions.

To substantiate our predictions, we further explored the relationship between enhanced engagement and improved outcomes in combined prediscussion activities and discussions. The current study sought to determine the influence of overt cognitive-behavioral engagement on students' learning, utilizing protocols established by Lutz et al. (2006). We did not determine engagement through direct cues, such as biological signals (e.g., Imai et al., 1992; Rosenberg & Ekman, 1994); rather, we rated students' moment-by moment engagement using successive segments of videotaped discussions. Student engagement during discussions was scrutinized to enable researchers to offer detailed evidence on this issue. The rubrics for engagement are shown in **Appendix M**. Subsequently, we examined the effects of such engagement in discussions on student learning through moderated mediation analysis.

Method

Participants and design

The study was approved by the Institutional Review Board of Seoul National University. This study was conducted in the second semester of 2021. It was conducted in an in-person regular class. We conducted an experimental study involving four introductory classes by four different instructors at a university in South Korea. All participants were Asian. All participants (N = 305; female student = 116) were recruited from an undergraduate course. The average age of participants is 19.35 years (aged 18 – 21 years). The majors of the students were diverse, including economics, geography and psychology. We excluded 47 participants who did not participate in the discussion from the outset, did not interact with other group members, and whose engagement could not be examined due to an inferior video recording quality. The study used a 2 x 2 experimental design examining the type of prediscussion activity (lecture or self-study) and discussion formats (student-led or teacher-led). Participants were then randomly assigned to four groups: student-led discussion after a lecture (LSD group, $n = 64$), teacher-led discussion after a lecture (LTD group, $n = 73$), student-led discussion after self-study (SSD group, $n = 63$), and teacher-led discussion after self-study (STD group, $n = 58$).

Procedure

All students were first apprised that they could either agree or disagree with the experiment and the video recording. Students were also explained that they could quit during the experiment and that test scores in the experiment would not affect their grades in this class; students who agreed, signed a consent form. After that, students were randomly assigned to different groups that worked in separate classrooms, and all experiments occurred simultaneously.

In the classroom, the experiment procedure was explained to the students in detail. All students first completed a background questionnaire. The background knowledge questionnaire used a 7-point Likert scale, ranging from 1 (no knowledge) to 7 (expert knowledge), for six subjects. The questionnaire included two subjects relevant to this study and four subjects irrelevant to this study, such as criminal procedure code. (1) The LSD group watched the 20-minute video lecture and engaged in discussion for 15 minutes in groups of three or four students. The student-led discussion lasted 15 minutes and the instructor did not intervene. For example, question and answer were not allowed. (2) The LTD group watched the 20-minute video lecture and discussed for 15 minutes in groups of three or four students. Students were free to ask questions to the instructor. The instructor moved around the classroom and observed the students' discussions. (3) Students assigned to the SSD group studied by themselves the texts for 20 minutes and thereafter engaged in discussion for 15 minutes in groups of three or four students. During self-study, students were asked to study on their own with written text. Students were free to perform activities such as underlining written text or taking notes. As with the LSD group, the students in the SSD groups discussed with their peers using only written text, and were not subjected to scaffolding such as asking questions to the instructor. (4) Finally, the students in the STD group studied on their own for 20 minutes and discussed for 15 minutes in groups of three or four students. As with the LTD group, students were able to freely interact with the instructor while discussing.

All discussions (whether student-led or teacher-led) were accomplished on camera. Discussions were recorded, and after the discussions were completed, all students in different four groups took a 20-minute same final examination on the learning content.

Teacher participant

The student-led discussion groups (the LSD and SSD groups) were not provided with any support or scaffolding from the instructor. Overall, the student-led discussion group was considered decentralized to the extent that students' interaction was encouraged. The instructor began with briefing about discussion session and the final test that followed. Thus, the instructors of the LSD and SSD classrooms were faithful to the experimentation, as were the experimenters.

The major distinction between two conditions was the manner in which the discussion was held. The instructor in the teacher-led classroom (the LTD and STD groups) encouraged the students to refer to their questions during the discussion. With reference to van de Pol et al.'s (2014) levels of teacher degree of control in the classroom, the instructors in the LTD and STD classrooms gave students low or medium control and support. The teacher-led discussion group was considered centralized in which meaning is constructed when students ask questions to the teacher about the content and teacher respond to their questions, or teacher asks questions to students and solicit and evaluate responses from students. Thus, the instructors did not provide new content, but they elicited a short response or an elaborate response, such as elaboration and explanation of why something was in a certain way. They also asked a more detailed or open question to students. For example, an Instructor 1 said “Well, I think, why don’t you reconsider the types of covalent bond?” to a student.

However, there was no difference in support such as additional learning materials, between the two groups.

Instructional material

The present study involved the following instructional materials: (i) monologue-style lecture and written text, and (ii) final tests.

Instructional lecture. We used an actual video lecture comprising a 20-minute long monologue-style lecture that delivered life science -related learning content, specifically on the covalent bond. Video lectures were conducted on one screen in different classrooms for each group (the LSD, and LTD groups). The video lecture featured a male instructor who was not assigned to the classroom. The video of the LSD and LTD groups was the same.

The six-page long written text for the class were derived from the contents of the video lecture and encompassed information on the same topic. All four groups were given the same instructional written text.

Final test items. The final test items used real case scenarios and accomplished a quantifiable assessment of the extent to which transfer of knowledge had occurred. The final test comprised a total of 11 questions worth 33 points. All the questions asked on the final test questions comprised transfer-type items, and every question required students to display their high-level of comprehension of the content such as application and transfer. We intended students to apply learning that surpassed the delivered content. Thus, students were required to deploy the acquired conceptual knowledge to extended situations that transcended the original context. The test items included short-answer, multiple-choice, and open-ended descriptive questions. All the instructional materials are shown in **Appendix N**.

First author and two raters scored 20% of short-answer and open-ended answer sheets to reduce subjectivity, and the measured interrater agreement analysis [ICC (3, k)] was .84. Accordingly, the remaining answer sheets were scored by the first author.

Data analysis

Data were analyzed using descriptive statistics [mean (M), standard deviation (SD), and standard error (SE)]. Analysis of variance (ANOVA) was performed between the four different groups to examine differences in learning outcome measured by transfer items, and p-values of $< .05$ were deemed to indicate statistically significant differences.

We adapted the ICAP framework on distinguishing levels of engagement through close observation of classroom videos (Chi & Wylie, 2014). However, the data sets that we analyzed sometimes lacked information to differentiate between Interactive, Constructive, and Active modes, and it seemed prudent not to force a distinction. A moment-by-moment real-time measure was then employed by three independent raters to allow a more direct and process-oriented analysis of engagement in discussions (see Lutz et al., 2006; Wu et al., 2013). The three raters watched the 15-minute discussion videos and evaluated an aggregate of 30 successive 30-second segments of the video on the same 5-point Likert-like scale ranging from 1 (*not at all engaged*) to 5 (*very much engaged*) (see **Appendix M**).

The group-related differences in engagement across the discussion times were examined using a linear mixed-effect model (LMM) analysis through the nlme package in R (Version 4.1.2). The discussion times, groups, and the interactions between these two elements were included in our statistical model as the fixed effects; gender and age were set as covariates. The regression coefficients of the interaction terms indicated the group effects on the engagement across all discussion times. The random effect structure encompassed a random intercept for the participant and a random time slope for the participant. To test our model, a quadratic term for the discussion time was included after inspecting visually raw data at the individual level and identifying the model yielding the lowest Akaike information criterion and Bayesian information criterion among different curve-fitting regression models (linear, quadratic, and cubic), which predicted the dependent variable (i.e., engagement) through a single independent variable (i.e., discussion time). A first-order autoregressive or AR (1) covariance structure was incorporated into the model to account for individual autocorrelation.

For further analysis of the causality between engagement and learning, the moderated mediation analysis was conducted using the PROCESS macro for SPSS (Model 7) (Hayes, 2017). A percentile bootstrap estimation with 5,000 replications was used to generate a 95% confidence interval (CI), in which the effects are estimated significant if zero is not included between lower limit and upper limit for the CI. Demographic variables (i.e., age and gender) were included as covariates in this model.

Table 3.8
Descriptive statistics for ANOVA

Leaning gain	Mean (*SD*)	*F*	*p*	η_p^2
Condtion 1:				
Prediscussion activity		88.57	<.001	0.26
Lecture	18.73 (*3.55*)			
Self-study	22.61 (*3.82*)			
Condtion 2:				
Discussion format		5.80	.017	0.02
Teacher-led	21.15 (*3.38*)			
Student-led	20.14 (*4.41*)			
Condition 1 × 2		13.40	<.001	0.05

Data represent values of means (and *SD*) in transfer-type items for main effects and interaction effect.

Result

Students learning

ANOVA was performed to examine the differences between the four experimental groups in accordance with our predictions. The gender and age yielded no significant differences in scores. In addition, the level of self-reported background knowledge for the target learning content was "almost having no knowledge" ($M = 2.21$, $SD = 1.40$), and there was no significant difference among groups in the background knowledge questionnaire, $p = .571$.

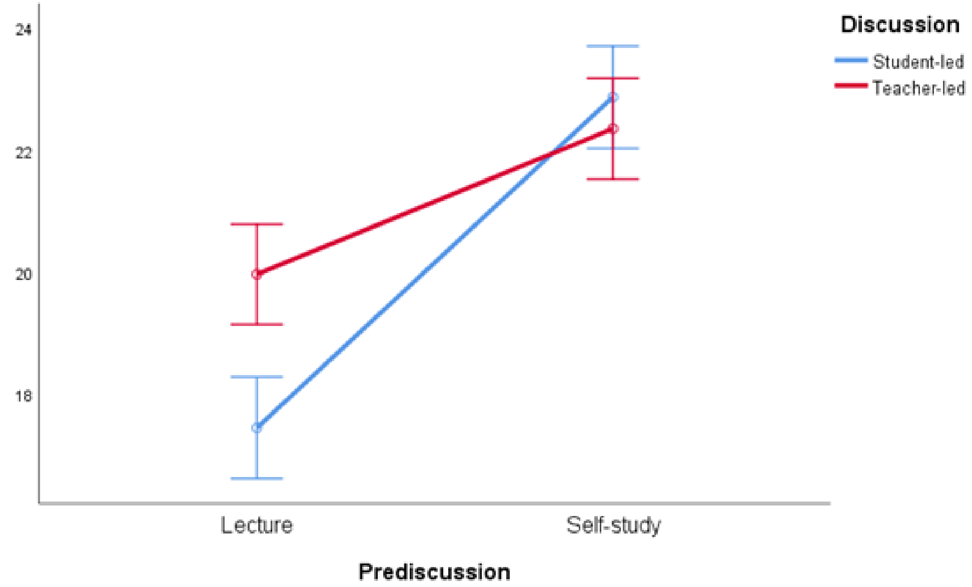
Groups were used as independent variables, and final test scores (i.e., transfer-type items) were used as dependent variables. The main effect of the prediscussion activity condition, lecture, or self-study was significant. The main effect of the discussion condition, teacher-led discussions, or student-led discussions revealed significant difference. The interaction effect of these two factors was significant in terms of gains by students (**Table 3.8** and **Figure 3.1**).

In particular, we examined the differences between all four groups. **Table 3.9** presents the original means and standard deviations of the final test scores for transfer-type items. The SSD group achieved significantly higher scores in transfer-type items than the LTD group, and STD group, followed by the LSD group, $ps < .05$. The LTD group registered significantly higher scores than the LSD group, $p < .001$. Further, the STD group had significantly higher scores than the LSD group, $p < .001$, but not the LTD group, $p = 1.000$.

Engagement during discussions

The agreement in engagement ratings registered by a set of raters (Raters 1, 2, and 3) vis-à-vis the video-recorded discussions was calculated as pooled across individuals. Of the total videos, 30% were randomly selected and were assessed by three raters, all of whom were independent of this study. According to the

Figure 3.1
Main effects and interaction effect for the students' learning



Note. L = Lecture group; S = Self-study group; LR = Lecture Review group; LD = Lecture Discussion group; SD = Self-study Discussion group. All types of items consist of verbatim, paraphrased, and transfer items. Error bars are $\pm 2SE$. The top of the scale reflects the possible highest score on that scale. .

Table 3.9
Descriptive statistics for four experimental groups

Groups ($N = 258$)	**Mean**	***SD***
LSD ($n = 64$)	17.45	3.65
LTD ($n = 73$)	21.19	3.37
SSD ($n = 63$)	22.87	3.30
STD ($n = 58$)	21.10	3.43

Data represent values of means and SDs of transfer-type items and are detailed for four groups. LSD, lecture and student-led discussion; LTD, lecture and teacher-led discussion; SSD, self-study and student-led discussion; and STD, self-study and teacher-led discussion.

rubric, each rater assessed individual students in the discussion groups every successive 30 seconds. Interrater agreement among the three raters was adequate [ICC (3, k) = .97, $p < .001$]. Thus, the remaining videos were assessed by Rater 1.

We estimated the internal consistency of engagement ratings by correlating the scores of the odd and even segments in the 15-minute video before applying the Spearman–Brown prophecy formula. The internal consistency reliability was calculated at .91. LMM was performed to identify the group effects on engagement across discussion times after controlling the covariates (**Table 3.10**). No significant difference was noted in students' engagement according to gender and age. The main effect of the discussion group was significant: the LTD, SSD, and STD groups were more engaged in discussions than the LSD group. The linear term for discussion time was significant and positive, suggesting constantly increasing engagement. However, the quadratic term was negative, indicating that this increase was not sustained and that the engagement decreased over time. These results imply an inverted U-shaped relationship between engagement and discussion time. The interaction effects between the groups and discussion times were also significant for the LTD group's linear and quadratic terms, the SSD group's linear term and quadratic terms, and the STD group's linear and quadratic terms.

These significant interactions were specifically identified using a Bonferroni-corrected post hoc test (**Figure 3.2** and **Appendix O**). The SSD group displayed the best engagement for almost the entire time (4 and a half minutes to 15 minutes) even though this group exhibited lesser engagement for the first 4 minutes than the STD group, $ps < .001$. The STD group engaged more actively at the beginning of the discussion than other groups. However, this trend was reversed: the engagement of the STD group declined steeply from the second

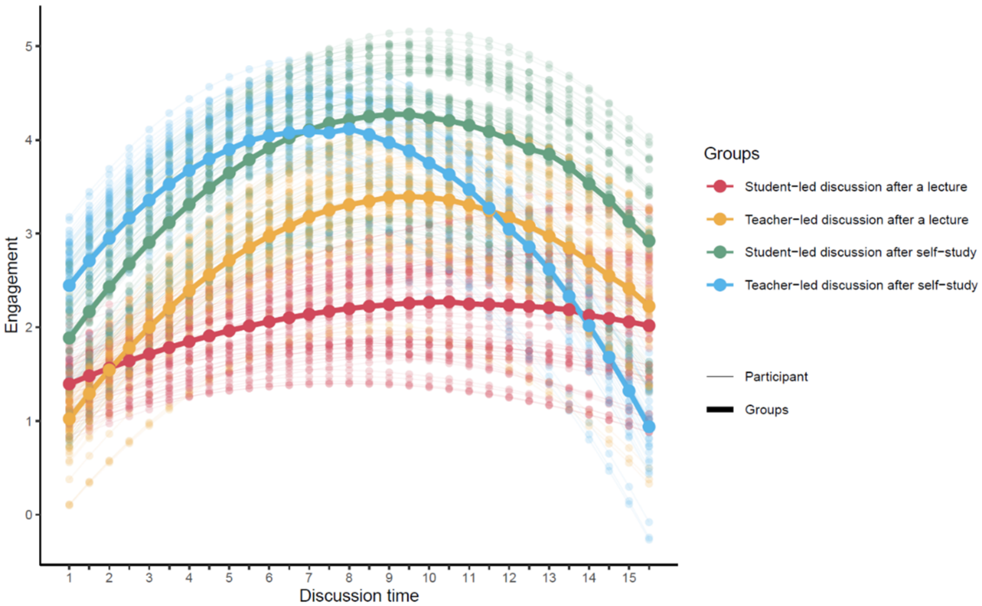
Table 3.10
LMM analysis of the group-based differences in engagement across discussion times

Engagement	*b*	*SE*	*df*	*t*	*p*
(Intercept)	1.71	0.80	6839.00	2.13	.033
Age	0.03	0.04	252.00	0.66	.511
Gender	-0.01	0.07	252.00	-0.17	.862
LSD			Reference		
LTD	1.10	0.12	252.00	9.36	<.001
SSD	2.04	0.12	252.00	16.81	<.001
STD	1.78	0.13	252.00	14.10	<.001
Time	0.02	0.01	6839.00	5.05	<.001
$Time^2$	0.00	0.00	6839.00	-5.47	<.001
LSD × Time			Reference		
LTD × Time	0.02	0.01	6839.00	2.41	.016
SSD × Time	0.01	0.01	6839.00	2.02	.044
STD × Time	-0.09	0.01	6839.00	-11.84	<.001
LSD × $Time^2$			Reference		
LTD × $Time^2$	-0.01	0.00	6839.00	-10.46	<.001
SSD × $Time^2$	-0.01	0.00	6839.00	-10.82	<.001
STD × $Time^2$	-0.01	0.00	6839.00	-14.89	<.001

Note. Time, linear term for discussion time; $Time^2$, quadratic term for discussion time; LSD, lecture and student-led discussion; LTD, lecture and teacher-led discussion; SSD, self-study and student-led discussion; and STD, self-study and teacher-led discussion.

half of the discussion. This decline was more pronounced than the LSD (for the last 1 and a half minutes) and LTD (after 12 and a half minutes) groups, $ps < .001$. The LTD group was less engaged in the discussion than the LSD group for 2 and a half minutes after the beginning but became more engaged for 12 and a half minutes leading up to the end of the discussion, $ps < .001$. Finally, the LSD group demonstrated the least engagement for almost the entire time (3 to 13 and a half minutes), $ps < .001$ (**Appendix O**).

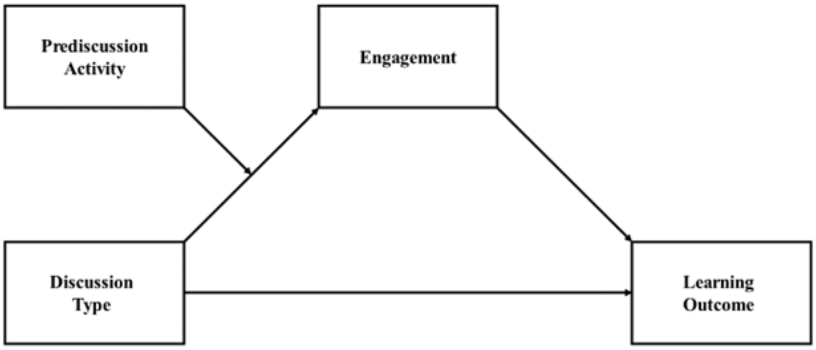
Figure 3.2
Differences in the engagement of four experimental groups across the discussion times.



Engagement leading to students' learning outcomes

The above analyses indicate that students' learning may be influenced by engagement during discussions. In addition, the engagement induced by different combinations of prediscussion activities and types of discussion could interact with student learning. To substantiate this interpretation, we subsequently analyzed the relationship between student engagement and outcome by conducting the moderated mediation analysis (**Figure 3.3**).

Figure 3.3
Moderated mediation model for examining relations among prediscussion activity, discussion type, engagement, and students' learning outcome.



As we predicted, student engagement was modulated by different combinations of prediscussion activities and discussion types, which had impacts on students' learning. First of all, the explained variance (R^2) of the model predicting the engagement was 0.48, $[F(5, 252) = 45.85,\ p < .001]$. In this model, the main effects of type of discussion ($b = -0.62$, $p < .001$, 95% CI [0.83, 0.41]) and prediscussion activities ($b = 0.60$, $p < .001$, 95% CI [0.38, 0.82]) were significant. Moreover, the interaction effect between type of prediscussion activities and type of discussions was significantly associated with the engagement ($b = 0.95$, $p < .001$, 95% CI [0.64, 1.26]), and the incremental variance was 0.08 in the engagement: $\Delta R^2 = .08$, $[F(1, 252) = 36.33,\ p < .001]$. Thus, student-led

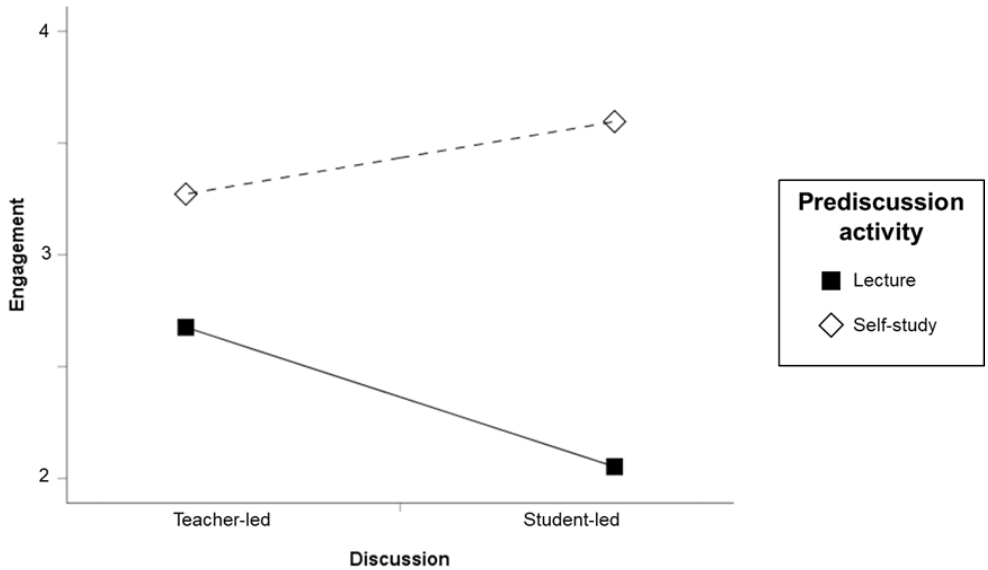
Table 3.11
Conditional effect of varied types of discussions on engagement

Prediscussion activity	b	t	p	LLCI	ULCI
Lecture	-0.62	-5.84	<.001	-0.83	-0.41
Self-study	0.32	2.84	.005	0.10	0.55

Note. b, unstandardized coefficient; LLCI, 95% lower limit confidence interval; ULCI, 95% upper limit confidence interval.

discussion rather than teacher-led discussion was significantly related to the decrease in engagement in the lecture conditions, but increased engagement in the self-study condition (**Table 3.11** and **Figure 3.4**).

Figure 3.4
Association of discussion and engagement moderated by prediscussion activity.



Regarding the overall model predicting learning outcome through the engagement [R^2 =.53, $F(4, 253) = 72.19$, $p < .001$], as shown in **Table 3.12**,

Table 3.12
Conditional indirect effect of varied types of discussions on learning via engagement

Prediscussion activity	Effect	SE_boot	LLCI	ULCI
Lecture	-2.06	0.36	-2.78	-1.37
Self-study	1.07	0.40	0.26	1.88
Index of moderated mediation	3.13	0.56	2.05	4.26

Note. LLCI, 95% lower limit confidence interval; ULCI, 95% upper limit confidence interval.

the index of moderated mediation (Hayes, 2017) of prediscussion activities was significant. More specifically, this result suggests that student-led discussion, relative to teacher-led discussion, can reduce the learning effect of students by decreasing engagement when the prediscussion activity is a lecture, but it can enhance the learning effect by increasing the engagement when the activity is self-study.

3.2.3 Discussion

In this study, we examined a two-part instructional sequence of prediscussion activity (lecture or self-study) and discussions (teacher-led or student-led) in higher education. We also investigated whether interaction of prediscussion activities and discussions would affect student engagement during subsequent discussions, leading to student learning outcome. The results revealed that students assigned to the SSD group recorded the highest scores, followed in descending order by the LTD, STD, and LSD groups. As Lam and Kapur (2018) assuming that generative preparation helped students learn from subsequent collaborative learning, the self-study group could benefit from greater opportunities for self-generated learning through novel learning materials.

We confirmed further that specific prediscussion activities could be appropriately paired with certain discussion formats: interestingly, the self-study groups scored higher for student-led discussions without the teacher's scaffolding, whereas the lecture groups registered higher scores when they engaged in teacher-led discussions that allowed the teacher's scaffolding. This result could be attributed to the students in lecture groups being mentally prepared for a teacher in mind when they listen to a lecture. Thus, the instructor's support could have functioned to structure learning for students who listened to lectures that were insufficient in helping them construct understanding on their own. Conversely, students who studied on their own built knowledge and comprehension in a personal manner, so they could share their understanding during the discussion and ask their peers to clarify uncertainties. The extant research has elucidated that classroom discussions to construct new knowledge do not occur naturally (McNamara et al., 1996; Murphy et al., 2016); thus, students could develop a prior fundamental understanding of the requisite materials through self-study and access this knowledge during discussions.

The instructor's support could exert a positive effect to bridge the gap with lecture groups during discussion. Conversely, the instructor's support could disrupt self-regulated or self-paced learning leading to discussions in self-study groups. Therefore, our results pertain more to the interactions between the different prediscussion activities that facilitate profound learning in students than to simple combinations of different activities.

We subsequently sought to determine the relationships between student engagement and learning outcome, focusing on how the interactions between individual prediscussion activities and type of discussion promoted outcomes. We videotaped student discussions and analyzed videotaped discussions using three independent methods. The results revealed that the students in all the discussion groups gradually became engaged as the discussion time elapsed, but their engagement decreased gradually at the end of the discussion period. A significant interaction was noted between the discussion group and the discussion time.

Specifically, student engagement in the LSD group increased gradually over time but no substantial change was observed between the beginning and the end of the discussion. This outcome could indicate that students used to the delivery of lecture content or passive absorption of lectures could have felt awkward in the interactions between learners and that they needed more time or support from teachers and other tutors. This construal can be confirmed through the results achieved by the LTD group. Unlike the LSD group, students in the LTD group could avail themselves of instructor scaffolding during their discussion and became significantly more engaged in their discussion. Thus, as with students' learning outcome, these results also indicate that students who listen to classroom lectures require discussions to be carefully scaffolded by instructors.

The two self-study and discussion groups (i.e., the SSD and STD groups) were more engaged from the beginning of the discussion period than the two lecture and discussion groups. Presumably, this result reflects that students could clearly distinguish between what they did or did not know when they studied on their own. In particular, students assigned to the SSD group were engaged throughout the discussion and displayed active overt cognitive-behavioral engagement. Students in the STD group were more engaged in the discussion than students in the SSD group at the beginning of the discussion. However, their engagement declined rather vertically over time. Even the STD group was less engaged than the LTD group during the first 13 minutes of the discussion and less engaged than the LSD group one minute before the end of the discussion.

In other words, the engagement of self-regulated learners diminished with the teacher's support during the discussion because they considered peer discussion as a means of accomplishing the work assigned by the teacher.

However, we are not arguing that autonomous and self-regulated students are negatively affected by their teachers' guidance. Indeed, students in the self-study group also became more engaged with their teacher's help at the beginning of the discussion. This means that teachers must intervene at an appropriate time and extent during classroom discussions according to their instructional approach and their students' characteristics.

Following this, we conducted the moderated mediation analysis to investigate the path that student engagement took according to the combination of prediscussion activity and type of discussion, which leads to student learning. In our model, type of prediscussion activity and type of discussion were significant, implying that the engagement increased in a self-study condition compared to the lecture condition and in a teacher-led discussion condition rather than in a student-led discussion condition. Furthermore, compared to teacher-led discussions, student-led discussions can reduce student engagement after lectures, but can increase student engagement after self-study activities to increase learning. Thus, student engagement during discussions exerted a significant effect on student learning.

Therefore, the present study offers practical implications on how teachers can use discussions in their classroom; prompting participants to engage in significant interactions during the discussion, and giving summative feedback afterwards are not easy for instructors. Thus, our study identifies how they can employ adequate preliminary activities timely and recommends appropriate discussion formats that could be deployed to encourage engagement and learning in their students.

Limitations and future directions

The current study acknowledges several limitations. This study seems to show that self-study can represent a superior instructional strategy to lectures. In practice, the results of the experiment disclosed superior student engagement and outcomes in the self-study groups than in the lecture groups. However, a dichotomy between self-study and lectures in this context could overlook the fact that lectures can encourage students to learn. The problem occurs when teachers continue to widely use only the conventional lecture format in academic disciplines such as science, technology, engineering, and mathematics (STEM) (e.g., Freeman et al., 2014; Schmidt et al., 2015; Stains et al., 2018; Yannier et al., 2021). Similarly, one-way didactic lectures have certain constraints: students cannot sustain their attention (e.g., Poh et al., 2010) and are not encouraged to think broadly or intensively (e.g., Bonawitz et al., 2011). Thus, many researchers suggest the application of active learning through or in combination with lectures (e.g., Chi, 2009; Chi & Wylie, 2014; Deslauriers et al., 2019).

Therefore, our study recommends that teachers and educational stakeholders should be cautious about using only lectures in their classrooms. Although they provide didactic lectures in their classes, teachers can trigger student engagement and performance by combining lectures with teacher-facilitated discussions that benefit from their support.

We did not assess the prior or contextual knowledge and engagement of the participating students. Thus, we could not determine whether our findings would be validated if previously existing variations were considered. However, the study did not intend to explore the degree to which each applied methodology influenced student engagement and students' learning. Therefore, future studies should evaluate the baselines of their participants before initiating the experimental intervention. Further, prospective investigations should test for the long-term effects of such interventions.

Moreover, it will also be interesting to see how the students are interviewed after the discussion. Analyzing student interviews and then deriving various findings, such as what is required for this design in a narrow sense or the form of discussion in a broad sense, can be helpful for learning science research.

In conclusion, there is no longer any doubt that classroom discussions represent an effective strategy in education. Universities have recognized for decades that an andragogical approach is appropriate for students undertaking higher education. The core tenets of higher education incorporate self-direction, self-motivation, and active learning (Knowles, 1975). Therefore, we hope that educators will utilize more varied discussion formats combined with their own instructional strategies in educational settings to make their students lifelong learners.

Conclusion

A series of experiments of this thesis began as an effort to explore alternatives to traditional lectures or PBL. The proposed method was to promote student participation by combining individual effort and cooperation. Specifically, the effectiveness of the self-study and discussion combination was examined.

Chapter 1 investigated the learning effect of discussions and the advantage of the combining self-study with discussion. Three experiments were conducted using law material, which is an unfamiliar subject for undergraduate students. The result revealed that self-study and discussion group obtained the highest scores, followed by the lecture and discussion group, then the lecture and review group. However, there was no difference between the self-study and lecture groups in the level of learning before holding a discussion. Considering the results of previous research (e.g., Chi & Wylie, 2014), it is not surprising that discussion activity is superior to reviewing learning materials. However, while there was no difference between the self-study and lecture groups, why was there a difference in the subsequent discussion? My speculation is that, compared to listening to a lecture, self-study can provide students with relatively more time for reflection. In other words, students who listened to the lecture would have progressed their learning at the pace set by the instructor and felt they had completed their studies. On the other hand, self-study provided an opportunity for students to distinguish what they knew and did not know, and to raise questions, which led to active and productive discussions.

In Chapter 2, three experiments were conducted with different academic domains and age groups to generalize the benefits of combining self-study and discussion. First, students from various majors were introduced to topics in introductory biology. The result showed that the discussion groups had higher scores than the review groups, even when the review group were given additional materials to study. Among the discussion groups, the self-study group was once again superior to the lecture group in the final test score. Next, science and engineering students studied concepts in physics. Between the two discussion groups, the self-study group obtained significantly higher scores than the lecture group. Moreover, the self-study and discussion group was superior to the lecture and discussion group in terms of the amount of learning gains. Analyses of discussions revealed that the self-study group elicited more constructive and productive content than the lecture group. The third experiment was conducted with high school students. The result showed that the self-study and discussion group outperformed the lecture and discussion group in both tests taken immediately after learning or one week later. In short, the self-study and discussion group showed the highest performance compared to other groups regardless of age or academic subject.

In Chapter 3, in order to examine the need for teacher intervention in the discussion, two discussion formats of student-led and teacher-led discussions were crossed with two individual preparation activities, lecture and self-study. Analysis of the posttest scores revealed that the self-study and student-led discussion group obtained the highest score, followed by self-study and teacher-led discussion group, lecture and teacher-led discussion group, then lecture and student-led discussion group. Analysis of the discussion showed that the self-study and discussion group elicited more constructive and productive content than other groups. In addition, when the levels of cognitive-behavioral engagement were compared, it was found the self-study groups engaged more in the discussion than the lecture groups. Thus, these results provides another evidence for encouraging discussions among students after self-study.

This thesis has two major contributions to learning science. One is that can be applied in class. The most important finding is that the combination of self-study and discussion enhance learning effectiveness. This finding is highly practical in that it can be applied directly to classes as an effective alternative to lectures. Even if this combination is not used as it is, the finding that discussion enhances learning and that discussion does not necessarily require teacher intervention is also worth using in practice. However, I should hasten to add that we need more follow-up studies to expand the scope of its application. For example, it is necessary to explore whether this method is applicable to middle school students or upper elementary school students.

Another contribution is that it sets a new direction for research in the learning sciences. Despite the seminal work of Schwartz and Bransford (1998) showing that some combinations of activities are superior to others, research on combining two different learning activities has not been actively conducted, considering its importance. Although my focus was on the self-study and discussions, there is room to explore the effects of other combinations, such as the combination of writing and discussion, or even the triple combination of self-study, writing, and discussion. I hope that these subsequent studies will shed a new light on better learning and education.

References

Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education*, *126*, 334–345. https://doi.org/10.1016/j.compedu.2018.07.021

Almasi, J. F. (1995). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussions of literature. *Reading Research Quarterly*, 314–351. https://doi.org/10.2307/747620

Almasi, J. F., & Garas-York, K. (2009). Comprehension and discussion of text. *Handbook of Research on Reading Comprehension*, 470–493. https://doi.org/10.4324/9781315759609-36

Andrews, J. J., & Rapp, D. N. (2015). Benefits, costs, and challenges of collaboration for learning and memory. *Translational Issues in Psychological Science*, *1*(2), 182. https://doi.org/10.1037/tps0000025

Asterhan, C. S., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cognitive Science*, *33*(3), 374–400. https://doi.org/10.1111/j.1551-6709.2009.01017.x

Bae, C. L., & DeBusk-Lane, M. (2018). Motivation belief profiles in science: Links to classroom goal structures and achievement. *Learning and Individual Differences*, *67*, 91–104. https://doi.org/10.1016/j.lindif.2018.08.003

Bae, C. L., & Lai, M. H. (2020). Opportunities to participate in science learning and student engagement: A mixed methods approach to examining person and context factors. *Journal of Educational Psychology*, *112*(6), 1128. https://doi.org/10.1037/edu0000410

Bae, C. L., Mills, D. C., Zhang, F., Sealy, M., Cabrera, L., & Sea, M. (2021). A systematic review of science discourse in k–12 urban classrooms in the united states: Accounting for individual, collective, and contextual factors. *Review of Educational Research*, *91*(6), 831–877. https://doi.org/10.3102/00346543211042415

Barrows, H. S., Tamblyn, R. M., et al. (1980). *Problem-based learning: An approach to medical education* (Vol. 1). Springer Publishing Company.

Beck, I. L., McKeown, M. G., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the author: A yearlong classroom implementation to engage students with text. *The Elementary School Journal*, *96*(4), 385–414. https://doi.org/10.1086/461835

Beyer, B. K. (1987). *Practical strategies for the teaching of thinking.* Allyn and Bacon. https://doi.org/10.1177/019263658807250083

Bielaczyc, K., & Kapur, M. (2010). Playing epistemic games in science and mathematics classrooms. *Educational Technology*, 19–25.

Bielaczyc, K., Kapur, M., & Collins, A. (2013). Cultivating a community of learners in k–12 classrooms. *International Handbook of Collaborative Learning*, 233–249. https://doi.org/10.4324/9780203837290.ch13

Billings, D. M., & Halstead, J. A. (2019). *Teaching in nursing e-book: A guide for faculty.* Elsevier Health Sciences.

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review*, *30*, 703–725. https://doi.org/10.1007/s10648-018-9434-x

Bligh, D. A. (1998). *What's the use of lectures?* Intellect books.

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *Interna-*

tional Journal of Educational Research, *31*(6), 445–457. https://doi.org/10.1016/s0883-0355(99)00014-2

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.00

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom*. 1991 ASHE-ERIC Higher Education Reports. ERIC Clearinghouse on Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183.

Boud, D., & Feletti, G. (1998). *The challenge of problem-based learning*. Psychology Press. https://doi.org/10.4324/9781315042039

Bowman, M. N. (2021). *Topics in science education*. Nova Science Pub Inc.

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*(1), 61–100. https://doi.org/10.3102/0091732x024001061

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, *20*, 1350–1356. https://doi.org/10.3758/s13423-013-0442-z

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, *6*(3), 271–315. https://doi.org/10.1207/s15327809jls0603_1

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, M. T., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M.,

. . . Yaghmourian, D. L. (2018). Translating the icap theory of cognitive engagement into practice. *Cognitive Science*, *42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chi, M. T., Kang, S., & Yaghmourian, D. L. (2017). Why students learn more from dialogue-than monologue-videos: Analyses of peer interactions. *Journal of the Learning Sciences*, *26*(1), 10–50. https://doi.org/10.1080/10508406.2016.1204546

Chi, M. T., Kristensen, A. K., & Roscoe, R. (2012). Misunderstanding emergent causal mechanism in natural selection. *Evolution Challenges: Integrating Research and Practice in Teaching and Learning about Evolution*, 145–173. https://doi.org/10.1093/acprof:oso/9780199730421.003.0007

Chi, M. T., Roy, M., & Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, *32*(2), 301–341. https://doi.org/10.1080/03640210701863396

Chi, M. T., & Wylie, R. (2014). The ICAP Framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chin, C., & Osborne, J. (2010). Students' questions and discursive interaction: Their impact on argumentation during collaborative group discussions in science. *Journal of Research in Science Teaching*, *47*(7), 883–908. https://doi.org/10.1002/tea.20385

Cho, Y. H., Caleon, I. S., & Kapur, M. (2015). *Authentic problem solving and learning in the 21st century: Perspectives from singapore and beyond.* Springer. https://doi.org/10.1007/978-981-287-521-1_1

Cohen, A. D. (1994). *Assessing language ability in the classroom.* Boston, Ma: Heinle and Heinle.

Craig, S. D., Chi, M. T., & VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of Educational Psychology*, *101*(4), 779. https://doi.org/10.1037/a0016601

Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level-reasoning-question effect: The role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction*, *24*(4), 565–591. https://doi.org/10.1207/s1532690xci2404_4

Davin, K. J. (2013). Integration of dynamic assessment and instructional conversations to promote development and improve assessment in the language classroom. *Language Teaching Research*, *17*(3), 303–322. https://doi.org/10.1177/1362168813482934

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, *116*(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

Dunlosky, J., & Rawson, K. A. (2019). How cognitive psychology can inform evidence-based education reform. *The Cambridge Handbook of Cognition and Education*, 1–13. https://doi.org/10.1017/9781108235631.001

Eccles, J. S. (2016). Engagement: Where to next? *Learning and Instruction*, *43*, 71–75. https://doi.org/10.1016/j.learninstruc.2016.02.003

Engelmann, T., Tergan, S.-O., & Hesse, F. W. (2009). Evoking knowledge and information awareness for enhancing computer-supported collaborative problem solving. *The Journal of Experimental Education*, *78*(2), 268–290. https://doi.org/10.1080/00220970903292850

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, *20*(4), 399–483. https://doi.org/10.1207/s1532690xci2004_1

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*, 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fischer, F., Kollar, I., Stegmann, K., Wecker, C., Zottmann, J., & Weinberger, A. (2013). Collaboration scripts in computer-supported collaborative learning. *The International Handbook of Collaborative Learning*, 403–419. https://doi.org/10.4324/9780203837290.ch23

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, *95*(1), 148. https://doi.org/10.1037/0022-0663.95.1.148

Gadgil, S., & Nokes-Malach, T. J. (2012). Overcoming collaborative inhibition through error correction: A classroom experiment. *Applied Cognitive*

Psychology, *26*(3), 410–420. https://doi.org/10.1002/acp.1843

Garas-York, K., & Almasi, J. (2017). Constructing meaning through discussion. In S. Israel (Ed.), *Handbook of research on reading comprehension* (2nd ed., pp. 500–518). The Guilford Press.

Glogger, I., Schwonke, R., Holzäpfel, L., Nückles, M., & Renkl, A. (2012). Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. *Journal of Educational Psychology*, *104*(2), 452. https://doi.org/10.1037/a0026683

Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction*, *39*, 72–87. https://doi.org/10.1016/j.learninstruc.2015.05.001

Graffam, B. (2007). Active learning in medical education: strategies for beginning implementation. *Medical Teacher*, *29*(1), 38–42. https://doi.org/10.1080/01421590601176398

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., . . . Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, *96*(3), 403. https://doi.org/10.1037/0022-0663.96.3.403

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford publications. https://doi.org/10.1111/jedm.12050

Haywood, H. C., & Tzuriel, D. (2002). Applications and challenges in dynamic assessment. *Peabody Journal of Education*, *77*(2), 40–63. https://doi.org/10.1207/s15327930pje7702_5

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. https://doi.org/10.1016/j.compedu.2015.09.005

Hung, W. (2011). Theory to reality: A few issues in implementing problem-based learning. *Educational Technology Research and Development*, *59*, 529–552. https://doi.org/10.1007/s11423-011-9198-1

Hutner, T. L., & Sampson, V. (2015). New ways of teaching and observing science class. *Phi Delta Kappan*, *96*(8), 52–56. https://doi.org/10.1177/0031721715583964

Imai, M., Anderson, R. C., Wilkinson, I. A., & Yi, H. (1992). Properties of attention during reading lessons. *Journal of Educational Psychology*, *84*(2), 160. https://doi.org/10.1037/0022-0663.84.2.160

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? how can technologies help? *Educational Psychologist*, *51*(2), 247–265. https://doi.org/10.1080/00461520.2016.1158654

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, *38*(5), 1008–1022. https://doi.org/10.1111/cogs.12107

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, *51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of* *the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Kirch, S. A., & Siry, C. A. (2012). “maybe the algae was from the filter”: Maybe and similar modifiers as mediational tools and indicators of uncertainty and possibility in children’s science talk. *Research in Science Education*, *42*, 261–280. https://doi.org/10.1007/s11165-010-9197-y

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, *4*(4), 298–310. https://doi.org/10.1187/05-06-0082

Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers.* Chicago, IL: Follett. https://doi.org/10.1177/105960117700200220

Kolluri, S. (2018). Advanced placement: The dual challenge of equal access and effectiveness. *Review of Educational Research*, *88*(5), 671–711.

Konicek-Moran, R., & Keeley, P. (2015). *Teaching for conceptual understanding in science.* NSTA Press, National Science Teachers Association Arlington. https://doi.org/10.2505/9781938946103

Kosslyn, S. M., & Nelson, B. (2017). Why we need a new kind of higher education. In S. M. Kosslyn & B. Nelson (Eds.), *Building the intentional university—minerva and the future of higher education* (pp. 5–18). Cambridge, MA, USA: MIT Press. https://doi.org/10.7551/mitpress/9780262037150.003.0001

Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents’ thinking. *Psychological Science*, *22*(4), 545–552. https://doi.org/10.1177/0956797611402512

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. do recent studies falsify or verify earlier findings? *Educational Research Review*, *10*, 133–149. https://doi.org/10.1016/j.edurev.2013.02.002

Lam, R., & Kapur, M. (2018). Preparation for future collaboration: Cognitively preparing for learning from collaboration. *The Journal of Experimental Education*, *86*(4), 546–559. https://doi.org/10.1080/00220973.2017.1386156

Lam, R., & Muldner, K. (2017). Manipulating cognitive engagement in preparation-to-collaborate tasks and the effects on learning. *Learning and Instruction*, *52*, 90–101. https://doi.org/10.1016/j.learninstruc.2017.05.002

Lee, J., & Choi, H. (2019). Rethinking the flipped learning pre-class: Its influence on the success of flipped learning and related factors. *British Journal of Educational Technology*, *50*(2), 934–945. https://doi.org/10.1111/bjet.12618

Liang, L. A., & Dole, J. A. (2006). Help with teaching reading comprehension: Comprehension instructional frameworks. *The Reading Teacher*, *59*(8), 742–753. https://doi.org/10.1598/rt.59.8.2

Lim, J., Ko, H., Park, J., & Ihm, J. (2022). Effect of active learning and online discussions on the academic performances of dental students. *BMC Medical Education*, *22*(1), 312. https://doi.org/10.1186/s12909-022-03377-9

Lim, J., Ko, H., Yang, J. W., Kim, S., Lee, S., Chun, M.-S., ... Park, J. (2019). Active learning through discussion: ICAP framework for education in health professions. *BMC Medical Education*, *19*(1), 1–8. https://doi.org/10.1186/s12909-019-1901-7

Lim, J., & Park, J. (2023). Self-study enhances the learning effect of discussions. *Journal of the Learning Sciences*, 1–22. https://doi.org/10.1080/10508406.2023.2185148

Linnenbrink-Garcia, L., Rogat, T. K., & Koskey, K. L. (2011). Affect and engagement during small group instruction. *Contemporary Educational Psy-*

chology, *36*(1), 13–24. https://doi.org/10.1016/j.cedpsych.2010.09.001

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, *29*, 693–715. https://doi.org/10.1007/s10648-016-9379-x

Lombardi, D., Bailey, J. M., Bickel, E. S., & Burrell, S. (2018). Scaffolding scientific thinking: Students' evaluations and judgments during earth science knowledge construction. *Contemporary Educational Psychology*, *54*, 184–198. https://doi.org/10.1016/j.cedpsych.2018.06.008

Lombardi, D., Shipley, T. F., Astronomy Team, C. T. E. T. G. T. G. T., Biology Team, & Team, P. (2021). The curious construct of active learning. *Psychological Science in the Public Interest*, *22*(1), 8–43. https://doi.org/10.1177/1529100620973974

Lutz, S. L., Guthrie, J. T., & Davis, M. H. (2006). Scaffolding for engagement in elementary school reading instruction. *The Journal of Educational Research*, *100*(1), 3–20. https://doi.org/10.3200/joer.100.1.3-20

Manz, E., & Renga, I. P. (2017). Understanding how teachers guide evidence construction conversations. *Science Education*, *101*(4), 584–615. https://doi.org/10.1002/sce.21282

Mayer, R. E., & Alexander, P. A. (2011). *Handbook of research on learning and instruction.* Routledge New York. https://doi.org/10.4324/9780203839089

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

McTighe, J., & Lyman, F. T. (1988). Cueing thinking in the classroom: The promise of theory-embedded tools. *Educational Leadership*, *45*(7), 18–24.

Mende, S., Proske, A., & Narciss, S. (2021). Individual preparation for collaborative learning: Systematic review and synthesis. *Educational Psychologist*, *56*(1), 29–53. https://doi.org/10.1080/00461520.2020.1828086

Menekse, M., & Chi, M. T. (2019). The role of collaborative interactions versus individual construction on students' learning of engineering concepts. *European Journal of Engineering Education*, *44*(5), 702–725. https://doi.org/10.1080/03043797.2018.1538324

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, *102*(3), 346–374. https://doi.org/10.1002/jee.20021

Mercer, N. (2000). *Words and minds: How we use language to think together.* Psychology Press.

Meyer, D. K. (1993). What is scaffolded instruction? definitions, distinguishing features, and misnomers. *National Reading Conference Yearbook*, *42*, 41–53.

Miyake, N., & Norman, D. A. (1979). To ask a question, one must know enough to know what is not known. *Journal of Verbal Learning and Verbal Behavior*, *18*(3), 357–364. https://doi.org/10.1016/s0022-5371(79)90200-7

Moje, E., Afflerbach, P. P., Enciso, P., Lesaux, N. K., & Kwok, M. (2020). *Handbook of reading research, Volume V.* Routledge. https://doi.org/10.4324/9781315676302

Moore, J. C., & Rubbo, L. J. (2012). Scientific reasoning abilities of nonscience majors in physics-based courses. *Physical Review Special Topics-Physics Education Research*, *8*(1), 010106. https://doi.org/10.1103/physrevstper

.8.010106

Morocco, C. C., & Hindin, A. (2002). The role of conversation in a thematic understanding of literature. *Learning Disabilities Research & Practice*, *17*(3), 144–159. https://doi.org/10.1111/1540-5826.00041

Muldner, K., Lam, R., & Chi, M. T. (2014). Comparing learning from observing and from human tutoring. *Journal of Educational Psychology*, *106*(1), 69. https://doi.org/10.1037/a0034448

Murphy, P. K., Firetto, C. M., Wei, L., Li, M., & Croninger, R. M. (2016). What really works: Optimizing classroom discussions to promote comprehension and critical-analytic thinking. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 27–35. https://doi.org/10.1177/2372732215624215

Murphy, P. K., Greene, J. A., Firetto, C. M., Hendrick, B. D., Li, M., Montalbano, C., & Wei, L. (2018). Quality talk: Developing students' discourse to promote high-level comprehension. *American Educational Research Journal*, *55*(5), 1113–1160. https://doi.org/10.3102/0002831218771303

Murphy, P. K., Wilkinson, I. A., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology*, *101*(3), 740. https://doi.org/10.1037/a0015576

Narciss, S., Proske, A., & Koerndle, H. (2007). Promoting self-regulated learning in web-based learning environments. *Computers in Human Behavior*, *23*(3), 1126–1144. https://doi.org/10.1016/j.chb.2006.10.006

National Research Council (NRC). (2012). *A framework for k-12 science education: Practices, crosscutting concepts, and core ideas.* National Academies Press. https://doi.org/10.17226/13165

National Research Council (NRC). (2013). *Next generation science standards: For states, by states.* National Academies Press. https://doi.org/

10.17226/18290

Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? insights from research on collaborative learning. *Educational Psychology Review*, *27*, 645–656. https://doi.org/10.1007/s10648-015-9312-8

Park, J. (2017). Class p rep: A peer review system for class preparation. *British Journal of Educational Technology*, *48*(2), 511–523. https://doi.org/10.1111/bjet.12390

Patall, E. A., Pituch, K. A., Steingut, R. R., Vasquez, A. C., Yates, N., & Kennedy, A. A. (2019). Agency and high school science students' motivation, engagement, and classroom support experiences. *Journal of Applied Developmental Psychology*, *62*, 77–92. https://doi.org/10.1016/j.appdev.2019.01.004

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. *Handbook of Research on Student Engagement*, 259–282. https://doi.org/10.1007/978-1-4614-2018-7_12

Poh, M.-Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, *57*(5), 1243–1252. https://doi.org/10.1109/tbme.2009.2038487

Raphael, T. E., & McMahon, S. I. (1994). Book club: An alternative framework for reading instruction. *The Reading Teacher*, *48*(2), 102–116. https://doi.org/10.1598/rt.48.2.1

Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, *36*(4), 257–267. https://doi.org/10.1016/j.cedpsych.2011.05.002

Reznitskaya, A., Glina, M., Carolan, B., Michaud, O., Rogers, J., & Sequeira,

L. (2012). Examining transfer effects from dialogic discussions to new tasks and contexts. *Contemporary Educational Psychology*, *37*(4), 288–306. https://doi.org/10.1016/j.cedpsych.2012.02.003

Reznitskaya, A., & Gregory, M. (2013). Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist*, *48*(2), 114–133. https://doi.org/10.1080/00461520.2013.775898

Riener, C., & Willingham, D. (2010). The myth of learning styles. *Change: The Magazine of Higher Learning*, *42*(5), 32–35. https://doi.org/10.1080/00091383.2010.503139

Rosenberg, E. L., & Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cognition & Emotion*, *8*(3), 201–229. https://doi.org/10.1080/02699939408408938

Schalk, L., Schumacher, R., Barth, A., & Stern, E. (2018). When problem-solving followed by instruction is superior to the traditional tell-and-practice sequence. *Journal of Educational Psychology*, *110*(4), 596. https://doi.org/10.1037/edu0000234

Schmidt, H. G., Cohen-Schotanus, J., Van Der Molen, H. T., Splinter, T. A., Bulte, J., Holdrinet, R., & Van Rossum, H. J. (2010). Learning more by being taught less: a "time-for-self-study" theory explaining curricular effects on graduation rate and study duration. *Higher Education*, *60*, 287–300. https://doi.org/10.1007/s10734-009-9300-3

Schmidt, H. G., Loyens, S. M., Van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on kirschner, sweller, and. *Educational Psychologist*, *42*(2), 91–97. https://doi.org/10.1080/00461520701263350

Schmidt, H. G., Van der Molen, H. T., Te Winkel, W. W., & Wijnen,

W. H. (2009). Constructivist, problem-based learning does work: A meta-analysis of curricular comparisons involving a single medical school. *Educational Psychologist*, *44*(4), 227–249. https://doi.org/10.1080/00461520903213592

Schmidt, H. G., Wagener, S. L., Smeets, G. A., Keemink, L. M., & van Der Molen, H. T. (2015). On the use and misuse of lectures in higher education. *Health Professions Education*, *1*(1), 12–18. https://doi.org/10.1016/j.hpe.2015.11.010

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475–5223. https://doi.org/10.1207/s1532690xci1604_4

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759. ttps://doi.org/10.1037/a0025140

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129–184. https://doi.org/10.1207/s1532690xci2202_1

Schwartz, D. L., Sears, D., & Chang, J. (2007). Reconsidering prior knowledge. *Thinking With Data*, 319–344.

Sharp, A. M. (1995). Philosophy for children and the development of ethical values. *Early Child Development and Care*, *107*(1), 45–55. https://doi.org/10.1080/0300443951070106

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist*, *50*(1), 1–13. https://doi.org/10.1080/00461520.2014.1002924

Slavin, R. E. (2012). Cooperative learning and achievement: Theory and research. *Handbook of Psychology, Second Edition*, *7*. https://doi.org/10.1002/9781118133880.hop207008

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, *323*(5910), 122–124. https://doi.org/10.1126/science.1165919

Soter, A. O., Wilkinson, I. A., Murphy, P. K., Rudge, L., Reninger, K., & Edwards, M. (2008). What the discourse tells us: Talk and indicators of high-level comprehension. *International Journal of Educational Research*, *47*(6), 372–391. https://doi.org/10.1016/j.ijer.2009.01.001

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., . . . Young, A. M. (2018). Anatomy of stem teaching in north american universities. *Science*, *359*(6383), 1468–1470. https://doi.org/10.1126/science.aap8892

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613. https://doi.org/10.1037/0003-066x.52.6.613

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology*, *69*(5), 797. https://doi.org/10.1037/0022-3514.69.5.797

Strobel, J., & Van Barneveld, A. (2009). When is pbl more effective? a meta-synthesis of meta-analyses comparing pbl to conventional classrooms. *Interdisciplinary Journal of Problem-Based Learning*, *3*(1), 44–58. https://doi.org/10.7771/1541-5015.1046

Sun, J., Anderson, R. C., Lin, T.-J., Morris, J. A., Miller, B. W., Ma, S., . . . Scott, T. (2022). Children's engagement during collaborative learning and direct instruction through the lens of participant structure. *Contemporary Educational Psychology*, *69*, 102061. https://doi.org/10.1016/j.cedpsych.2022.102061

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist*, *42*(2), 115–121. https://doi.org/10.1080/00461520701263426

Taber, K. S., & Akpan, B. (2016). *Science education: An international course companion*. Springer. https://doi.org/10.1007/978-94-6300-749-8

Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? *Discourse, Tools and Reasoning: Essays on Situated Cognition*, 361–384. https://doi.org/10.1007/978-3-662-03362-3_16

Traum, D. R., & Heeman, P. A. (1996). Utterance units in spoken dialogue. In H. Niemann, R. De Mori, & G. Hanrieder (Eds.), *Dialogue processing in spoken language systems* (pp. 125–140). Springer Verlag. https://doi.org/10.1007/3-540-63175-5_42

Tsovaltzi, D., Judele, R., Puhl, T., & Weinberger, A. (2015). Scripts, individual preparation and group awareness support in the service of learning in facebook: How does cscl compare to social networking sites? *Computers in Human Behavior*, *53*, 577–592. https://doi.org/10.1016/j.chb.2015.04.067

Van Amelsvoort, M., Andriessen, J., & Kanselaar, G. (2007). Representational tools in computer-supported collaborative argumentation-based learning: How dyads work with constructed and inspected argumentative diagrams. *Journal of the Learning Sciences*, *16*(4), 485–521. https://doi.org/10.1080/10508400701524785

Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction*, *10*(4), 311–330. https://doi.org/10.1016/s0959-4752(00)00002-5

van de Pol, J., Mercer, N., & Volman, M. (2019). Scaffolding student understanding in small-group work: Students' uptake of teacher support in subsequent small-group interaction. *Journal of the Learning Sciences*, *28*(2), 206–239. https://doi.org/10.1080/10508406.2018.1522258

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Waggoner, M., Chinn, C., Yi, H., & Anderson, R. C. (1995). Collaborative reasoning about stories. *Language Arts*, *72*(8), 582–589.

Wang, M.-T., & Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction*, *28*, 12–23. https://doi.org/10.1016/j.learninstruc.2013.04.002

Webb, N. M., Nemer, K. M., & Ing, M. (2006). Small-group reflections: Parallels between teacher discourse and student behavior in peer-directed groups. *Journal of the Learning Sciences*, *15*(1), 63–119. https://doi.org/10.1207/s15327809jls1501_8

Wei, L., & Murphy, P. K. (2017). Teacher and student roles: Walking the gradually changing line of responsibility. *Classroom Discussions in Education*, (pp.30–53). Routledge. https://doi.org/10.4324/9781315174594-2

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*, *17*(4), 416–426. https://doi.org/10.1016/j.learninstruc.2007.03.007

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In

M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). Orlando, FL: Academic Press. https://doi.org/10.1016/b978-012109890-2/50045-7

Wortham, S. (2006). *Learning identity: the joint emergence of social identification and academic learning.* Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511810015

Wu, X., Anderson, R. C., Nguyen-Jahiel, K., & Miller, B. (2013). Enhancing motivation and engagement through collaborative discussion. *Journal of Educational Psychology*, *105*(3), 622. https://doi.org/10.1037/a0032792

Yannier, N., Hudson, S. E., Koedinger, K. R., Hirsh-Pasek, K., Golinkoff, R. M., Munakata, Y., ... Brownell, S. E. (2021). Active learning: "hands-on" meets "minds-on". *Science*, *374*(6563), 26–30. https://doi.org/10.1126/science.abj9957

Zimmerman, B. J., & Schunk, D. H. (2008). An essential dimension of self-regulated learning. *Motivation and Self-Regulated Learning: Theory, Research, and Applications*, *1*, 1–30. https://doi.org/10.4324/9780203831076

Appendix A. Examples of topics covered in the lecture and learning material about the Korean code of criminal procedure

1. Accusation: an accusation made by a victim of a crime or a person representing the victim, to report the crime to the investigative authorities and seek punishment for the offender.
2. Complaint: charge or claim made by a third-party (any person other than the criminal or the victim) about a crime to seek punishment against an offender.
3. Recognition: the most fundamental way that the investigative authority initiates an investigation.

cf. Report: Complaint expresses the intent to seek punishment for the offender, while report refers to reporting the offense.

4. Investigative authority: people authorized by law to investigate crimes (e.g. prosecutor, judicial officer).
5. Accusation capacity: ability to deal with accusation in legal proceedings. It refers to competence to understand the offense and the social interests related to the accusation. A thorough knowledge of civil law is not required. Proper communication ability is sufficient.
6. Offense subject to complaint: An offense that requires an accusation by victims or other persons for prosecution.

e.g., Contempt, Infringement on privacy

(1) Absolute offense subject to complaint: Offense subject to complaint that is established based on criminal fact regardless of status

(2) Relative offense subject to complaint: Relationship exists between criminals and victims such as kinship, or marriage relationship.

Appendix B. Samples of three types of test questions: verbatim, paraphrased, and transfer items

Examples of verbatim items

1. Given that there is no one who filed an accusation against a crime subject for prosecution, prosecutors must designate a person who can file the complaint within (   ) days upon the request of the stakeholders. *(Answer: 10)*

2. Who does not have the right to file a complaint? *(Answer: number 2)*

(1) Legal representatives

(2) Property manager

(3) Guardian

(4) A representative of the corporation which is the victim

7. According to Article 223, a victim of a crime has the right to accuse the offender. However, although the victim cannot accuse the one's relatives nor his/her spouse, (   ) is an exception. *(Answer: sexual crime)*

Examples of paraphrased items

11. Explain who the entitled person with the right to file a complaint is.

(Answer: Provided that there is no one to make the accusation (in case of an offense subject to complaint), prosecutors shall designate the person with the right to file a complaint within 10 days upon request by stakeholders.)

12. Define who the victim is and explain the cases where he/she is unable to make an accusation.

(Answer: A victim is a person whose life, body, property, or fame has been infringed upon or threatened. The victim cannot accuse the offender if he/she is one's or his/her spouse's linear ascendants.)

13. Explain the following concepts: accusation, complaint, recognition. Discuss their common features.

(Answer: Accusation: an expression of will by the victim of a crime or a person representing the victim; Complaint: an expression of will by the third party or witness; Recognition: the most basic way that the investigative authority initiates an investigation. Common feature: they occur at the beginning of an investigation.)

Examples of transfer items

17. The under-aged victim (V) accused the offender (D) of contempt, and then withdrew his accusation on July 26th, 2017. Afterwards V's mother (M), the legal representative of V, accused D on August 3rd, 2017. D was charged with contempt and was found guilty on the first trial. However, D made an appeal claiming M's complaint is not valid because V has already withdrawn his complaint, and thus, the prosecutor's indictment is against the provisions of the

law. Will the Court of Appeals accept D's claim? *(Answer: A legal representative of an under-aged victim can independently file a complaint regardless of whether the victim's complaint is nullified. Such complaint can even go against the victim's stated will. Thus, even if victim V withdraws his accusation, the complaint of V's legal representative M is still effective. In conclusion, the Court of Appeals will reject D's claim.)*

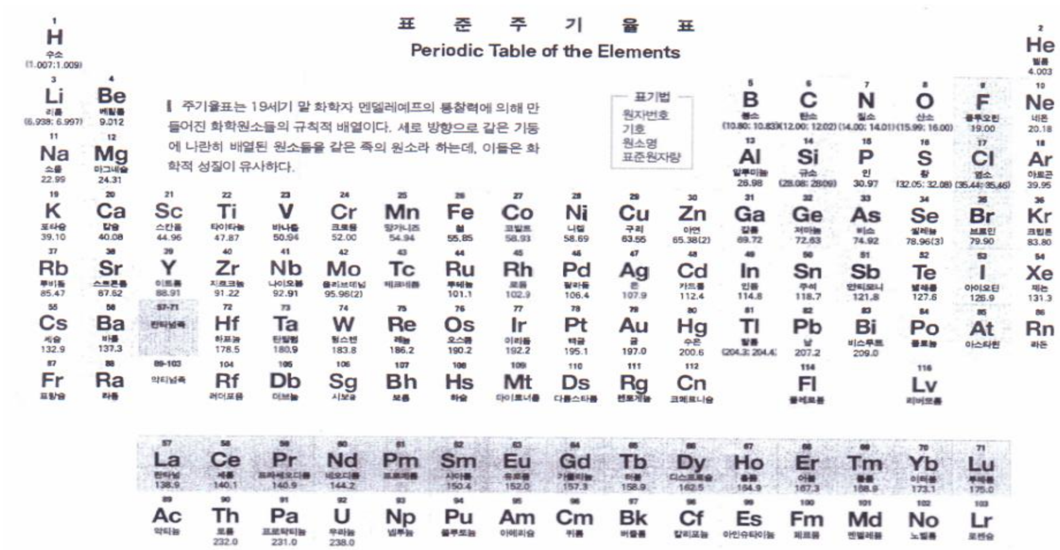
18. On May 1st, 2016 the father (F) of a famous under-aged celebrity (V), accused person (D) of infringement on privacy claiming that D has hacked V's computer. However, F's parental rights on V were terminated on January 1st, 2017. D was charged with infringement on privacy and was convicted guilty on the first trial. Eventually, D made an appeal claiming that F has no right to file a complaint because his parental rights were terminated. Thus, D claims that the accusation should be dropped. Will the Court of Appeals accept D's claim? *(Answer: The status of a legal representative must be determined based on the day of the accusation. If the status was valid on the day of accusation, the loss of status at the time of the crime or after the accusation does not invalidate the accusation. In conclusion, even if F's parental rights on V were terminated after the day of accusation, F's complaint is still valid. Thus, the Court of Appeals will reject D's claim.*

Appendix C. Examples of interaction and non-interaction episodes

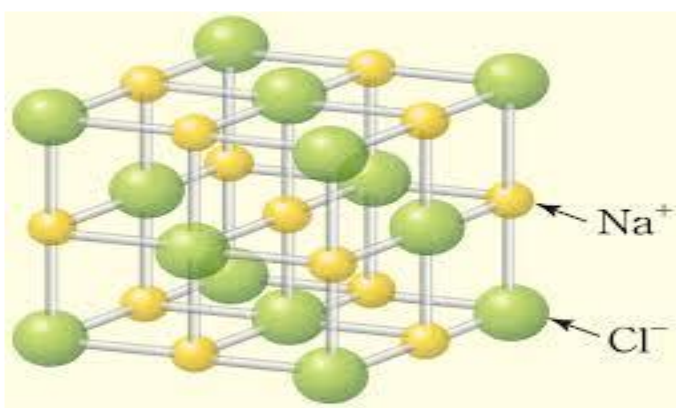
Interaction Episode 1	Student 1: Are there crimes that correspond to relative offense subject to complaint?	*Turn* 1	*Co-Constructive turn*
	Student 2: It doesn't say that there are crimes related to relative or absolute offense subject to complaint. Rather crimes such contempt, is a prime example of offense subject to complaint.	*Turn* 2	
	Student 3: In other words, doesn't it mean that offenses subject to complaints are categorized by relationships rather than the type of crime? Therefore, contempt can either be absolute or relative offense subject to complaint.	*Turn* 3	*Co-Constructive turn*
Interaction Episode 2	Student 4: I'm curious why the crime not prosecuted against objection subject to charge. Since the prosecutor can indict without accusation, it cannot be charge, can it be categorized as recognition?	*Turn* 1	*Co-Constructive turn*
	Student 5: I think it can be recognition or accusation. You can initiate the investigation with accusation, or with recognition without accusation.	*Turn* 2	*Co-Constructive turn*

	Student 6: That’s right, I understand it this way. That there are two types of crime, offenses subject to complaint or not subject to complaint or not subject to complaint. Sexual offenses are in a separate category.	*Turn* 3
Non-Interaction Episode 1*	Student 7: I don't understand what dual liability is. Here in this part about offense subject to complaint. Student 8: Me, neither.	*Turn 1*
Non-Interaction Episode 2*	Student 9: Are minors really seen as incapable of making regal decisions? Student 10: Once they reach high school age, they can be seen as capable. Student 11: Yeah, high school age ...	*Turn 1*

Appendix D. Samples of written materials and text items



- All the elements in our universe are listed on the periodic table that Mendeleev produced with his insight in the late 19th century.



- Salt and sodium chloride (NaCl) are ion salts composed of ion bonds, and the binding energy between two ions of Na+ and Cl- is stronger than the O--H hydrogen bonding energy of water molecules. But ice is harder to break than salt. Why is salt more brittle? (4 points)

Appendix E. Samples of written materials and test items (2)

a. Written instructional material

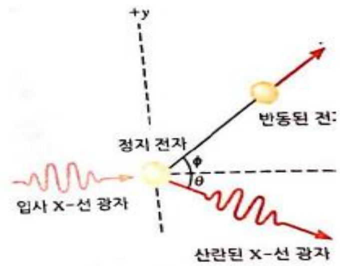


그림 29.10 콤프턴이 수행한 실험에
선 광자는 정지된 전자와 충돌한다.
광자와 반동된 전자는 충돌에 의해
향으로 멀어진다.

American physicist Compton explained that electrons in graphite scatter X-rays with photon models. X-rays are electromagnetic waves of large frequencies and are composed of photons like light. Figure 29.10 shows how X-rays collide with electrons in a graphite piece. X-ray photons scatter in one direction after the collision, and electrons bounce in another direction. Compton observed that the scattered photons have a frequency f´ less than the frequency f of the photons when they enter.

b. Test items

Question 1. Ultraviolet light (intensity [intensity] 1.00 [W/m^2]) with a wavelength of 350 [nm] is directed to the surface of potassium. [8]

A) Find the maximum value of the kinetic energy (KE) of photoelectrons. [3]

B) If 0.5[%] of incident photons generate photoelectrons, find how many photons are emitted per second if the area of the potassium surface is 1.00[cm^2]. [5]

Appendix F. Examples of interaction and non-interaction episodes

	Line				
Interaction Episode 3	1	Student 1:	Is there any reason to use the law of conservation of angular momentum here?	Turn 1 (Sub)	
	2	Student 3:	Wouldn't the same conclusion be reached using the law of conservation of momentum?	Turn 2 (Sub)	Co-constructive Turn
	3	Student 2:	Yeah. This will be the same as the conservation of angular momentum or the conservation of momentum by calculating all the angle changes.	Turn 3 (Sub)	Co-constructive Turn
	4	Student 1:	Is the calculation taking into account the fact that the electrons are distorted?	Turn 4 (Sub)	Co-constructive Turn
Interaction Episode 7	1	Student 6:	Why is energy *hf*? If *n* is just an integer, I think that it can be not only *hf* but also *ehf*.	Turn 1 (Sub)	
	2	Student 4:	So, isn't it just talking about the *hf* value, not by tying up the size?	Turn 2 (Sub)	Co-constructive Turn

	3	Student 5:	I think so.	Turn 3
Non-Interaction Episode 4	1	Student 7:	Okay, we...	Turn 1
	2	Student 8:	We are taking the test soon?	Turn 2
	3	Student 9:	Uh huh.	Turn 3
Non-Interaction Episode 11	1	Student 10:	(reading the material) Light is an electromagnetic wave, and those waves have a continuous shape in electric and magnetic fields.	Turn 1
	2	Student 12:	Will it come out as a test question?	Turn 2
	3	Student 10:	Uh...umm. Maybe not.	Turn 3
	4	Student 11:	Shall we talk about what each of us thinks is important?	Turn 4 (Sub)
	5	Student 12:	Mmhm...	

Appendix G. Samples of learning materials and test items

i) Written materials

- *Reporting a crime*

A complaint is a report of a criminal offence. In general, the criminal offence must be specified, but the degree of specificity is only required to ensure that the complainant's intention is to specify which criminal offence he or she is seeking punishment for. In other words, it is not necessary for the complainant himself or herself to specify the criminal offence, pointing out in detail the time, place and method of the offence.

Therefore, even if the name of the perpetrator is not known exactly, or if there is a mistake in the report, or if the time, place, and method of the crime are not clear or incorrect, it does not affect the effectiveness of the complaint.

Also, since a complaint is a report of a crime, it is not necessary to specify who the perpetrator is. However, in the case of "relative confession" between close relatives such as immediate blood relatives or spouses, the identity of the perpetrator must be revealed in light of the nature of the crime.

ii) Test items for immediate/long-term retention

Who can't sue if the victim is deceased? (1 point)

1. the victim's brother

2. the victim's wife

3. the victim's mother-in-law

4. the victim's father

iii) Test items for immediate/long-term transfer

The under-aged victim (V) accused person (D) of rape, but eventually withdrew her accusation on February 1st, 2017. Afterwards V's father (F), legal representative of V, accused D on February 10th, 2017. D was charged with rape and was convicted of the crime on the first trial. However, D made an appeal claiming that F's complaint should not take effect since V already withdrew her complaint, and thus, the prosecutor's indictment was against the provisions of the law. Will the Court of Appeals accept D's claim? (4 points)

Appendix H. Excerpts of interaction episode

	Line				
Interaction Episode 4	1	Student 1:	Did you get to the end of this?	Turn 1	
	2	Student 3:	I couldn't read the back of it.	Turn 2	
	3	Student 1:	Here it says inherent right, but I don't understand what inherent right is. It says that the next of kin has the right to sue, and it also says that the legal representative has the right to sue, but I don't understand.	Turn 3 (Sub)	
	4	Student 3:	I think it's related to this part about the victim's next of kin being able to sue independently.	Turn 4 (Sub)	Co-constructive Turn
	5	Student 2:	I think it might have something to do with the fact that you said the prosecutor can assign it if it's an orphan,	Turn 5 (Sub)	Co-constructive Turn

Interaction Episode 11	17	Student 2:	Is there something you're not sure about?	Turn 1 (Sub)	
	18	Student 3:	Insulting... Well, do you know the difference between insult and libel?	Turn 2 (Sub)	Co-constructive Turn
	19	Student 1:	I thought insults were lesser offences, or so I've been told.	Turn 3 (Sub)	Co-constructive Turn

Appendix I. Samples of learning materials and test items for Experiment 1

a. Learning materials

- Recognition: the most fundamental way that the investigative authority initiates an investigation.

cf. Report: Complaint expresses the intent to seek punishment for the offender, while report refers to reporting the offense.

- Investigative authority: people authorized by law to investigate crimes (e.g. prosecutor, judicial officer).

b. Example of questions for shallow knowledge

According to Article 223, a victim of a crime has the right to accuse the offender. However, although the victim cannot accuse the one's relatives nor his/her spouse, (   ) is an exception (1 point)

c. Example of questions for transfer knowledge

On May 1st, 2016 the father (F) of a famous under-aged celebrity (V), accused person (D) of infringement on privacy claiming that D has hacked V's computer. However, F's parental rights on V were terminated on January 1st, 2017. D was charged with infringement on privacy and was convicted guilty on the first trial. Eventually, D made an appeal claiming that F has no right to file a complaint because his parental rights were terminated. Thus, D claims that the accusation should be dropped. Will the Court of Appeals accept D's claim? (5 point)

Appendix J. Samples of scoring rubrics for transfer items

On May 1st, 2016 the father (F) of a famous under-aged celebrity (V), accused person (D) of infringement on privacy claiming that D has hacked V's computer. However, F's parental rights on V were terminated on January 1st, 2017. D was charged with infringement on privacy and was convicted guilty on the first trial. Eventually, D made an appeal claiming that F has no right to file a complaint because his parental rights were terminated. Thus, D claims that the accusation should be dropped. Will the Court of Appeals accept D's claim? (*Answer: The status of a legal representative must be determined based on the day of the accusation (1). If the status was valid on the day of accusation, the loss of status at the time of the crime or after the accusation does not invalidate the accusation (1). In conclusion, even if F's parental rights on V were terminated after the day of accusation, F's complaint is still valid (1). Thus, the Court of Appeals will reject D's claim (2)* (5 point)

- The score is evaluated for each sentence, and even if the correct term is not used, it is considered correct if the students seem to have understood.

1) Did students apply the concepts they learned to the questions? 0 or 1, respectively)

   The status of a legal representative must be determined based on the day of the accusation (1) even if F's parental rights on V were terminated after the day of accusation, F's complaint is still valid (1)

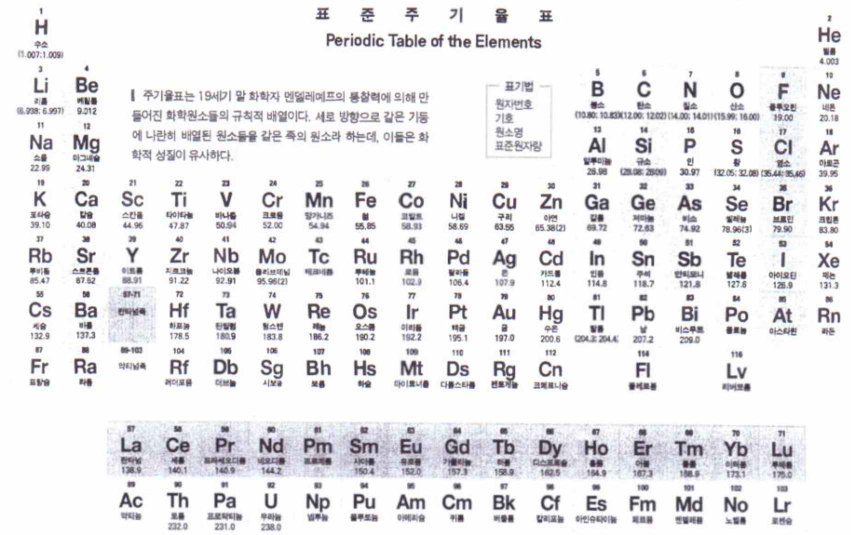
2) Is there any evidence for what they described? (0 or 1, respectively)

   If the status was valid on the day of accusation, the loss of status at the time of the crime or after the accusation does not invalidate the accusation (1)

3) Is the conclusion correct? (0 or 2 pts)

   Thus, the Court of Appeals will reject D's claim (2)

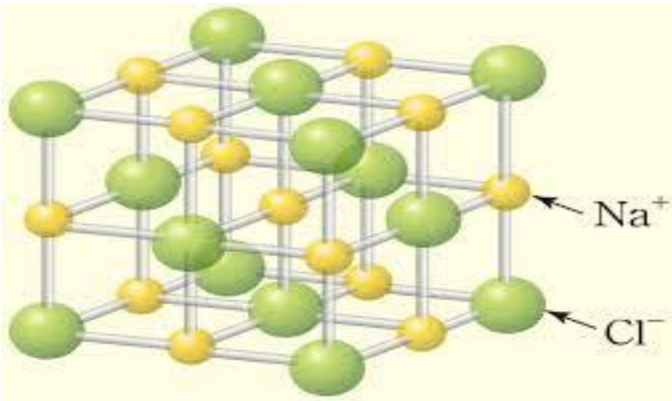
Appendix K. Samples of learning materials and test items for Experiment 2



a. Learning materials

- All the elements in our universe are listed on the periodic table that Mendeleev produced with his insight in the late 19th century.

b. Example of questions for transfer knowledge



Salt and sodium chloride (NaCl) are ion salts composed of ion bonds, and the binding energy between two ions of Na+ and Cl- is stronger than the O--H hydrogen bonding energy of water molecules. But ice is harder to break than salt. Why is salt more brittle? (4 points)

Appendix L. Coding schemes for discussions

1. Substantive comment

A substantive comment pertains to an idea relevant to the concepts being taught, regardless of whether the comments are correct or not (Chi et al., 2017).

Example: Can you explain this? What exactly is the van der Waals force? (to other students in group)

2. Substantive question

A substantive question is determined whether the question (a) required generative rather than factual knowledge and verbatim recall and (b) elaborative more than a yes/no response:

1) Generative question: students ask the instructor questions that are not directly in the instructional material.

Example: As the electrons drop off, the elements are released and then filled with electrons to follow the octet rule again, forming new bonds between the elements. Is this right? (to the instructor)

2) Elaborative question: students answer and explain the questions provided by the instructor in their own language or build an understanding beyond the content of the learning material.

Example: What does it mean exactly? (to the instructor)

3. Elaborative feedback from instructor

An instructor adds explanations beyond saying that the student's answer/explanation is incorrect or incomplete.

Example: Excellent. This second electron shell is called the outermost electron shell.

4. Others

1) Simple answer: students answer the instructor's question, but does not develop into further interaction.

Example: According to the octet rule, maybe.

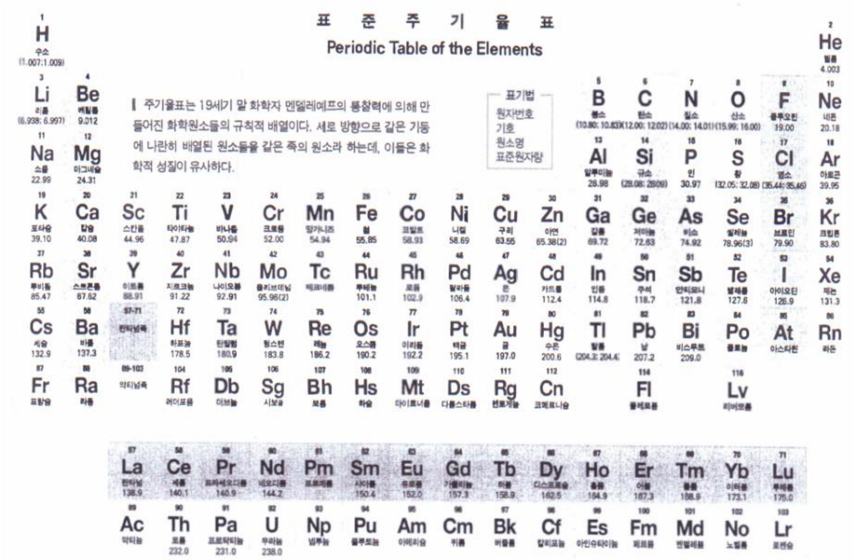
2) Meta-talk and off-talk: students talk about the task itself, instructions etc.

Example: How much discussion time is left?

3) Agreement: students answer and agree with the instructor’s questions.

Example: Oh, we agree with.

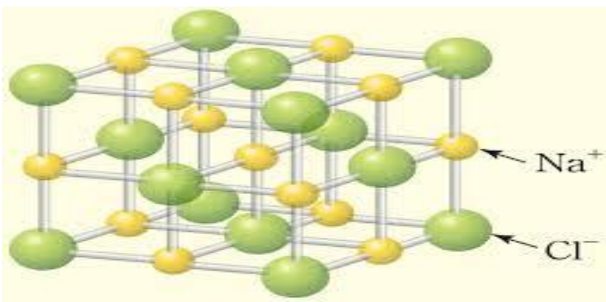
Appendix M. Samples of instructional materials and test items for Experiment



a. Learning materials

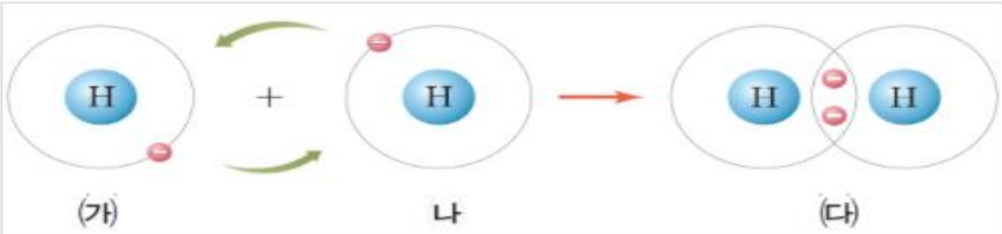
- All the elements in our universe are listed on the periodic table that Mendeleev produced with his insight in the late 19th century. The elements in the rightmost column of the periodic table are chemically very stable inert elements, which are admired by all elements. Everyone admires the arrangement of electrons that inert elements have in a stable way. Inert elements have 2 or 8 electrons in their outermost shell.

b. Transfer type items



- Salt and sodium chloride (NaCl) are ion salts composed of ion bonds, and the binding energy between two ions of Na+ and Cl- is stronger than the O--H hydrogen bonding energy of water molecules. But ice is harder to break than salt. Explain in 3 - 4 lines why salt is more brittle (4 points)

- The following figure shows a model of the process in which hydrogen atoms combine to form hydrogen molecules. Which of the following descriptions of the model is correct in <Example>? (2 points)



<Example>

a. (A) and (B) are the same particles.

b. (C) is a more stable particle than (A).

c. (A) and (B) share a pair of electrons.

(1) a (2) c (3) a, b (4) b, c (5) a, b, c

- Ionic bonds are not covalent bonds, but they can all be described by octet rules (O, X) (1 point)

Appendix N. Scoring rubrics for engagement behavior during discussions

(Adapted at Lutz et al., 2006)

1. Totally distracted or entirely not engaged; head completely down on desk (i.e., not participating in task)
2. Distracted by something unrelated to task; hard to judge whether student is truly behaviorally engaged prolonged yawn
3. Not off task, but does not appear particularly involved; eyes may or be on something, but does not seem to really be following discussion or actively engaged in activity; may be slouching
4. Clearly on-task, as suggested by eye movement and posture toward peer (or speaker); raising hand (perhaps just briefly); writing; speaking; clearly listening (suggesting that student is attentive at least behaviorally)
5. Waving hand; hand "shoots" into air to answer question; making noises that suggest great enthusiasm and eagerness to participate; otherwise seems "super-engaged"

Appendix O. Descriptive statistics and Bonferroni-corrected post-hoc test on experimental groups' engagement across discussion time

Time	Groups	Mean	(*SD*)	Post-hoc test
1	LSD	1.36	(0.52)	STD>SSD>LSD>LTD
	LTD	1.29	(0.46)	
	SSD	1.97	(0.67)	
	STD	2.05	(0.76)	
2	LSD	1.50	(0.56)	STD>SSD>LSD>LTD
	LTD	1.51	(0.58)	
	SSD	2.13	(0.81)	
	STD	2.47	(0.63)	
3	LSD	1.58	(0.66)	STD>SSD>LSD, LTD
	LTD	1.49	(0.60)	
	SSD	2.46	(1.15)	
	STD	2.66	(0.98)	
4	LSD	1.73	(0.57)	STD>SSD>LSD, LTD
	LTD	1.36	(0.51)	
	SSD	2.84	(0.97)	
	STD	3.22	(1.11)	
5	LSD	1.84	(0.86)	STD>SSD>LSD>LTD
	LTD	1.74	(0.55)	
	SSD	2.54	(1.19)	
	STD	3.74	(1.41)	
6	LSD	1.88	(0.81)	STD>SSD>LTD>LSD
	LTD	1.96	(0.75)	

	SSD	3.30	(1.03)	
	STD	3.95	(1.26)	
7	LSD	1.94	(0.81)	STD>SSD>LTD>LSD
	LTD	2.42	(0.85)	
	SSD	3.46	(1.27)	
	STD	4.05	(1.30)	
8	LSD	1.80	(0.82)	STD>SSD>LTD>LSD
	LTD	2.52	(0.77)	
	SSD	3.59	(1.29)	
	STD	4.17	(0.82)	
9	LSD	1.89	(0.99)	SSD, STD>LTD>LSD
	LTD	2.63	(0.74)	
	SSD	3.76	(1.25)	
	STD	4.31	(0.82)	
10	LSD	1.94	(0.94)	SSD, STD>LTD>LSD
	LTD	2.70	(0.83)	
	SSD	3.59	(1.38)	
	STD	4.16	(0.87)	
11	LSD	1.92	(0.82)	SSD, STD>LTD>LSD
	LTD	3.05	(0.80)	
	SSD	4.06	(1.22)	
	STD	4.20	(1.00)	
12	LSD	2.16	(0.93)	SSD, STD>LTD>LSD
	LTD	3.14	(0.75)	
	SSD	3.92	(1.47)	

	STD	4.05	(1.27)	
13	LSD	2.05	(0.88)	SSD, STD>LTD>LSD
	LTD	3.21	(0.88)	
	SSD	3.84	(1.66)	
	STD	4.25	(0.92)	
14	LSD	2.08	(0.86)	SSD, STD>LTD>LSD
	LTD	3.41	(0.84)	
	SSD	3.70	(1.24)	
	STD	4.00	(1.36)	
15	LSD	2.13	(0.93)	SSD, STD>LTD>LSD
	LTD	3.39	(1.08)	
	SSD	3.66	(1.18)	
	STD	4.02	(1.24)	
16	LSD	2.36	(1.07)	SSD, STD>LTD>LSD
	LTD	3.35	(0.97)	
	SSD	3.86	(0.97)	
	STD	4.15	(1.07)	
17	LSD	2.27	(1.00)	SSD, STD>LTD>LSD
	LTD	3.38	(0.91)	
	SSD	4.42	(0.86)	
	STD	3.57	(1.22)	
18	LSD	2.34	(1.09)	SSD>STD>LTD>LSD
	LTD	3.47	(1.03)	
	SSD	4.49	(0.90)	
	STD	3.80	(0.95)	

19	LSD	2.34	(1.04)	SSD>STD>LTD>LSD
	LTD	3.25	(1.08)	
	SSD	4.50	(0.79)	
	STD	3.50	(0.82)	
20	LSD	2.30	(1.06)	SSD>STD>LTD>LSD
	LTD	3.47	(1.24)	
	SSD	4.50	(0.85)	
	STD	3.53	(1.01)	
21	LSD	2.26	(1.24)	SSD>LTD, STD>LSD
	LTD	3.26	(1.09)	
	SSD	4.54	(0.79)	
	STD	2.98	(1.02)	
22	LSD	2.30	(1.17)	SSD>LTD, STD>LSD
	LTD	3.42	(1.00)	
	SSD	4.50	(0.83)	
	STD	3.33	(1.26)	
23	LSD	2.51	(1.35)	SSD>LTD, STD>LSD
	LTD	2.91	(1.25)	
	SSD	3.91	(1.05)	
	STD	2.60	(0.94)	
24	LSD	2.34	(1.08)	SSD>LTD, STD>LSD
	LTD	3.02	(1.10)	
	SSD	4.36	(1.00)	
	STD	2.23	(0.84)	
25	LSD	2.28	(1.21)	SSD>LTD>STD>LSD
	LTD	3.12	(1.10)	

	SSD	3.77	(1.03)	
	STD	2.14	(0.87)	
26	LSD	2.30	(1.10)	SSD>LTD>LSD, STD
	LTD	3.15	(0.97)	
	SSD	3.91	(0.86)	
	STD	2.14	(0.87)	
27	LSD	2.11	(0.88)	SSD>LTD>LSD, STD
	LTD	2.97	(1.04)	
	SSD	3.65	(0.97)	
	STD	2.22	(0.72)	
28	LSD	2.06	(1.02)	SSD>LTD>LSD>STD
	LTD	2.75	(1.21)	
	SSD	3.17	(1.36)	
	STD	1.97	(0.77)	
29	LSD	1.85	(0.96)	SSD>LTD>LSD>STD
	LTD	2.30	(1.03)	
	SSD	3.02	(1.25)	
	STD	1.67	(0.76)	
30	LSD	1.94	(0.98)	SSD>LSD, LTD>STD
	LTD	1.91	(0.87)	
	SSD	2.61	(0.89)	
	STD	1.47	(0.56)	

Note. p-values of all significant group differences are less than 0.001. LSD, student-led discussion after a lecture; LTD, teacher-led discussion after a lecture; SSD, student-led discussion after self-study; and STD, teacher-led discussion after self-study.

국문초록

대한민국의 교실에서 토론은 활발하게 이루어지지 않고 있다. 그 교육적 효과를 보여주는 연구가 많음에도 불구하고 초등학교에서부터 대학에 이르기까지 수업 시간은 전통적인 강의가 중심이다. 하지만, 전통적 강의의 학습 효과가 크지 않기 때문에, 수업에서 강의를 줄이고 토론을 사용하려는 변화가 필요하다. 이런 맥락에서, 본 논문에서는 토론을 수업에서 활용하는 방안을 탐구한다. 특히, 어떤 준비가 효과적인 토론으로 이어지는지를 살펴보고, 그 결과를 일반화하기 위한 일련의 연구를 수행하였다.

제 1장에서는 대학생을 대상으로 세 집단, 강의와 복습, 강의와 토론, 그리고 자율학습과 토론 집단을 비교하였다. 사전 지식의 영향을 최소화하기 위해 법학 컨텐츠가 사용되었다. 세 개의 실험의 결과, 토론 집단이 복습 집단보다 유의미하게 높은 점수를 받는다는 것을 확인하였다. 게다가, 자율학습과 토론 집단이 강의와 토론 집단보다 더 좋은 성과를 얻었다는 것도 발견하였다. 토론의 대화 내용 분석 결과, 자율학습 집단이 강의 집단보다 토론에서 더 적극적이고 생산적인 상호작용한 것을 확인하였다. 이러한 결과는 자율학습 과정에서의 어려움과 궁금증을 후속 토론에서 해결하였기 때문인 것으로 보인다.

제 2장에서는 자율학습과 토론 조합이 갖는 이점을 새로운 학습 자료와 새로운 참여자 집단에 대해 일반화하고자 하였다. 먼저, 과학 교육에서 대학생을 대상으로 두 개의 실험을 수행하였다. 그 결과 사후평가에서 자율학습과 토론 조합이 강의와 토론 조합, 강의와 복습 조합보다 유의미하게 높은 점수를 받는다는 것을 확인하였다. 학습의 증가량에서도 자율학습과 토론 조합이 다른 조합에 비해 우월하였다. 다음으로, 고등학생을 대상으로 법학 컨텐츠를 사용하여 실험을 진행한 결과, 강의와 토론 조합과 비교해 자율학습과 토론 조합이 단기 및 장기적으로 암기와 전이 문제에서 더 높은 수행을 보였다. 요컨대 자율학습과 토론 조합은 다양한 학습

자료와 둘 이상의 연령 집단에 걸쳐 가장 높은 성과를 보였다.

제 3장에서는 토론에 교수자의 개입이 필요한지를 알아보기 위해, 토론 방식이 조작되었다. 토론 방식은 학생주도와 교사주도 두 수준으로, 토론의 준비 활동인 강의 혹은 자율학습과 조합되었다. 대학생으로 이루어진 네 집단의 사후 평가 결과를 비교한 결과, 자율학습 후 학생주도 토론, 자율학습 후 교사주도 토론, 강의 후 교사주도 토론, 그리고 강의 후 학생주도 토론 순으로 점수가 높았다. 그 원인을 파악하기 위해 토론의 대화 내용과 토론 중 관찰된 인지-행동적 몰입도를 분석하였다. 그 결과, 강의 집단에 비해 자율학습 집단은 토론에서 높은 몰입을 보였고, 그 몰입이 학생주도 토론에서는 높게 유지된 반면, 교사주도의 토론에서는 점차 떨어졌다. 반면, 강의 집단은 교사주도 토론에서 상대적으로 높게 몰입했지만, 학생주도 토론에서는 제대로 몰입하지 못했다. 이 결과는, 수업 목적에 따라 두 활동을 적절히 조합해야 할 필요가 있지만, 자율학습 후 학생 간 토론을 더 장려할 근거를 제공한다.

본 논문은 학습 과학 분야에서 두 가지 면에서 중요한 기여를 한다. 하나는 자율학습 후 학생주도의 토론이 다른 조합에 비해 높은 학습 효과를 보인다는 발견이다. 이 발견은 실제 수업에서 강의에 대한 효과적인 대안이라는 점에서 실용적 시사점이 크다. 본 연구의 또 다른 기여는 연구 방향성을 제시한다는 점이다. 본 연구에서는 자율학습과 토론 조합에 초점을 두었지만, 다른 가능한 조합들 예를 들면, 글쓰기와 토론, 혹은 자율학습, 글쓰기 그리고 토론이라는 삼중 조합이 학습에 미치는 영향을 탐색할 가능성을 제공한다. 이와 관련된 후속 연구가 학습과 교육의 미래에 새로운 빛을 제공하기를 기대한다.

주요어: 토의, 자율학습, 조합, 준비 학습 활동, 강의

학번: 2019-39236

Acknowledgements

It has been six years since I enrolled with vague hopes and dreams of becoming a researcher. It is not enough to simply say thank you to the many people who have helped me along the way. From the bottom of my heart, I am especially indebted to Professor Jooyong Park, who has been supportive of my career goals and worked actively to provide me with the protected academic time to pursue those goals. As my advisor and mentor, he has taught me more than I could ever give him credit for here. He has shown me what it means to be a good researcher and person. I am forever grateful for his guidance, encouragement, and unwavering support.

I am grateful to all of those with whom I have had the pleasure to work during this thesis and other related projects. Each of the members of my thesis committee, Professors Sowon Hahn, Byoung-Tak Zhang, Hee Seung Lee, and Soojeong Jeong, has provided me with extensive professional guidance and comments. I deeply appreciate their invaluable insights, expertise, and support throughout this journey.

In the pursuit of my chosen path, nobody has been more important to me than my family members, Hyojeong Lim, Jaehyeon Ban, and Jaejoon Ban. Most importantly, I would like to thank my parents, who have always been there for me with love and support. I am also deeply grateful to my amazing friends, Junsang Yu, Daewon Kim, Inkyeol Seong, Hyeonjin Yun, Kang Il Choi, Youngbeen Son, Haneul Kim, Jeonghyeon Kim, Woncher Im, and Wongi Kim, who have provided me with unending inspiration and support. A big thank you to Bongju Shin, Soul Han, Jaeho Oh, Jongwoo Lee, Chankyu Kim, Eun Taek Im, Kyung Gun Joo, Hakgyun Lee, Tae Geun Kim, and Deok Hyeon Hwang who generously gave their time and shared their experiences with me.

I am deeply grateful to Taehyeong Kim, Sungjae Cho, Youjin Hwang, Gi-Cheon Kang, Hwiyeol Jo, Yongmin Shin, Junho Lee, and May Jorella Lazaro who shared this journey with me. Their friendship, support, and encouragement have been invaluable to me.

In conclusion, I would also like to express my heartfelt gratitude to my laboratory colleagues, Jung Ae Park, Gyeongmee Gim, Songeui Kim, Jungyeon Park, Ji Won Yang, Min Hae Song, Sunlin Kwon, Hwimin Kim, Kahyun Kim and Hyeonmyeong Jang. Their diverse insights and perspectives have greatly enriched my work and helped me to improve it.

This year, I think not even a day has gone by without me taking a stomach relief medicine. This is how much stressed I was for the past six months or so, without even realizing it. This stress may have stemmed from my desire to write good papers, or from my concerns about my next career move, but I know that it is a major issue that I will have to continue to address in the future. Through this process, not only me, but other members of the lab as well, have realized how difficult studying and researching is. Although I am still quite unsure about what the future holds, if possible, I would like to continue to work in this field in the future. I will always be hoping to see you often as we come and go, and always praying for the health of our lab members and looking forward to their good research.