이학석사 학위논문

How Can We Make Better Use of Discussions in Education?

학습 상황에서 토론 효과 증진 방안

2019 년 8 월

서울대학교 대학원

협동과정 인지과학전공

임 재 서

How Can We Make Better Use of Discussions in Education?

학습 상황에서 토론 효과 증진 방안

지도교수 박 주 용

이 논문을 이학석사 학위논문으로 제출함

2019 년 8 월

서울대학교 대학원

협동과정 인지과학전공

임 재 서

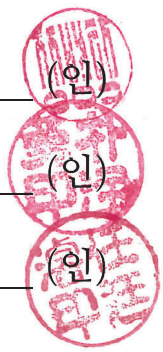
임재서의 이학석사 학위논문을 인준함

2019 년 7 월

위 원 장 장 병 탁 (인)

부위원장 박 주 용 (인)

위 원 임 정 준 (인)



Abstract

How Can We Make Better Use of Discussions in Education?

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

Although there is empirical evidence that students learn more when actively participating in learning, the current education is still lecture-centered. Even in science, technology, engineering, and mathematics, so-called STEM, where the amount of knowledge seems important, students learn more when they actively participate in classes than passively listening to a lecture. In this context, this study is conducted for two purposes. First, we seek to confirm that the discussion increases learning outcomes by comparing it with the review. The second purpose of the study is to find out whether the discussion effect is varied by the activity that precedes it. As a result of the experiment, among the groups who listened to the lecture, the discussion group performed better than the review group. In addition, among the groups in which the discussion was conducted, the learning outcomes of the groups discussed after self-study were superior to those discussed after the lecture. These results were due to self-study that made the discussion more plentiful. In conclusion, this study are important in that it provides educational and practical implications for how to change the current lecture-centered education.

Keywords: active study, discussion, education, lecture, self-study
Student Number: 2017-25197

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Limitations of Traditional Teaching

Education is still centered on lectures. The reason why teacher-centered style is so widespread is that people think not only that it is a more convenient way for both teachers and students, but also that providing more information is effective for learning. However, simply conveying a lot of information does not lead to learning and educational achievement. Based on a research comparing lectures and other teaching methods, Bligh (1972) argues that lectures are similar to other methods in conveying information, but are less effective than discussions in terms of promoting thought and less effective in changing attitudes or in acquiring skills. He added that it is professors' erroneous notion that lectures can draw the students' attention, stimulate their motivation, or improve academic level.

Hrepic, Zollman, and Rebello (2007) investigated the effect of video lecture. Experts and students responded to a set of questions, and they answered the same set of questions after watching a video lecture. The result showed that experts were able to learn new concepts after the lecture, while students were not able to for the complex concept questions. In response to the findings, Wieman and Perkins (2005) suggested that the amount of information might have been too overwhelming for students to digest. Poh, Swenson, and Picard

(2010) also supported this opinion. They had their participants to wear on their wrists a device called E4, a wristband which measures and records sympathetic nerve system's activity through skin conductive response. While they had the device on their wrist, they were told to record their activities regularly. By combining the activation level of nervous system and self-reported activity, they found that the activation level was high in doing homework, studying, or taking tests, but it was low in watching to a lecture, watching TV, or doing chores. The findings corresponds to Bligh's claim that it is illusionary to think lectures are effective for enhancing attention, increasing motivation, or improving academic level.

In addition, Carpenter and colleagues (2013, 2016) provided research findings that showed the eloquence of the lecturer is irrelevant to students' academic achievement. They had the same lecturer teach the same content once inarticulately and once eloquently. Two groups of students watched each video lecture and studied. They were asked to rate how much they think they learned, and were tested. The group who watched the eloquent lecture thought that they learned more than the other group, but the actual test result showed no difference between the two groups. Thus, the lecturer's eloquence made the student feel they learned a lot, yet it did not actually help students learn more. These studies show convergingly the limitation of traditional lecture-based class. With these in the background, it has been studied continuously to seek a new learning method to replace lectures.

1.2 Need for Interaction as an Alternative to Lectures

Various methods have been researched as alternatives to lectures. They are called active learning, and require students' cognitive intervention (Bonwell &

Eison, 1991). Active learning takes place through analyzing, integrating the given information, and testing over the content (Corno & Mandinach, 1983). Although much research was conducted comparing active learning and traditional classes, we can dichotomize them for the sake of discussion. One is a direct comparison of traditional lecture with active learning in actual class or laboratory settings (e.g., Hake, 1998; Stull & Mayer, 2007; Menekse et al., 2013; Freeman et al., 2014). The other line of research involves the debate on the effectiveness of problem-based learning (PBL) (e.g., Kirschner, Sweller, and Clark, 2006).

The two studies show findings that are too varied to conclude either method is better, and it does not seem likely to that a conclusion will be made any time soon. However, one can easily find successful cases of active learning. Let us take a look at two empirical studies on positive effect of active learning. Hake (1998) conducted a research comparing the effect of lecture-centered class and that using student response system (SRS). Utilizing SRS facilitates student participation since problems are offered during the class, the responses are collected through SRS, and they can be utilized right away. According to Hake, students from a lecture-centered class acquired only 30% of a new concept, while using SRS and thus increasing participation and interactivity helped students achieve 60%. A more surprising finding is that the same patterns are observed regardless of college professors' abilities, and the number of students. Freeman et al. (2014) conducted a meta-analysis on 225 studies comparing academic performance of traditional lecture-centered class and active learning in the field of science, technology, engineering, and mathematics (STEM). Academic performance was based on (1) equivalent test score, concept inventories, or other assessment, and on (2) the rate of D, F and withdrawal (DFW rate). As a result of the analysis, active learning showed higher performance level and lower DFW

rate than traditional lecturing. Even in STEM where the amount of information matters, active learning was more effective than traditional lecturing.

Still controversial is the debate on problem-based learning, having strong arguments from proponents and opponents. The reason for such controversy can be found in the analysis by Hung (2011). He compared interpretations of the problems used in problem-based learning, but match result was only 62% between the lecturer's intended academic goal and the students' interpretation of academic goal. Also he pointed out that the students' comprehension level, attitude toward learning, proficiency of lecturer, and small group interaction skills can have an influence on the effect of PBL. Thus, PBL can be successful only when lecturer and students agree on basic philosophy and assumptions, delicately design a problem, and investigate the procedures, going beyond of simply proposing a problem to the students. Hung's analysis implies that PBL is not a complete alternative, but something that needs to be refined.

Chi and colleagues (Chi, 2009; Menekse et al., 2013; Chi & Wylie, 2014) pointed out that active learning deals with an excessively broad field since it includes all the teaching methods except for lecturing. ICAP framework is proposed as a framework for specification. This framework categorizes based on learner's behavioral modes. Passive mode signifies learner is sitting still in a traditional lecture, and other three modes are specified modes from the active learning. Active mode emphasizes students' physically manipulating the information without adding new knowledge. For example, it consists of underlining, copying problem solution method. Constructive modes refers to students building knowledge more than the given information such as by drawing a table, explaining, asking questions, and coming up with solutions. Lastly, the interactive modes refers to two or more colleagues collaborating with each other. These co-constructing activities occur through asking questions and having a

conversation. Interactive mode allows students clarify each person's thoughts, and to create counter-statement against the other side, and it is constructive and productive (Berkowitz & Gibbs, 1982). Thus, Chi and Wylie (2014) concluded that academic achievement was the lowest at P, then A, C, and I in the ascending order.

Menekse et al. (2013) conducted an experiment where college students were told to study in four different ways and their final scores were compared at the end. At first, all of the participants in four groups read a two page long text. And then the passive learning group read additional eight page long learning material, while the active learning group read as they underlined the important parts. Constructive learning group interpreted a graph and a table with a relevant work sheet. Interactive learning group were given graph and table with a work sheet as they coupled to discuss and interpret. After each group studied in their own way, they were tested over the same content. As ICAP model predicted, passive group, active group, constructive group, and interactive group scored higher in the ascending order.

Another evidence that discussion facilitates learning is found in an experimental study comparing learning from an online unilateral lecture to learning from an online conversational video (Daradoumis, Bassi, Xhafa, & Cabellé, 2013; Hew & Cheung, 2014). Two groups of participants watched two different videos, studied with a workbook, and were tested over the content. Result showed the ones who watched the conversational video scored significantly higher than the others did. Moreover, they scored higher when they watched the video with someone else and discussed together than alone (Craig, Chi, & VanLehn, 2009; Craig, Driscoll, & Gholson, 2004; Driscoll, Craig, Gholson, Hu, & Graesser, 2003; Muller, Bewes, Sharma, & Reimann, 2008; Muller, Sharma, Eklund, & Reimann, 2007). In short, students learn more when the lecture

format is conversational and when the students interact with each other.

1.3 Present Study

Although discussions can generally facilitate learning, as Murphy and others' meta-analysis (2009) have proven, discussion alone does not guarantee a successful result. It requires prior preparation to activate basic knowledge about the content as well as to facilitate interactivity among students. Students can be asked to come up with questions before discussion in order to activate basic knowledge. More research must be done to examine what activities should precede discussions and also how discussion should be carried out in order to increase learning outcome through discussions.

This study seeks to confirm if discussions improve learning and to find out what preliminary activity can facilitate learning through discussions. In order to confirm if discussions actually improve learning, learning achievement after discussions is compared with that after review activity. Specifically, two groups watched a video lecture. Then, students in one group reviewed by themselves, while those in the other discussed together. In effect, the review group resembles a typical class environment. In line with previous literature, we hypothesized that the discussion group would score higher than the review group.

Next, to find out what preliminary activities improve academic outcome through discussions, video lectures and self-study method were compared. We hypothesized that self-study will produce a better result than simply watching the video lecture. Although Chi and Wylie categorized both reading and viewing a video as passive mode, sympathetic nerve system activity was measured higher during self-studying than during watching video lectures according to Poh. Therefore, even though they may be both passive modes, video-watching

was likely to be relatively more passive, thereby result in a worse academic outcome. With these two objectives in mind, the following two experiments were conducted.

Chapter 2

Experiment 1

Experiment 1 seeks to find out which activity produces more outcome after watching a video lecture: discussion or reviewing. In addition, it seeks to find out if watching a video and self-studying before discussion create a different result in the posttest. Participants were divided into three groups: reviewing after video-watching (Lecture Review: LR), discussing after video-watching (Lecture Discussion: LD), and discussing after self-studying (Self-study Discussion: SD). We compared three groups of posttest scores.

2.1 Methodology

Participants and design. Seoul National University undergraduate students participated in this experiment for course credits. 90 students participated but 2 left during the experiment, thus leaving 37 men and 51 women. The participants were randomly assigned to three different groups: LR ($N = 30$), LD ($N = 29$, number of discussion groups ($G = 9$), and SD ($N = 29$, $G = 9$). Three or four students formed a discussion group.

Background knowledge survey. Background knowledge (Beyer, 1987; Miyake & Norman, 1979) and interest (Cuccio-Schirripa & Steiner, 2000) can have an influence on questions and the quality of questions in an academic setting. Thus,

we identified background knowledge level that may have an effect on the experimental results. The participants were tested over two topics related with the learning information and four different topics that is irrelevant with. The total correct score is 7 points over 6 topics.

Test questions. Seven pages of learning material were given during the learning period. The test questions could be divided into three categories: (1) Rote memory type questions. That is multiple choice and short answers format. Students had to simply memorize given information for they were asked straightforward information from the material. They are ten points for 10 questions. (2) Paraphrased type questions. These consisted of items with similar concepts in materials. Students had to explain concepts from the given material. Twenty-two points for 6 questions. (3) Transfer type questions. These items required students to utilize their understanding the whole content and apply it in new situations. Eighteen points for 4 questions (i.e., questions with dissimilar scenarios). We considered the deeper transfer type questions.

Content covered in the video lecture. The instructional video lecture used in the experiment corresponded to lecture-style monologue sessions. The videos were lectures on law available on the university web site. The lecture dealt with accusation, charge, and recognition as criminal procedure code. The subject matter was chosen because it was less likely to be affected by background knowledge since study of law is not available as an undergraduate course, and because it can produce simple memory type questions, similar concepts questions, and deeper transfer type questions. The video was 18 minutes long.

Procedure. First, the participants took a background knowledge survey. They were to rate their knowledge level from 1 (no knowledge) to 7 (expert knowl-

edge) for 6 categories including criminal procedure code, accusation, and charge. The LR group viewed the video lecture and self-studied for 18 minutes with the provided learning material. The LD group viewed the video lecture and discussed for 18 minutes in groups consisting of 3 to 4 students. Lastly, the SD group discussed after self-studied the learning material. All three groups then took 20-minute test.

2.2 Results and Discussion

Measurement model. A total of 90 students participated in Experiment 1. Thirty were assigned to each LR, LD, and SD group. However, LD and SD groups each had one participant leaving the experiment, thus leaving only 88 participants for the experiment. The average and standard deviation are provided in Figure 2.1. Analysis of variance (ANOVA) was conducted to examine the differences between groups. The groups were compared over their analysis of the total scores from the test consisted of rote memory, paraphrased, and transfer type questions. First, there were no significant differences in age among the groups. LR group ($M = 20.47$, $SD = 1.94$), LD group ($M = 20.31$, $SD = 2.05$), and SD group ($M = 19.69$, $SD = 1.53$), $F(2, 85) = 1.43$, $p = .245$. Next, the background knowledge survey ranged from 1 (no knowledge) to 7 (expert knowledge). The participants who marked 5 (know a little) or above on criminal procedure code, accusation, and charge sections were ruled out from the experiment. The average score of criminal procedure code section was the following: LR group ($M = 1.55$, $SD = .69$), LD group ($M = 1.47$, $SD = .90$), and SD group ($M = 1.82$, $SD = 1.03$). Almost all of them nearly had no relevant knowledge, and there was no significant difference, $F(2, 85) = 1.33$, $p = .270$. In addition, the average of accusation and charge sections was the

following: LR group ($M = 2.40$, $SD = 1.07$), LD group ($M = 2.00$, $SD = 1.00$), and SD group ($M = 2.34$, $SD = 1.08$). They showed little knowledge in these sections as well and had no significant difference, $F(2, 85) = 1.25$, $p = .293$.

Impact of condition on learning. As proposed by Vacha-Haase and Thompson (2004), we reported the effect size on major effect. Cohen (1988) suggested the following guideline on effect size: as small when $\eta^2 = .01$ or $d = .02$, medium when $\eta^2 = .06$ or $d = .05$, and large when $\eta^2 = .14$ or $d = .08$. Now to find out in earnest if there is a difference between groups in academic performance (learning outcome), variance analysis was conducted. As a result, significant major effect was found between groups in rote memory, transfer type questions, and total scores. Total scores ($F(2, 85) = 10.01$, $p = .000$, $\eta^2 = .19$), and transfer type questions ($F(2, 85) = 13.86$, $p = .000$, $\eta^2 = .25$) can be said to have a large effect size, while rote memory type questions ($F(2, 85) = 4.82$, $p = .010$, $\eta^2 = .10$) had a medium level of effect size.

And based on the hypothesis, mutual comparison was conducted between groups through orthogonal planned comparisons. As a result of analyzing the total score in order to examine the overall trend, SD group ($M = 38.41$, $SD = 4.74$) had a significantly higher score than LD group ($M = 34.62$, $SD = 3.92$), $t(85) = 3.26$, $p = .012$, $d = .88$, and LR group ($M = 33.47$, $SD = 4.58$), $t(85) = 4.29$, $p = .000$, $d = 1.05$. According to Cohen's standard, the difference between these groups have large effect size in difference. But, the difference LR and LD had insignificant difference, $t(85) = 1.00$, $p = .320$, $d = .26$.

For the rote memory type questions, SD group ($M = 7.35$, $SD = 1.20$) scored significantly higher than LD group ($M = 6.35$, $SD = 1.29$), $t(85) = 2.76$, $p = .007$, $d = .64$, and LR group ($M = 6.40$, $SD = 1.61$), $t(85) = 2.63$, $p = .010$, $d = .46$. The difference between these groups is medium. But LR and LD groups

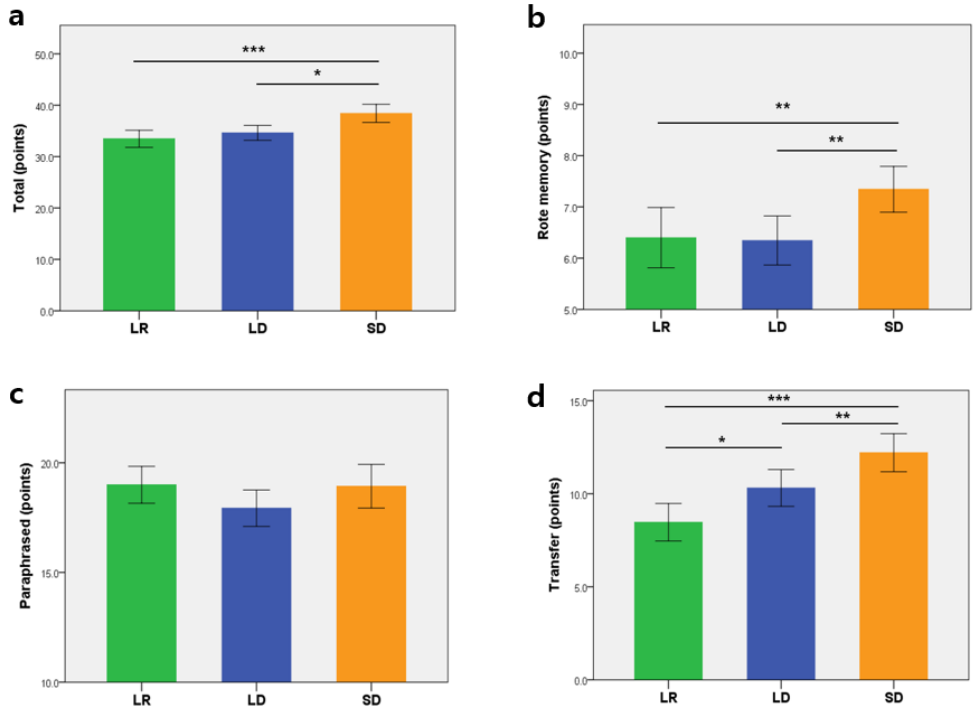


Figure 2.1: LR, LD, and SD groups and their scores for all types of questions: total, rote memory, paraphrased, and transfer type questions. (**a**) Three groups and their total score; $^{*}p < .05$, $^{***}p < .001$. (**b**) Three groups and their scores for rote memory type questions; $^{**}p < .01$. (**c**) Three groups and their scores for paraphrased questions; *n.s.* (**d**) Three groups and their scores for transfer questions; $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. Error bars are $\pm 2SE$.

showed no significant difference, $t(85) = .153$, $p = .879$, $d = .20$.

For the paraphrased type questions, all three groups; SD ($M = 18.93$, $SD = 2.69$), LD ($M = 19.00$, $SD = 2.30$), and LR ($M = 17.93$, $SD = 2.24$) showed no significant difference, $p > .05$.

For the transfer type questions, the three groups shows significant difference. SD group ($M = 12.21$, $SD = 2.76$) scored higher than LD group ($M = 10.31$,

$SD = 2.76$), $t(85) = 2.65$, $p = .010$, $d = .69$, and LR group ($M = 8.47$, $SD = 2.76$), $t(85) = 5.27$, $p = .000$, $d = 1.32$. In addition, LD group was found to have scored significantly higher than LR group, $t(85) = 2.60$, $p = .012$, $d = .65$. That is, the difference of size effect between all three groups is large.

The result so far supports the hypothesis of experiment. First, in correspondence with the hypothesis that the discussing together will result in a higher learning outcome than reviewing after the lecture, the discussing group; SD, and LD groups scored superbly higher than the other group in the posttest. Also corresponding with the result of Menekse et al. (2013), discussion after the lecture which is an interactive mode resulted in a higher academic effect than reviewing alone which is active or constructive mode. The two findings had a different methodology to confirm the effect of discussion, yet they show a positive influence of discussion in a convergent manner.

Next, in correspondent with the second hypothesis, self-studying improved the academic performance through discussion more than watching a lecture before discussion. The findings are in correspondence with ICAP hypothesis in that self-studying scored higher than watching a lecture possibly, because the former involves active or constructive activity such as underlining or summarizing important concepts while the latter is passive. Especially, transfer type items showed a large difference, it sheds a large practical implication, considering that transfer is challenging to achieve (Muldner, Lam, & Chi, 2013). This means students need more opportunities for self-regulated learning.

However, there could arise a counter-argument regarding the second hypothesis in Experiment 1 findings. The possible argument could say that the reason why the group of discussing after self-studying scored higher on the test than the group of discussing after a lecture is not that discussion was involved, but that self-studying was more effective than watching to a lecture. In other

words, the difference in the test scores between the two groups has nothing to do with discussion, but it is due to the learning level's difference before having a discussion. In order to confirm this possibility, Experiment 2 was conducted.

Chapter 3

Experiment 2

Experiment 2 was conducted to examine partial findings in Experiment 1. That is, to confirm that the reason the SD group scored higher than the LD group was because discussion was involved, not because the preliminary activity was different before having the discussion. In order to test the possibility that the difference in learning outcome occurred simply because of the different preliminary activity, a test was given after the activity (self-study or lecture). Lack of significant difference would support our hypothesis that there is interaction between the preliminary activity and subsequent discussion.

3.1 Methodology

Participants, materials and procedure. Seoul National University undergraduate students who took a psychology lecture participated in this experiment for course credits. Total participants were 52 consisted of 38 men and 12 women. Participants were randomly assigned to each group in half: Lecture group (Lecture: L), and Self-studying (Self-study: S).

Except for taking a survey during discussion, other procedures were the same as Experiment 1. The survey was about image evaluation, irrelevant to learning information, which lasted about 15 minutes.

3.2 Results and Discussion

Measurement model. Fifty-two students participated in Experiment 2. Two of them were excluded, with 50 used for the analysis. Twenty-five were assigned to each group of the lecture, and self-studying. The average and standard deviation are provided in Figure 3.1. T verification was conducted to examine the differences between groups. The groups were compared over their analysis of the total scores from the test consisted of rote memory, paraphrased, and transfer-type questions. First, there were no significant differences in age among the groups. First after analyzing prior knowledge, the average score of criminal procedure codes for each group is the following: L group ($M = 1.56$, $SD = .71$) and S group ($M = 1.92$, $SD = 1.00$) showed little relevant knowledge and had no significant difference, $t(48) = 1.47$, $p = .148$. In addition, the average score of accusation and charge is the following: L group ($M = 2.00$, $SD = .91$) and S group ($M = 2.40$, $SD = .87$) showed little knowledge as well and no significant difference, $t(48) = 1.59$, $p = .119$.

Impact of condition on learning. Like the Experiment 1, we reported on the effect size on major effects. As expected, L group and S group showed no significant difference in total scores, $t(48) = .00$, $p = 1.00$, $d = .00$. And there was no significant difference in rote memory type questions, $t(48) = 1.49$, $p = .143$, $d = .39$, paraphrased type questions, $t(48) = .84$, $p = .407$, $d = .23$, and transfer type questions, $t(48) = .73$, $p = .732$, $d = .01$. Thus, all the result shows there was no significant difference.

We carried out a post-analysis using Tukey in order to analyze the difference between the previous groups in Experiment 1: LR, LD, and SD, and two new groups of L, and S. There was no significant difference between L group and S group in total scores, rote memory, paraphrased, and transfer type questions.

However, the total scores of L and S groups had significantly lower than the three groups of Experiment 1, $p = .000$. L group scored significantly lower than all the other groups in rote memory type questions, $p < .01$, S group scored only lower than SD group, $p = .000$. L and S groups scored significantly lower than the other three groups in paraphrased type questions, $p < .05$. Interestingly, they scored significantly lower than the two discussing groups, $p < .01$, but showed no significant difference from LR group.

Experiment 2 started from the possibility that a better learning outcome comes from self-studying than from watching to a lecture, regardless of having a discussion. Unlike the Experiment 1, it compared the L and S groups, taking discussion out of the equation. As a result, total score of the two groups showed no significant difference, and moreover there was no significant difference in all types of questions even including transfer type questions. That is, the different outcome from LD and SD groups was not because of the preliminary activity's learning level difference, but because of the discussion accompanied together. The findings imply that the group of SD scored higher than that of LD in the final test not because of different preliminary activities but because of interactivity with discussion. Active and self-regulated study can improve academic performance even if having a same discussion.

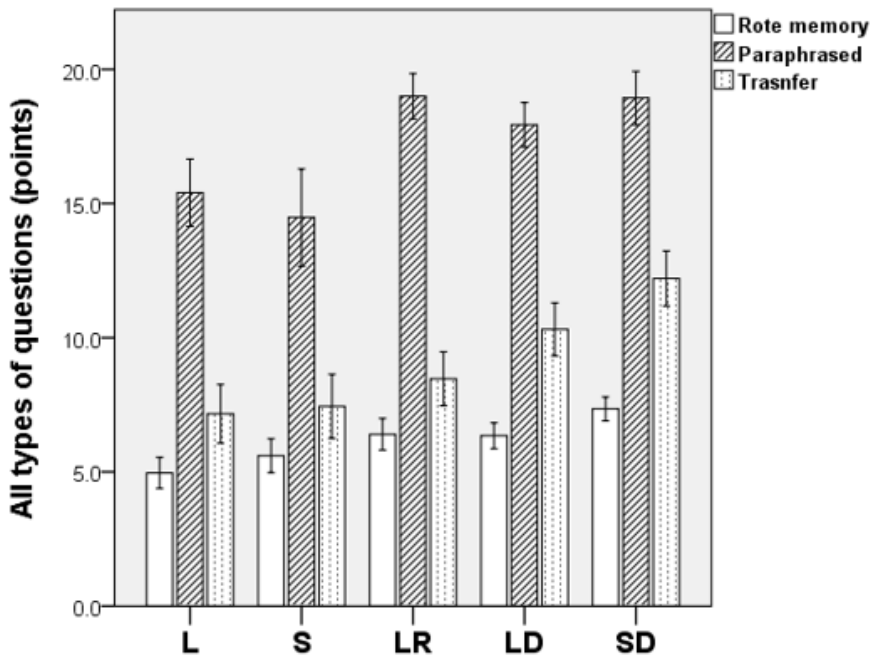


Figure 3.1: L, S, LR, LD, and SD groups with each scores for all types of questions: total, rote memory, paraphrased, and transfer type questions. Error bars are $\pm 2SE$.

Chapter 4

Experiment 3

Through Experiments 1 and 2, it was confirmed that rather than review discussions, and rather than lecture before discussions, self-studying increased learning. In Experiment 3, by repeatedly verifying the results of Experiment 1 to other participants, to find out why performance increased when self-study comes first compared to lecture, the discussions content was analyzed. For analysis, the method used by Chi et al. (2017) was applied. They divided the content of the conversations to C and I, and each actions applicable were set to substantive comments, interaction episodes and co-constructive turns. Given the considerations above, we used same video lecture as earlier Experiment 1, thereby ensuring that our results would not be tied to an individual participants and increasing the generalizability of our findings.

4.1 Methodology

Participants, materials and procedure. Seoul National University undergraduate students who took a psychology lecture participated in this experiment for course credits. Total participants were 74 and consisted of 32 men and 42 women. The participants were randomly assigned to LD group ($N = 37$, $G = 10$) and SD group ($N = 37$, $G = 10$).

Except for LR groups, they were the same as experiment 1 procedures.

4.2 Results and Discussion

Measurement model. There was no significant difference of LD group (N = 37, 19 men) age (M = 20.30, SD = 1.47), and SD group (N = 37, 13 men) age (M = 20.54, SD = 2.30), $t(72) = .54$, $p = .590$. The average regarding the code of criminal procedure provision was all unknown in terms of LD group (M = 1.97, SD = .90), SD group (M = 1.76, SD = .93) and there was no significant difference between groups, $t(72) = 1.02$, $p = .311$. Moreover, the average regarding indictment and charge provisions was all unknown in terms of LD group ($M = 2.35$, $SD = .98$), and SD group ($M = 2.24$, $SD = 1.04$) and there was no significant difference, $t(72) = .46$, $p = .646$.

Impact of condition on learning. Having found out if there was difference between groups in the final test, there was significant difference between groups in the rote memory type, transfer type, and total score. Firstly, it can be said that the overall score had a large effect size, $t(72) = 3.57$, $p = .001$, $d = .83$. Next, rote memory types questions, $t(72) = 3.65$, $p = .000$, $d = .85$. Transfer type questions had large effect size too, $t(72) = 4.16$, $p = .000$, $d = .97$. However, paraphrased type questions did not gave significant difference between groups, $t(72) = .19$, $p = .851$, $d = .05$. This was shown in the Figure 4.1 below.

The above results, like the results of the Experiment 1, show that the SD group had better performance compared to the LD group. Especially, it had very similar results with Experiment 1 in that it showed significant difference in total score, rote memory type questions, and transfer type questions. This means the analyses presented here further illustrate how ICAP can guide the design of new learning environment as well as provide evidence in support of ICAP's assumption that higher modes of engagement, the constructive and interactive modes, are responsible for deeper learning, as reflected in the transfer questions

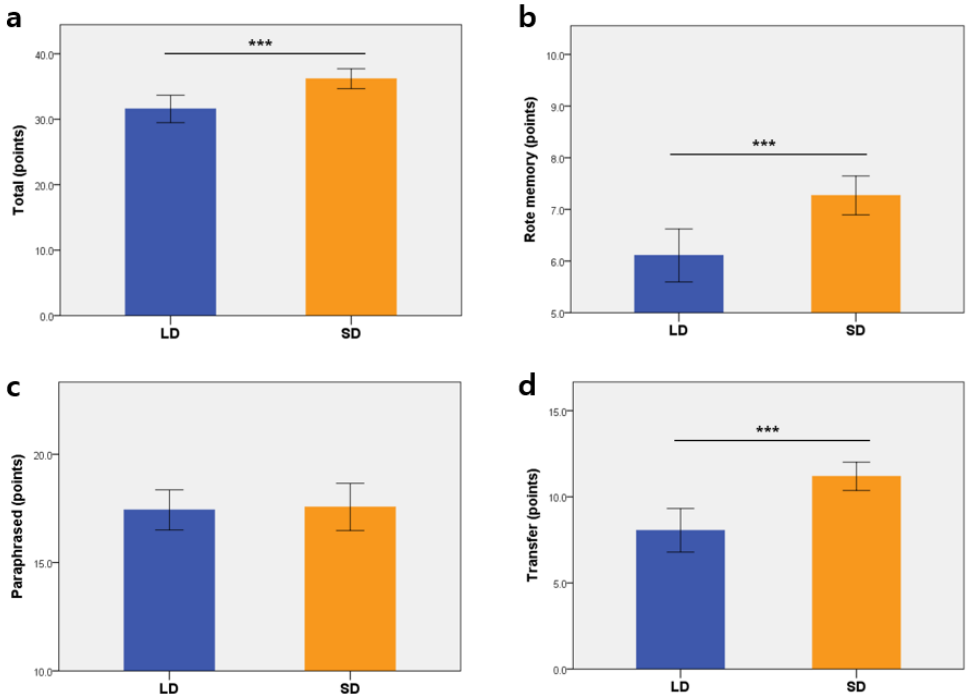


Figure 4.1: LD and SD groups and their scores for all types of question: total, rote memory, paraphrased, and transfer type questions. (**a**) Two groups and their total score; $^{***}p = .001$. (**b**) Two groups and their scores for rote memory type questions; $^{***}p < .001$. (**c**) Three groups and their scores for paraphrased questions; *n.s.* (**d**) Three groups and their scores for transfer questions; $^{***}p < .001$. Error bars are $\pm 2SE$.

we used in our assessment of learning.

To continue, we investigated why the results from the experiment would show such. The conversations in discussions had been transcribed for analyzes in the experiment. Thus, we analyzed the groups' conversations while participating in discussions, which were segmented into statements, and coded (Chi et al., 2017; Muldner et al., 2014).

4.3 Why Is Self-Study Before Discussions More Effective?

This analysis analyzed discussions of each group through the coding methods used in Chi et al. (2017). Transcription was conducted for discussions analysis and the reliability with other assessors was calculated. There were 10 transcripts per group.

Constructive: substantive comments. First, for the standard to assess the constructiveness of discussions, we assessed the substantiveness of the comments on the discussions conversation (Muldner et al, 2014; Chi et al., 2017). The sentences were divided into substantive comments and those, which were not. Substantive comments refer to questions and conversations related to concepts of the lecture. Yet, the right and wrong of comments were not important and the relevance of comments and study concept was important. Simultaneously, if those comments remind other people of those concepts, or if it induced questions, it was referred to as substantive. For example, substantive/relevant comments at the phrase level are as follows: "What exactly is offence subject to complaint? Does it mean punishment is possible if there are charges?" (Related to concept of offence subject to complaint), "Then if the relative is the suspect, is that the case? 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 independent victim?"(Related to legal representative concept of victim). In contrast, non-substantive comments were either comments irrelevant to the topic of materials or comments such as "yes", "I don't' know.", or "I think so." As a result, the SD group (37.80 Total per discussion group), had significantly high frequency of substantive comments compared to the LD

group (15.20) and this had large effect size ($p = .020$, $d = 1.18$). Interrater reliability for this coding was conducted with other assessors ($\kappa = .95$ for 20% of the transcripts). This data is in the first line of Table 1.

Interactive: episode and turn. The conversation transcripts of each discussions group was analyzed to know how additionally interactive it was. As previously, using the transcripts of each discussions group conversation, we segmented the episodes. An episode was a portion of the in-group members' multi-turn conversation on the same topic and line of thought that included at least one substantive comment from either member of the group, with substantive comments as defined in the preceding analysis. Yet, the episode of main concepts of the lectures is the 12 classified in Appendix 1.

Each episode was additionally named as interaction episode and this when 2 or more members in the group make at least one substantive comment. For example, in the interaction episode 1 of Appendix 2, Female 1 made a substantive comment about the offense subject to complaint and the male 1 too made a substantive comment that urges explanation and questions about the offense subject to complaint. Ultimately, as there was substantive comment consecutively within one episode, this can be called interaction episode. Meanwhile, if like 1 non- interaction episode 1, female 2 and female 1 made comments that were not substantive, it is not classified as interaction episode. Ultimately, episode with 1 or more substantive comment was seen as interaction episode. Using this coding rubric, seeing the interaction episode by discussions group, SD group had 5.50 per discussions and LD group had 3.80 per discussions, to not have significant difference, $p = .21$, $d = .59$ Interrater reliability for this coding was conducted with other assessors ($\kappa = .95$ for 20% of the transcripts). Although the average number of interaction episode of the two groups did not

have significant difference, it probably results from the number of episodes of lecture data and limited discussion time. Therefore, this suggests that perhaps a finer grained analysis of interaction quality was needed to reveal difference between the LD group and SD group. Accordingly, the richness of the interactions was captured in the following more fine-grained ways that reflect their co-constructive quality.

Moreover, we conducted an additional 2 analyses and first number of substantive comments in interaction episode was analyzed. This showed more substantive comments, interaction episode has more content-relevant information and it can be seen as having a richer discussion. Having analyzed this, substantive comments per interaction episode had 6.85 per SD group and 2.43 of LD group, being more significantly higher and having a very large effect size, $p < .01$, $d = 1.60$. Therefore, at the interaction episode, the SD group is more constructive overall in their interaction.

Next, from the ICAP framework, we expected that there will be more co-constructive turns at interaction episode. This means the more co-constructive turns at interaction episode, the discussion becomes richer. This hypothesis was based on the idea that co-constructing generally means that each member of group builds on, refines, or challenges the ideas of the other member (Chi & Wylie, 2014). Hence, we set the frequency of turn based on the co-construction and this also contributed to consecutive substantive comments.

A turn is defined as only change in speakers. However, the co-constructive turn we set is set as more than 2 speakers from substantive contribution rather than only change in speaker. For example, in interaction episode 2 of Appendix 2, as male 2 urged questions regarding the offense subject to complaint, this is substantive comment. Thus, male 1 made a substantive comment that explains the offense subject to complaint and this becomes sequential turn of substantive

comment. Ultimately, as substantive comment after substantive comment was formed, this turn becomes co-constructive turn. Hence, a turn in which both speakers contribute substantive comments is one co-constructive turn.

Per interaction episode, the number of co-constructive turn taken by SD group exceeded the number of co-constructive turns taken by LD group (4.45 vs. 1.79 per interaction episode, respectively). It also has large effect size, $p = .03$, $d = 1.10$.

To summarize, although overall frequencies of interactions did not differ significantly between LD group and SD group, closer inspections of the quality of the interactions revealed that SD group interacted in richer way in that they generated more substantive comments and engaged in a greater number of co-constructive turns than LD group. The patterns of total results as such is more interactive more students for SD group than LD group, it shows they are have constructive discussions. Therefore, the more activity in the process before discussions for students, the conversation within discussions are constructive and interactive.

Exploratory regression analyses. The above analyses and interpretation suggest that substantive comments may have direct impact on outcome of transfer type items, since so many of the difference in number between SD group and LD group. To substantiate this interpretation, we analyzed the relationship of substantive contributions and academic performance of transfer type questions for SD group versus, LD group more closely, by running an exploratory linear regression, with average in-group member's score of transfer type item as the dependent variable and substantive comments as explanatory variables. Because we were interested primarily in analyzing how being substantive comments influenced learning in LD group and SD group. Consequently, the effect of sub-

stantive comments on transfer type questions was significant, $F(1, 18) = 5.38$, $p = .030$, $R^2 = .23$.

Given that the generation of substantive contributions has been shown to be associated with learning (Chi et al., 2008), this data provides a possible explanation for our previous results. Overall, for Experiment 1, the students of SD group conditions had higher significant outcome compared to LD group and especially it was well shown in the transfer problem. In Experiment 2, the SD group, had more significant substantive comments compared to LD group and at each interaction episode, there were more substantive comments and co-constructive turns. Moreover, the substantive comments that symbolize constructive activity caused significant effect on the transfer type problems. Therefore, generation of substantive comments may also explain the superior learning of the participants in SD group in this experiment compared to the participants in LD group, since the pattern of learning and the pattern of substantive contributions correspond.

Table 4.1: Mean frequencies (SD) of in-group members' engagement behaviors per discussion group after lecture and after self-study.

Behavior	**Discussion type**				
Conversations	After lecture		After self-study	p	d
Frequency of substantive comments per discussion group	15.20 (10.34)	$<$	37.80 (25.13)	.02	1.18
Frequency of interaction episodes per discussion group	3.80 (3.01)	$=$	5.50 (2.76)	.21	.59
Substantive comments per interaction episode	2.43 (1.78)	$<$	6.85 (3.47)	$<$ .01	1.60
Co-constructive turns per interaction episode	1.79 (2.30)	$<$	4.45 (2.53)	.03	1.10

Chapter 5

General Discussion

In education, although students are to be the major role and teachers the supporting role, lecture-centered classes are widespread. This is because lecture is the most familiar and convenient form of learning for both students and teachers. However, numerous studies have proven that lecture is not the best way of teaching method. Not that lecture is useless in and of itself, but that lecture-centered teaching should be avoided. The lecture-only attitude should be avoided, and activities where students can actively participate should be more encouraged. One of them is discussion.

This study conducted three experiments in order to empirically confirm the effect of discussion. The first experiment had three conditions compared: reviewing after the lecture, discussing after the lecture, and discussing after self-studying. To summarize the two findings from the first experiment, the group of discussing after the lecture scored higher than that of reviewing after the lecture, and the group of discussing after self-studying scored higher than that of discussing after the lecture. Experiment 2 reviewed the possibility that preliminary activities, not discussion, could cause different outcomes. The lecture group and self-studying group took the final test without having a discussion, and they showed no significant difference. Also reviewing after the lecture had no influence on transfer, thus showing that current teaching method

is ineffective.

The findings so far correspond with ICAP model of Chi & Wylie. That is the learning outcome was better in the interactive mode of discussion than in the active, constructive mode of reviewing. Taking a step further than ICAP model, we confirmed that even in the interactive mode of discussion, the preliminary activity could cause a different result. The findings were confirmed by having one condition of integrating lecture and discussion and another of integrating self-studying and discussion. Through this procedure, even though there was no difference between watching a lecture and self-studying, the learning outcome was better through follow-on discussion. Especially, it is important that transfer-type item showed a significant difference. According to Goldstone & Day (2012), although transfer is difficult to achieve, it is important in the practical perspective to use a learned knowledge into a new situation. Thus, knowledge-change process from interactive activities facilitated transfer through learner's deep comprehension of learning content.

Although more various groups and contents have to be dealt to generalize, it is important that it sheds a practical implication of how to change the current lecture-centered teaching method. The current educational emphasis is on teacher-centered lecturing followed by student's reviewing. The findings of this study suggests changing to methodology where students study by themselves and then gather to have a discussion. Lecture will have its place when it is necessary, but lecture time should be minimized, while student's discussion time should be lengthened. Follow-on studies should be conducted to determine how much of discussion is necessary depending on learner's characteristics.

It needs to be pointed out that the students had discussion without teacher's interruption in this study. Many teachers and professors can easily think discussions during lecture should be centered on themselves. However, they end up

not having discussion when there are too many students to control the discussion. However, this study confirmed discussion proved to have better learning outcome than reviewing, even though discussion was done by the students who share the same preliminary activity. Thus, it needs to be reviewed to activate discussion between students while minimizing teacher's interruption. For example, we can start from letting the students actually have a discussion by securing their discussion time and minimizing lecture time.

The participants were college students. Future studies could involve younger students to research if discussion is effective and how to maximize the effect of discussion. One methodology is to provide a type of scaffolding to students. For example, it is to lead the students to leave substantive comment in discussion by implanting a tool where students can reflect on textbooks or videos (Ogan, Aleven, & Jones, 2009; Van Blankenstein, Dolmans, Van der Vleuten & Schmidt, 2011).

In addition, research could be done to confirm why self-studying before discussion is effective. Students can use the methodology to measure discussion. During self-studying, they can have relatively more time of self-reflection, which will help them have more substantive comments during discussion. Lecture-combined discussion already provides decisive knowledge to students, which will give relatively lesser time to think deeply, and consequently minimizing substantive comments in discussion. Instead of watching to a lecture, self-studying before discussion can help students share more co-constructive turns including substantive comments in the process where they share what they do not know. Therefore, instead of feeling knowledgeable from a lecture, pondering and thinking deeply through self-studying will likely to result in a better discussion.

Lastly, just like a current lecture-type class, the video used in the experiment had one professor relaying information unilaterally. Likewise, most online teach-

ers nowadays offer monologue-type lectures (Caspi, Gorsky, & Privman, 2005; Zhang, Zhou, Briggs, & Nunamaker, 2006). However, future studies could have a conversational video instead of monologue-type lecture. An empirical study is necessary since it is highly possible to have a more effective result through the new type of video than the old video.

The above-mentioned possible follow-on studies are expected to make a new difference in the future education.

References

Baumeister, R. F., Vohs, K. D., Aaker, J. L., & Garbinsky, E. N. (2013). Some key differences between a happy life and a meaningful life. *The journal of positive psychology*, *8*(6), 505–516.

Berkowitz, M. W., & Gibbs, J. C. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly (1982-)*, 399–410.

Beyer, B. K. (1987). *Practical strategies for the teaching of thinking.* ERIC.

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

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom. 1991 ashe-eric higher education reports.* ERIC.

Carpenter, S. K., Mickes, L., Rahman, S., & Fernandez, C. (2016). The effect of instructor fluency on students' perceptions of instructors, confidence in learning, and actual learning. *Journal of Experimental Psychology: Applied*, *22*(2), 161.

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*(6), 1350–1356.

Caspi, A., Gorsky, P., & Privman, M. (2005). Viewing comprehension: Students' learning preferences and strategies when studying from video. *Instructional Science*, *33*(1), 31–47.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework

for differentiating learning activities. *Topics in cognitive science*, *1*(1), 73–105.

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.

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.

Chi, M. T., & Wylie, R. (2014). The icap framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219–243.

Clark, D. B., Stegmann, K., Weinberger, A., Menekse, M., & Erkens, G. (2007). Technology-enhanced learning environments to support students' argumentation. In *Argumentation in science education* (pp. 217–243). Springer.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Corno, L., & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational psychologist*, *18*(2), 88–108.

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.

Craig, S. D., Driscoll, D. M., & Gholson, B. (2004). Constructing knowledge from dialog in an intelligent tutoring system: Interactive learning, vicarious learning, and pedagogical agents. *Journal of Educational Multimedia and Hypermedia*, *13*(2), 163–183.

Cuccio-Schirripa, S., & Steiner, H. E. (2000). Enhancement and analysis of

science question level for middle school students. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *37*(2), 210–224.

Daradoumis, T., Bassi, R., Xhafa, F., & Caballé, S. (2013). A review on massive e-learning (mooc) design, delivery and assessment. In *2013 eighth international conference on p2p, parallel, grid, cloud and internet computing* (pp. 208–213).

Driscoll, D. M., Craig, S. D., Gholson, B., Ventura, M., Hu, X., & Graesser, A. C. (2003). Vicarious learning: Effects of overhearing dialog and monologue-like discourse in a virtual tutoring session. *Journal of Educational Computing Research*, *29*(4), 431–450.

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.

Goldstone, R. L., & Day, S. B. (2012). Introduction to "new conceptualizations of transfer of learning". *Educational Psychologist*, *47*(3), 149–152.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, *66*(1), 64–74.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (moocs): Motivations and challenges. *Educational research review*, *12*, 45–58.

Hrepic, Z., Zollman, D. A., & Rebello, N. S. (2007). Comparing students' and experts' understanding of the content of a lecture. *Journal of Science Education and Technology*, *16*(3), 213–224.

Hung, W. (2011). Theory to reality: A few issues in implementing problem-

based learning. *Educational Technology Research and Development*, *59*(4), 529–552.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, *41*(2), 75–86.

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.

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.

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

Muller, D. A., Bewes, J., Sharma, M. D., & Reimann, P. (2008). Saying the wrong thing: Improving learning with multimedia by including misconceptions. *Journal of Computer Assisted Learning*, *24*(2), 144–155.

Muller, D. A., Sharma, M. D., Eklund, J., & Reimann, P. (2007). Conceptual change through vicarious learning in an authentic physics setting. *Instructional Science*, *35*(6), 519–533.

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.

Ogan, A., Aleven, V., & Jones, C. (2009). Advancing development of intercultural competence through supporting predictions in narrative video. *International Journal of Artificial Intelligence in Education*, *19*(3), 267–

288.

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.

Stull, A. T., & Mayer, R. E. (2007). Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of educational psychology*, *99*(4), 808.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of counseling psychology*, *51*(4), 473.

Van Blankenstein, F. M., Dolmans, D. H., van der Vleuten, C. P., & Schmidt, H. G. (2011). Which cognitive processes support learning during small-group discussion? the role of providing explanations and listening to others. *Instructional Science*, *39*(2), 189–204.

Wieman, C., & Perkins, K. (2005). Transforming physics education. *Physics today*, *58*(11), 36.

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker Jr, J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & management*, *43*(1), 15–27.

Appendix A

Key Concepts Relevant to Learning Material

1. 고소 : 고소란, 범죄의 피해자나 피해자와 일정한 관계에 있는 고소권자가 수사기관에 대하여 범죄사실을 신고하여 범인의 처벌을 구하는 의사표시.
2. 고발 : 제 3자, 즉 목격자가 행하는 것을 고발이라고 한다.
3. 인지 : 수사기관이 직권으로 수사를 개시하는 가장 기본적인 방법

   cf. 신고 : 고소는 범인의 처벌을 구하는 의사표시, 그러나 신고는 단순한 피해사실의 신고
4. 수사기관 : 고소는 수사기관인 검사 또는 사법경찰관에게 하여야 한다.
5. 고소능력 : 고소에 대한 소송행위 능력을 말한다. 피해를 받은 사실을 이해하고 고소에 따른 사회생활상의 이해관계를 알아차릴 수 있는 사실상의 행위 능력, 민법상의 행위능력이 없어도 무방하다, 즉 사실상의 의사능력이 있으면 충분하다.
6. 친고죄 : 검사가 공소제기를 위해서, 피해자 기타 고소권자의 고소가 있어야 하는 범죄. (예시) : 모욕죄, 비밀침해죄
   - 절대적 친고죄 : 신분 불문하고, 범죄사실 그 자체로 인해 친고죄
   - 상대적 친고죄 : 직계혈족, 배우자, 동거친족 등 범인과 피해자 사이의 일정한 관계

cf. 양벌규정 : 위법행위에 대하여 행위자를 처벌하는 외에 그 업무의 주체인 법인 또는 개인도 함께 처벌하는 규정. 쌍벌규정이라고도 한다.

7. 반의사불벌죄 : 고소가 없어도 검사가 공소제기를 할 수 있으나, 피해자가 처벌을 원하지 않는다는 의사 표시를 한 경우에는 처벌할 수 없는 범죄. 고소가 없더라도 얼마든지 수사기관의 인지에 의하여 수사를 개시하고 공소를 제기할 수 있다. (예시) : 폭행죄, 협박죄, 명예훼손죄, 과실치상죄

8. 성범죄 : 성범죄에 대해서는 친고죄도, 반의사불벌죄도 적용되지 않는다.

9. 법정대리인 : 피해자가 어린이 등으로 고소할 수 있는 방법을 모를 경우, 친권자나 후견인이 고소권을 가지게 된다. 고아인 경우, 검사가 고소할 사람을 지정해줄 수 있다. 단 법정대리인의 지위는 고소시점을 기준으로 판단하여야 한다. 고소 당시에 존재하였다면, 범죄 당시에 혹은 고소 후에 그 지위를 상실하였더라도 고소의 효력에는 영향이 없다.

   cf. 가족관계등록부에 상관없이, 출생을 하면 법률상의 친족관계에 있다 할 것이다.

10. 피해자 : 자기 혹은 배우자의 직계존속을 고소하지는 못한다. 단, 성폭력범죄는 예외이다. 법인과 법인격 없는 사단이나 재단도 포함될 수 있다. 이때에는 대표자가 고소권을 가진다. 중요한 것은, 고소권은 상속, 양도가 불가하다.

11. 고소권자 : 원칙적으로 피해자 혹은 그 법정대리인, 그러나 일정한 경우 배우자나 친족 등에게도 인정한다. 먼저, 피해자가 사망한 경우에는 배우자, 직계친족 또는 형제자매는 고소할 수 있다. 단, 피해자의 의사를 반하지는 못한다.

    cf. 고소권자 예외 : 재산관리인, 파산관재인, 법인의 대표자는 고소권자가 될 수 없다.

12. 지정고소권자 : 친고죄에 대하여, 고소할 자가 없는 경우 이해관계인의 신청에 자라 검사는 10일 이내에 고소권자를 지정한다. 여기서 이해관계인이란, 법률상 또는 사실상의 이해관계를 가진 자를 말한다.

Appendix B

Examples of Non-Interaction and Interaction Episodes

Non-interaction Episode 1*	여2 : 저는 양벌규정이 뭔지 잘 모르겠어요. 친고죄 부분에 여1 : 저도..	*Turn* 1
Non-interaction Episode 2*	여3 : 저도 뭐 생각났는데, 보통 친고죄의 예 중에 모욕죄와 비밀침해죄가 있다고 했잖아요. 그런데 보통 어떤 사람이 죄를 저지르면, 한 죄목만 가지고 처벌을 받는 것이 아니라, 다른 죄목이 있을 수도 있잖아요. 만약에 다른 죄목에 대해 처벌을 받고 있다면 굳이 다른 것에 대해 처벌할 필요가 없단.. 여1, 여2, 여4 : 아아	*Turn* 1
Non-interaction Episode 3*	여4 : 미성년자는 사실상의 의사능력이 있지 않다고 보는 건가요? 여3 : 고등학생만 되어도 있다고 할 수 있지 않을까. 여2 : 고등학생 정도만 되어도 ,,,,	*Turn* 1

*Non-interaction episodes contain no co-constructive turns.

(continued)

Interaction Episode 1	여1 : 여기 친고죄 부분에서 약간 이해가 안 되는데, 여기 피해자의 의사를 존중할 필요가 있고, 그래서 기타 고소권자의 고소가 필요한, 필요하다고 하는데, 그게 피해자와 고소권자인지, 고소권자 한 명이 되어도 되는 건지 명확하지 않은 것 같아요.	*Turn* 1	*Co-constructive turn*
	남1 : 그냥 이거 기타 고소권자가 원래는 피해자가 고소권잔데, 뒤에서 피해자가 어떤 특별한 경우에 넘어가는 사람들을 다 포함한 경우가 아닐까요?	*Turn* 2	*Co-constructive turn*
	남2 : 총괄적으로, 무조건 고소가 있어야 하는 고소권자를 그런다는..	*Turn* 3	
Interaction Episode 2	남2 : 친고죄가 정확히 무슨 말이죠? 그니깐 고소가 있어야 처벌이 가능하다는 그런건..가?	*Turn* 1	
	남1 : 일단 친고죄는 그러면 검사가 자기 재량껏 할 수 있는 것이 아니고, 고소가 들어와야 할 수 있는 거고, 반면에 반의사불벌죄는 자기 의사대로 수사를 시작할 수 있으나, 피해자가 원하지 않는다면 처벌을 할 수 없다는 그 차이인 것 같은데..	*Turn* 2	*Co-constructive turn*
	여1 : 절대적은 그럼 고소권자가 누가 되든지 딱히 제한은 없다는... 뜻인..?	*Turn* 3	*Co-constructive turn*
	남2 : 아 그런 것 같아요. 상대적은 밑에 나온 이런 결과여야만 하는 것 같고	*Turn* 4	*Co-constructive turn*

Each turn includes a substantive comment. 남 1/2, 여 1/2/3/4 denote various speakers.

국문초록

학생이 능동적으로 학습에 참여할 때 더 많이 배운다는 경험적 증거들이 제시되고 있음에도 불구하고, 현재의 교육은 여전히 강의 중심적이다. 심지어 지식의 양이 중요해 보이는 과학, 기술, 공학, 그리고 수학, 이른바 STEM 분야에서도 전통적 교수법인 강의보다 학생들이 능동적으로 수업에 참여하는 것이 더 많이 배운다. 이런 맥락에서, 본 연구는 다음의 두 가지 목적을 위해 수행되었다. 먼저, 복습과의 비교를 통해 토론이 학습 성과를 높인다는 것을 확인하고자 한다. 두 번째 목적은, 선행하는 활동에 의해 토론의 효과가 달라지는지를 알아보고자 한다. 실험 결과, 강의를 들은 집단 중에서, 복습을 한 집단보다 토론을 진행한 집단이 전반적인 학습 성과가 높았다. 또한 토론을 한 집단 중에서도, 강의 후에 토론한 집단보다 자습 후에 토론한 집단의 학습 성과가 우수했다. 이러한 결과는 자습이 토론을 더욱 풍성하게 한 것에서 기인했다. 결론적으로, 본 실험은 현재의 강의 중심의 교육 방식을 어떻게 바꾸어야 할지에 대한 교육적이고, 실용적인 시사점을 제공한다는 점에서 중요하다.

주요어: 강의, 교육, 능동적 학습, 자습, 토론

학번: 2017-25197

Acknowledgements

It seems like only yesterday that I entered the main entrance of the school with excitement and dreams, but it has already been two years. It was a time when I had to be grateful all the time. There are so many people I want to thank you from the bottom of my heart.

I would like to express my deepest gratitude to advisory professor Jooyong Park, who has been an invaluable mentor. His gift for conceptualization, his enduring encouragement, an his practical advice have been an inestimable source of support for me during this course. I am grateful to all of those with whom I have had the pleasure to work during this and other related projects. Each of the members of my dissertation committee professor Jung-Joon Ihm, and Byoung-Tak Zhang has provided me with professional comments.

I would also like to thank the colleagues in my laboratory, Jung Ae Park, Gyeongmee Gim, Songeui Kim, Jungyeon Park, Ji Won Yang, and Min Hae Song for helpful comments and discussion. Their varied perspectives have helped me to strengthen my work. I truly appreciate Sungjae Cho, Taehyeong Kim, Hyunwoong Ko, Youjin Hwang, Jisoo Jung, Seung Hee Yang, Gi-Cheon Kang, Junho Lee, and Jieun Lee who shared this journey with me.

Nobody has been more important to me in the pursuit of the path I have chosen than the member of family, Hyojeong Lim, Jaejoon Ban, and Jaehyeon Ban; and most importantly, I would like to thank my parents, whose love and guidance are with me in whatever I pursue. I wish to thank my supportive

friends, Hyeonjin Yun, Kang Il Choi, Youngbeen Son, Haneul Kim, Jeonghyeon Kim, Woncher Im, and Wongi Kim who gave me unending inspiration. Thank you to Bongju Shin, Soul Han, Jaeho Oh, Jongwoo Lee, Chankyu Kim, Eun Taek Im, and Kyung Gun Joo who gave their time and provided me with insight into their experience.