Optimal Group Size for High School Students' Collaborative Argumentation Using SNS for Educational Purposes

Michiko Sugai [*], Tatsuya Horita [*], Yuichi Wada [*]

Abstract

This study aims to identify the optimal group size for high school students' collaborative argumentation using SNS for educational purposes. The study is based on the participation of high school students; the participants were divided under two experimental conditions. The students were required to engage in one of two types of arguments: argumentation via SNS or face-to-face argumentation. Furthermore, the students were divided into small groups comprising three to five participants for each condition. Conversations and task performances were collected as the students were engaged in discussions. The transition probabilities between two talkers in the group, the number of utterances, and the task performances were analyzed. The results showed that the 3-person groups held discussions in a lively manner; however, obtaining the consensus in these groups were difficult. The 4-person groups did not converse much initially but eventually conversed enough to be compared with the other groups. It was also easy to obtain the group's consensus. The 5-person groups conversed the least, possibly because of social loafing. The findings suggested that a 4-person group is the optimal size for collaborative argumentation using SNS for educational purposes.

Keywords: collaborative argumentation, Computer-Supported Collaborative Argumentation (CSCA), SNS for educational purposes, group size.

1   Introduction

"Arguing to learn" is becoming a necessary element in education [1]. In Japan, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) emphasizes the necessity for "pro-active learning," "interactive learning," and "deeper learning." The MEXT's stance is based on the belief that teachers can cultivate in students' knowledge, qualities, and abilities through interactive learning [2]. Collaborative argumentation is a form of interactive learning that aims to enable participants to examine their opinions and points of common understanding. By doing so, collaborative argumentation enables participants to better integrate their opinions, which enables them to form better opinions [1]. The effects of collaborative argumentation on learning have at least the following five educational benefits: (a) Increased motivation for learning, (b) Deepened

[*] Graduate School of Information Sciences, Tohoku University, Miyagi, Japan

understanding of learning content, (c) Improved acquisition of general argumentation skills, (d) Improved acquisition of specific argumentation skills, and (e) Better knowledge-building practices [3].

One method of enhancing students' ability to argue effectively is the Computer-Supported Collaborative Learning (CSCL). CSCL systems can support and guide productive argumentation, thereby facilitating a deeper understanding among students involved in this process [1]. Notably, CSCL supporting for collaborative discussion is called Computer-Supported Collaborative Argumentation (CSCA) [4]. CSCA systems make possible the scaffolding of critical discourse and argumentation processes [5].

As one of the tools of CSCA, Social Networking Service (SNS) is thought to be useful for discussions. SNS is a form of social media in which users transmit and form information [6]. Social media provides many benefits, such as the creation and exchange of knowledge, deepened trust relationships, the creation of cultural values, collaborative problem solving, the acquisition of diverse perspectives, and the creation of industries and businesses. Depending on the usage, conflicts and confusion, discrimination and disparities, and concentration of power, an unfair use may occur [7]. SNS provides easy, platform-independent access to and almost unrestricted interactivity for sharing ideas and opinions. In addition, SNS can be conducive to online dialogic exchange and argumentative knowledge construction [8].

In elementary and secondary education, it is conceivable to use "SNS for educational purposes" (hereafter referred to as "SNS"). SNS enables teachers to pay attention to the words and actions students use to exchange ideas. In addition, it also provides secure communication. Holland and Muilenburg examined high school students' SNS-based discussions in a classroom setting. SNS called "Edmodo" was similar to Facebook in appearance and operation. Students had frequently used Facebook; thus, they could easily master the operation of Edmodo [9]. Holland and Muilenburg reported that SNS enabled students to share deeper insights because they required minimal teacher intervention [9]. Sugai et al. examined high school students asked to solve problems collaboratively and reported that students made fewer remarks in discussions held via SNS compared with face-to-face (FTF) discussions. Nevertheless, the quality of the students' work was observed to be similar for both types of discussions, and students' ideas were more original in SNS discussions [10].

Although some studies have focused on collaborative argumentation via SNS, little is known of the optimal initial conditions for discussions via SNS. According to Dillenbourg, one question teachers frequently ask about initial conditions for collaborative learning is "What is the optimal size of a group?" [11]. The greater the number of individuals in a group, the greater their resources when working in a group. By contrast, as the number of individuals increases, the contribution to the solution of the problem of each individual in the group decreases [12]. Thus, group size should be small because all students can be active and everyone can participate equally [12]. The typical size of collaborative learning groups in FTF meetings is two to four persons, based on the teacher's rule of thumb that "small group size is good." [12]. By contrast, empirical studies have demonstrated that optimal group size is six to eight persons [13]. Very few studies have compared the effects of group sizes on interactions in computer-supported group-based learning (CSGBL) settings [8]. Group size may affect group performance and argumentation patterns [14]. In addition, group size may improve or reduce learning outcomes [15]. Therefore, group size should be considered while designing group learning environments [16]. In a comparison of 2- and 4-person groups, the composition of the 4-person groups was more likely to induce cognitive conflict (disagreement and negotiation) [17]. In addition to 4-person groups, it is considered necessary to investigate the group size around four persons. Therefore, this study aims to identify the effects of group size on the initial conditions for collaborative argumentation. High school students

were divided into two groups; they were either required to discuss via SNS or discuss face-to-face. Small groups of three to five students were formed for each condition. Based on the utterances and task performances collected during the students' discussions, the effects of group size were examined in order to determine the optimal conditions for collaborative argumentation via SNS for educational purposes.

2   Method

2.1   Participants

A total of 238 Japanese second-year high school students from Miyagi Prefecture (males: 157, females: 81) participated in this study. All participants were enrolled in a public high school.

2.2   Conditioning by Types of Discussion

The participants were randomly assigned to one of three experimental conditions. The experimental conditions were based on a combination of two types of discussions. The first condition required participants to discuss via SNS both times (*n*=80); this group shall hereafter be referred to as the "SNS condition." The second condition required participants to discuss face-to-face both times (*n*=80), and this group shall hereafter be referred to as the "FTF condition." The third condition required participants to discuss face-to-face the first time, and via SNS the second time (*n*=78). The third condition was excluded from the analyses because it contained a hybrid of the two discussion conditions in which the same participants were engaged.

2.3   Small Groups for Discussion

We created small groups of three to five persons in both conditions (hereafter referred to as "group size"). Each participant was assigned to one of the small groups. We nominated a leader for each small group on the basis of our analysis of the questionnaire participants had completed before the lesson. The scales we used were "Discussion Skill Scale [18]" and "Personal Report of Communication Apprehension (PRCA-24) [19]." The Discussion Skill Scale is a scale to measure the degree of participant's discussion skills. Yasunaga et al. defined discussion skills as participants gaining a better understanding of certain matters and solving problems in a friendly atmosphere with the free exchange of information and opinions [18]. The Discussion Skill Scale has four subscales: "Advancing and handling the discussion," "Positive participation and self-assertion," "Understanding and consideration for others," and "Creating an approachable atmosphere." PRCA-24 is a psychological scale to measure anxiety in dialogue communication. The subscale comprises the four items: "group discussion," "meeting," "conversation," "speech," and using the total score of those. Participants who rated themselves as being good at facilitating discussions, creating an approachable atmosphere and not feel uneasy about talking to others were assigned as leaders. Next, participants who rated themselves as not good at discussions with others were assigned to each group. Other participants were randomly assigned to the small groups. For the SNS condition and the FTF condition, we created small groups in the following manner: 3-person groups × 4, 4-person groups × 4, and 5-person groups × 2. The numbers of male and female participants in each small group are presented in Table 1.

Table 1: Number of Person in Small Groups by Types of Discussion

Group size	Number of people	SNS	FTF
3-person groups	male × 3	2	2
	female × 3	2	2
	gender-mixed	0	0
4-person groups	male × 4	2	1
	female × 4	0	0
	gender-mixed (male × 2, female × 2)	2	3
5-person groups	male × 5	0	0
	female × 5	0	0
	gender-mixed (male × 3, female × 2)	2	1
	gender-mixed (male × 2, female × 3)	0	1

2.4 Lesson Schedule

This investigation was conducted as part of a course called "Information Study by A Scientific Approach" on the "basic ways of problem-solving" and "problem-solving utilizing information and communication networks." The lesson schedule is presented in Figure 1

2.5 Learning Environment

Lessons were conducted in the computer room. Participants accessed and answered the web-based questionnaire via Moodle, a learning support system. They registered their answers in Moodle's "feedback" module. For the SNS condition and discussions pertaining to this condition, we used an SNS designed for educational purposes called "ednity" (https://www.ednity.com/).

2.6 Data Collection

We analyzed the data obtained during the first, fourth and sixth periods of the lesson schedule. The details are presented as follows.

2.6.1 Typing Speed

In the first period, the number of characters of a Japanese sentence that participants accurately typed on a personal computer in 10 minutes was measured. This measurement was conducted for all the participants of the SNS and FTF conditions.

2.6.2 Conversations in Argument Practice

The SNS condition discussed via SNS, and their conversations were recorded in the SNS log. The FTF condition discussed face-to-face, and their conversations were recorded using an IC recorder; their conversations were later transcribed as text files.

(1) Theme of the First Argument Practice: During the first practice, the participants undertook a consensus exercise entitled "NASA exercise: Survival on the Moon" (e.g., https://www.nasa.gov/pdf/166504main_Survival.pdf). Consensus exercises are sometimes used in corporate training programs or group activities in school education. As the name suggests, consensus exercises mainly aim to reach a consensus among the participants. Typically, these exercises require participants to discuss a given problem as a group [20].

Period 1: Measurement of typing speed

▼

Period 2: Answering a prequestionnaire

▼

Period 3: Lesson on the way of argumentation

(1) Pretest on reasoning

(2) Lesson

(3) Posttest on reasoning

(4) Postquestionnaire

▼

Period 4: First argument practice

(1) Discussion
Theme: NASA exercise
Survival on the Moon
Discussion time: 20 minutes

(2) Postquestionnaire

▼

Period 5: Reflection on the first argument practice

(1) Reflecting on the previous lesson

(2) Postquestionnaire

▼

Period 6: Second argument practice

(1) Discussion
Theme: Idea generation and planning
A comfortable addition to our classroom
Discussion time: 25 minutes

(2) Postquestionnaire

▼

Period 7: Reflection on the second argument practice

(1) Reflecting on the previous lesson

(2) Postquestionnaire

▼

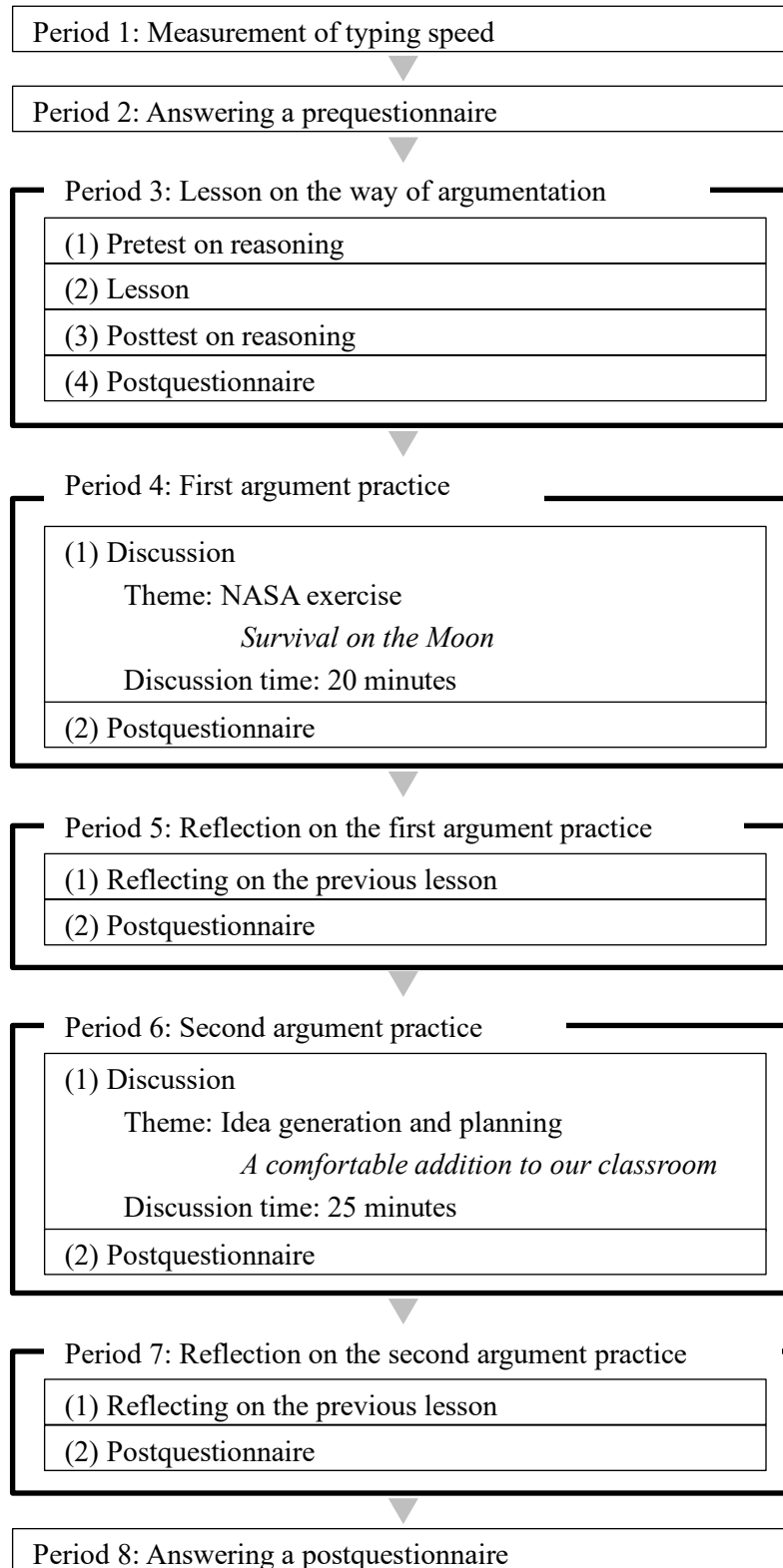Period 8: Answering a postquestionnaire

Figure 1: Flowchart of Lesson Schedule

(2) Theme of the Second Argument Practice: During the second practice, the participants were required to discuss the following theme: "What needs to be instilled in the classroom to make it more comfortable?" The aim of this practice was to enable participants to solve a familiar and open-ended problem.

2.6.3 Scores of the Consensus Exercise in the First Practice

In the first practice, participants were presented ten items, which had to be ranked in the order that would best enable them to survive a crash landing on the Moon. Participants were then re-quired to write the results on a worksheet. The participants completed the following steps and wrote the results on the worksheet. First, the participants ranked each item individually. Second, the groups discussed the reasons for their individual choices. Third, each group agreed on a final ranking for the items; this stage represented the consensus-building process. The exercise had one correct answer, and lower scores indicated better group performance. The submitted work-sheets contained the participants' individual choices and the groups' decisions. Figure 2 presents the worksheet filled out with model answers and the scoring methods used. The original work-sheet was written in Japanese. Teacher determined that a consensus was reached within the group when members of the same group wrote the same content on the worksheet.

3   Analyses and Results

3.1 Typing Speed

Two-way analysis of variance (ANOVA) for the effects of group sizes and gender were conducted based on typing speed (i.e., number of characters typed in 10 minutes) measured in the first period. This analysis was performed only for the SNS condition because the FTF condition did not use a personal computer in both argument practices. The means and standard deviations of typing speed per person in the SNS condition are presented in Table 2. This analysis revealed a main effect of gender ($F(1, 32)= 26.77$, $p<0.05$, $\eta_P^2=0.18$) but no main effect of group sizes and no interaction of group sizes and gender. Post hoc Holm comparisons demonstrated significant dif-ferences between males and females. The typing speed of females was faster.

Table 2: Means and Standard Deviations (SD) of the Number of Characters typed in 10 minutes per Person in the SNS Condition

Group size	Male M	Male SD	Female M	Female SD	Whole M	Whole SD
3-person groups	323	41	330	41	326	32
4-person groups	288	29	451	51	369	29
5-person groups	348	41	453	51	400	33
Whole	320	22	411	28	344	112

<u>Submit this worksheet</u>
<u>at the end of class</u>   Worksheet   Discussion date: ／

Grade:2   Class: No: Group No: Name:

【Model answer by NASA】

Lefted items	Order	Reason (Grounds and/or Warrant)	Model anser	Score
Matchbox	10	There is no oxygen to ignite the moon. There is no value at all.		
Condensed foods	4	It is an effective means of supplying body energy.		
Parachute	7	It helps to block the sun and protect yourself.		
Portable heater using solar heat	9	It is not necessary unless it is shaded.		
Oxygen tank (50kg)	1	It is the most necessary item for survival.		
Constellation map used on the moon	3	It is indispensable to confirm the present location and to decide the direction of the road.		
Water (20L)	2	It is necessary for survival after oxygen.		
Lighting bullets used in space	8	When you find a mother ship, you can send a distress signal.		
First aid kit with syringe	6	Vitamins and drugs can be injected through special holes in the space suit.		
Solar cell type FM receiver / ransmitter	5	You can only use a short distance, but it will help you to contact you when you approach the mother ship.		
			Total	

■How to calculate the score :
Score = Order by your ideas or Group's ideas － Order by Model Anser (In ablolute value)
■Example of Scoring
【Your ideas】

Lower scores are better !

Lefted items	Order		Model anser	Score
Matchbox	10		10	0
Condensed foods	3		4	1
Parachute	9		7	2
Portable heater using solar heat		The part where the order is not finished is calculated as 0.	9	9
Oxygen tank (50kg)	2		1	1
Constellation map used on the moon	5		3	2
Water (20L)	1		2	1
Lighting bullets used in space	4		8	4
First aid kit with syringe		The part where the order is not finished is calculated as 0.	6	6
Solar cell type FM receiver / ransmitter	6		5	1
			Total	27

Figure 2: Worksheet with model answers and scoring

3.2 Transition Probabilities between Two Talkers

The transition probabilities between two talkers (hereafter referred to as "transition probabilities") were obtained by converting the frequency of the transition in the turnover of the talkers into the probability. By calculating the transition probabilities, we determined whether a participant was good at inducing the utterance of another person and/or continuing with his or her topic after another person's utterance. This capacity was quantitatively indicated. In addition, it is also possible to extract a combination of participants who are likely to make a transition of utterances from individual transition tendency [21]. In this study, the transition probabilities were calculated by analyzing the interaction sequences. For this purpose, we used "GSEQ" (The Generalized Sequential Querier: http://bakeman.gsucreate.org/) software.

Before calculating the transition probabilities, the talkers in each small group were encoded as follows. Leaders of the small group were encoded as "A." Other members were encoded in descending order from "B" to "E," based on their self-evaluation of discussion skills. When two or more participants had the same self-evaluation in terms of their discussion skills, they were encoded according to their anxiety levels during conversations.

After calculating the transition probabilities, state transition diagrams were drawn for each type of discussion and all group sizes. These are presented in Figure 3-5. The nodes within the circle represent the members (participants) of the groups. The nodes were positioned clockwise in the descending order, based on participants' discussion skills. The solid line with the arrow indicates the transitions between two talkers with a significance level of 5% or 1% (e.g., "A->B" means A is followed by B). The value on the solid line indicates the transition probabilities.

A summary of the possible interpretations of the characteristics of the state transition diagrams for each condition is as follows.

3.2.1 3-person Groups' Discussion

(1) First Practice

SNS condition: The two members with the highest score for discussion skills (i.e., A and B) repeated remarks; they talked in turns. The member with the lowest skill score (i.e., C) rarely talked to a particular person in succession. Few exchanges of utterances were observed between the members (i.e., B and C).

FTF condition: Frequent individual interactions were observed between the leader and the other two members. Few exchanges of utterances were observed between the members (i.e., B and C).

(2) Second Practice

SNS condition: The leader's remarks followed the remarks of the member with the lowest skill score (i.e., C). The remarks of the other members (i.e., B) followed the leader's remarks. Few conversational exchanges were observed between B and C.

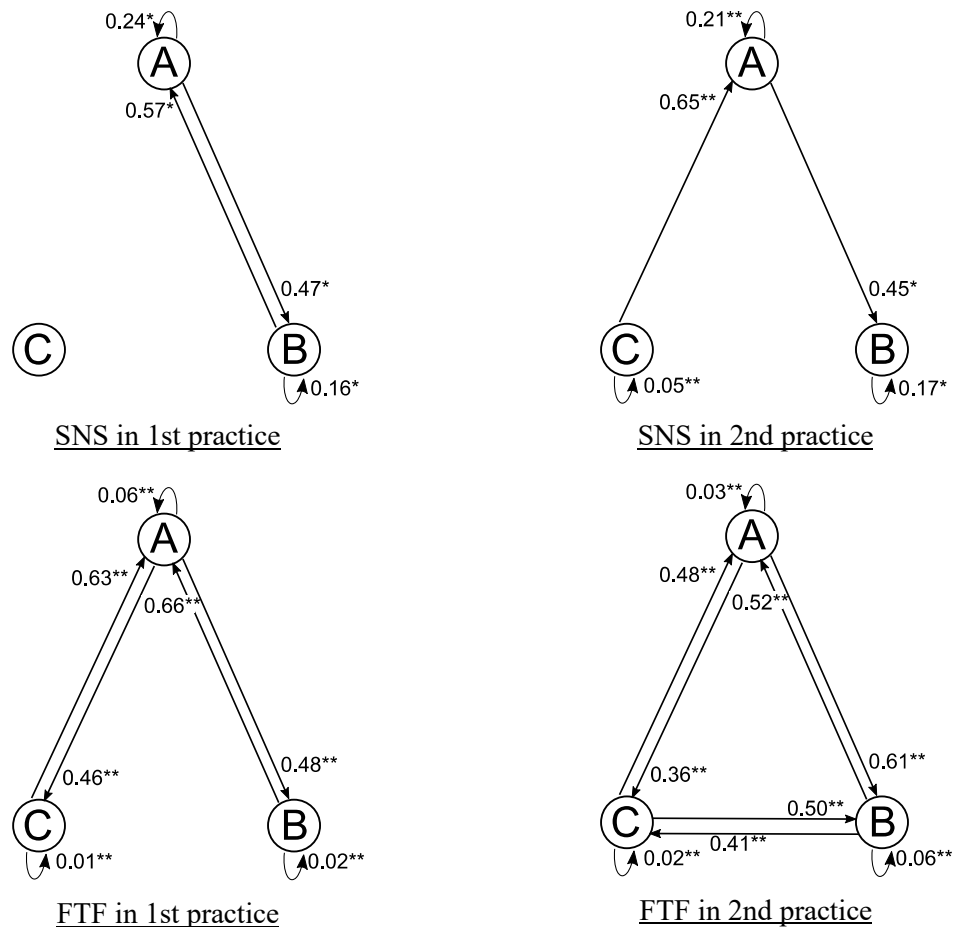FTF condition: All members maintained lively contact with each other.

Figure 3: Transition probabilities in 3-person groups

3.2.2 4-person Groups' Discussion

(1) First Practice

SNS condition: No significant transition pattern of remarks was observed.

FTF condition: Mutual remarks were observed between the leader and the other members. Few conversational exchanges were observed between B and C.

(2) Second Practice

SNS condition: The leader's remarks followed the remarks of the member with the lowest skill score (i.e., D). D's remarks followed the remarks of the member with the second-lowest skill score (i.e., C). The role of the member with the second-highest skill score (i.e., B) was not clear.

FTF condition: The pattern of transition was similar to that observed in the first practice. Different from the results of the first practice, the number of the utterances of the lowest skill member (i.e., D) who talked after the member with the second-highest skill (i.e., B) decreased. In addition, the number of the utterances of the member with the second-lowest skill score (i.e., C) who talked after the member with the second highest-skill score (i.e., B) increased.
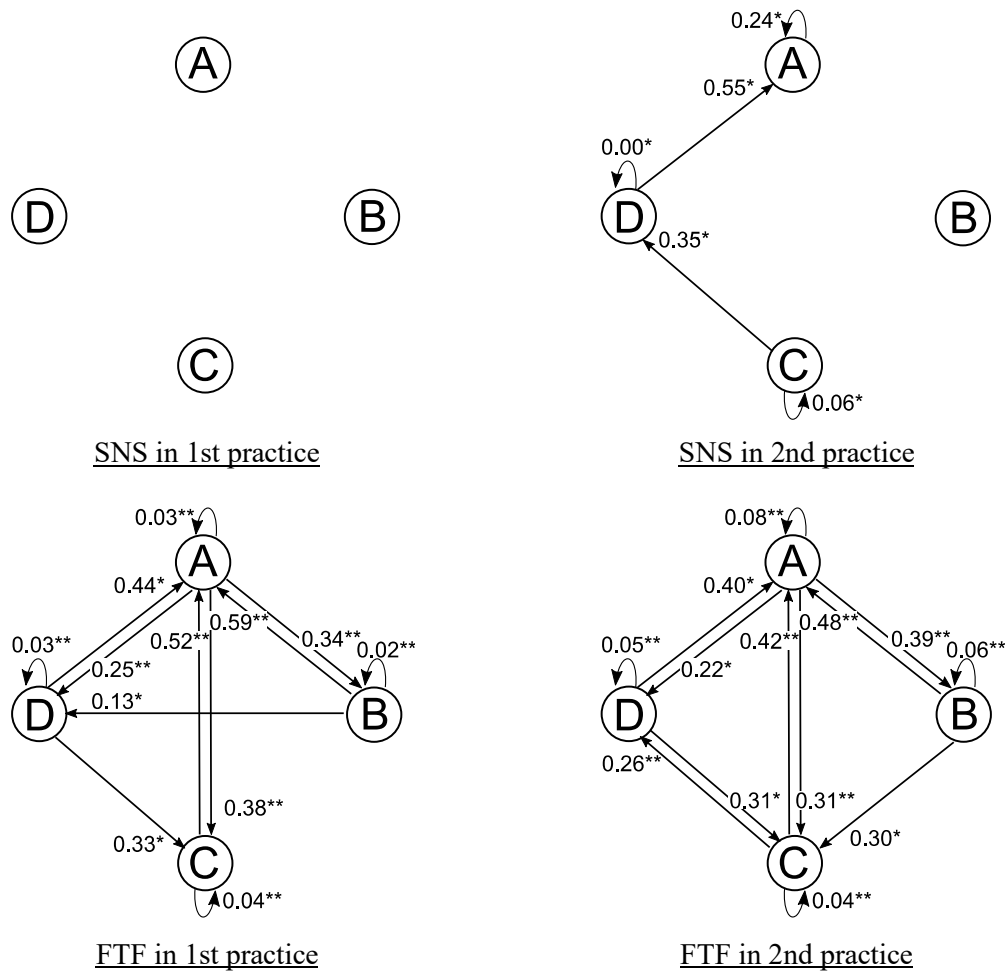
Figure 4: Transition probabilities in 4-person groups

3.2.3 5-person Groups' Discussion

(1) First Practice

SNS condition: The exchanges of remarks tended to start from the member with the lowest skill score (i.e., E). The role of the member with a medium-level skill score (i.e., C) was not clear.

FTF condition: Most of the members maintained lively contact with each other, except for D, who had the second-lowest discussion skill score.

(2) Second Practice

SNS condition: No significant transition pattern of remarks was observed.

FTF condition: The exchange of utterances between the leader (i.e., A) and the member with the second-highest discussion skill score (i.e., B) disappeared. Additionally, the pattern of transition was similar to that observed in the first practice.
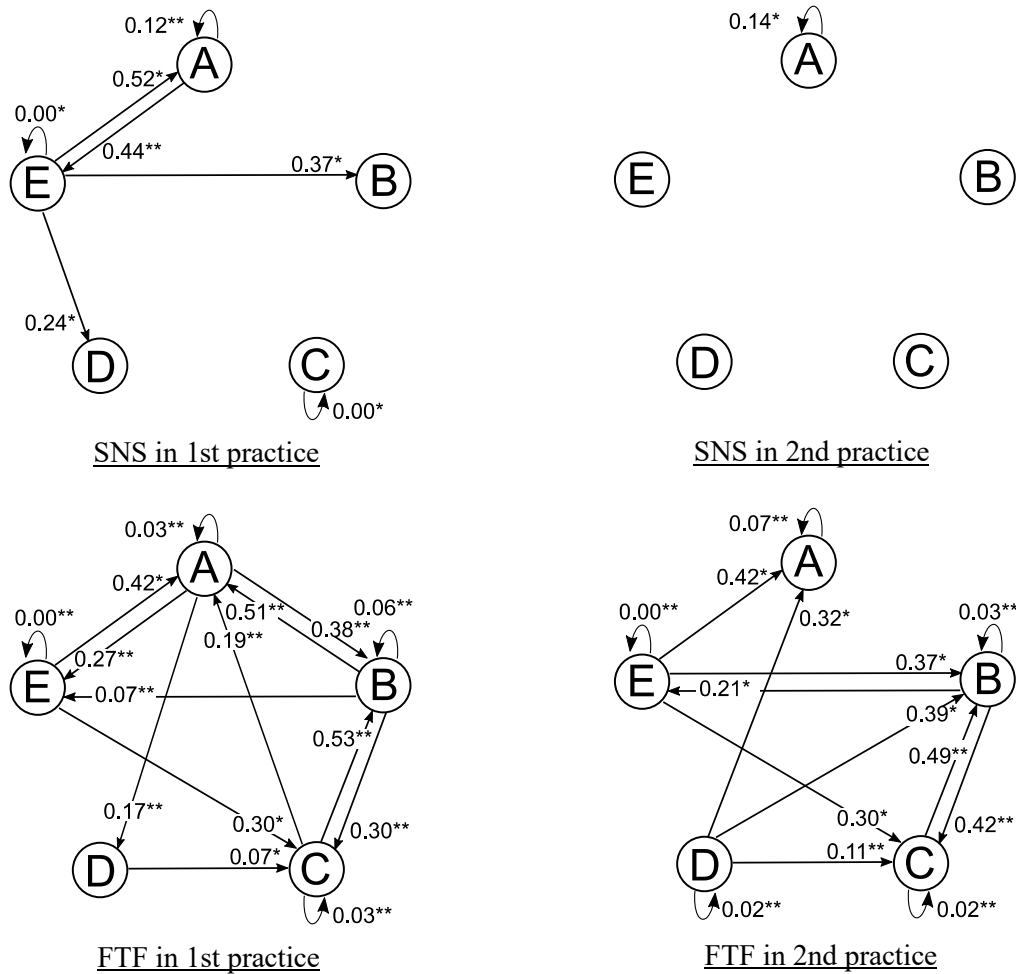
Figure 5: Transition probabilities in 5-person groups

3.3 Number of Utterances

Table 3 presents the means and standard deviations of the number of utterances for each condition.

By studying the differences between CMC (Computer-Mediated Communication) and FTF communication, Bordia pointed out that in a given time period, CMC produced fewer utterances than FTF communication [22]. Thus, we could not simply compare the number of utterances in the SNS condition to the number of utterances in the FTF condition. Therefore, we analyzed the SNS condition and the FTF condition separately.

3.3.1 Comparison by Group Size in the SNS Condition

We performed a two-way repeated-measures ANOVA to determine the effects of group size (3-person/4-person/5-person; as between-subject) and phase of practice (first practice/ second prac-tice; as within-subject). The ANOVA revealed that the main effects for group size were ($F_{(2, 35)}$=3.34, $p<0.05$, η_p^2=0.16); the main effects for the phase of practice was observed to be ($F_{(1, 35)}$=4.57, $p<0.05$, η_p^2=0.12). In addition, a significant interaction was observed between group size and phase of practice ($F_{(2, 35)}$=16.42, $p<0.001$, η_p^2=0.48). A post hoc Holm's test demonstrated that 4-person groups had significantly fewer utterances

Table 3: Means and Standard Deviations of the Number of Utterances per Person

Types of Discussion	SNS				FTF			
	1st		2nd		1st		2nd	
Group size	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
3-person groups	10.3	5.6	11.6	4.8	49.2	18.5	78.8	35.6
4-person groups	4.1	3.1	9.8	4.1	57.6	30.9	75.0	36.5
5-person groups	9.9	5.6	6.9	3.3	44.4	26.4	46.2	25.9

1st: first practice
2nd: second practice

than the other groups during the first practice. In addition, 5-person groups had significantly fewer utterances compared with 3-person groups during the second practice.

3.3.2 *Comparison by Group Size in the FTF Condition*

We performed a two-way repeated-measures ANOVA to determine the effects of group size (3-person/4-person/5-person; as between-subject) and phase of practice (first practice/second practice; as within-subject). The ANOVA revealed that the main effects for the phase of practice were ($F(1, 35)=13.15$, $p<0.01$, $\eta_p^2=0.27$). No significant interaction was observed between group size and the phase of practice. The second practice had the same number of utterances as the first practice.

3.4 Task Performances in the First Practice

Table 4 presents the means and standard deviations of the scores of the consensus exercise before and after discussions in the first practice.

3.4.1 *Comparison by Group Size in the SNS Condition*

We performed a two-way repeated-measures ANOVA to determine the effects of group size (3-person/4-person/5-person; as between-subject) and score (before discussion/after discussion; as within-subject). The ANOVA revealed that the main effects for group size were ($F(2, 35)=10.66$, $p <0.001$, $\eta_p^2=0.38$), and the main effects for the score were ($F(1, 35)=287.17$, $p <0.001$, $\eta_p^2=0.89$). In addition, a significant interaction was observed between group size and score ($F(2, 35)=3.64$, $p<0.05$, $\eta_p^2=0.17$). A post hoc Holm's test demonstrated that 4-person groups had sig-nificantly lower scores compared with 3-person groups and 5-person groups.

3.4.2 *Comparison by Group Size in the FTF Condition*

We performed a two-way repeated-measures ANOVA to determine the effects of group size (3-person/4-person/5-person; as between-subject) and score (before discussion/after discussion; as within-subject). The ANOVA revealed that the main effects for group size were ($F(2, 35)=5.34$, $p <0.01$, $\eta_p^2=0.23$).

Table 4: Means and Standard Deviations of Scores of the Consensus Exercise per Person

Types of Discussion	SNS				FTF			
	Before		After		Before		After	
group size	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
3-person groups	18.0	4.3	47.6	7.5	17.3	5.4	18.8	5.2
4-person groups	13.9	6.0	36.7	9.9	15.5	4.3	13.5	2.7
5-person groups	13.1	5.0	46.6	3.2	15.4	6.5	12.0	4.2

Before: Before discussion
After: After discussion

3.5 Variables Likely to Correlate with Typing Speed

Typing is indispensable for discussions using SNS. Bordia reported that typing speed was related to the number of utterances and the ability to reach consensus [22]. Based on this, we expected correlations between typing speed and such variables in this study. Therefore, the correlation between typing speed and number of utterances, and the correlation between typing speed and task performances in the first practice were analyzed.

3.5.1 Correlation between Typing Speed and Number of Utterances

To analyze whether typing speed affected the number of utterances, the correlation between the typing speed and the number of utterances was analyzed. A positive relationship was observed in 3-person groups in the second practice. The correlation coefficients between typing speed and number of utterances in the SNS condition are presented in Table 5.

Table 5: Correlation Coefficients (*r*) between Typing Speed
and Number of Utterances in the SNS Condition

Group size	1st practice	2nd practice
3-person groups	0.28 *n.s.*	0.69 *
4-person groups	-0.37 *n.s.*	-0.09 *n.s.*
5-person groups	-0.13 *n.s.*	-0.18 *n.s.*

*$p < .05$

3.5.2 Correlation between Typing Speed and Task Performances in First Practice

To examine whether typing speed affected the task performances, the correlation between typing speed and the score after the discussion was analyzed. No correlation was observed between the typing speed and the score after the discussion. The correlation coefficients between typing speed and the score after the discussion in the SNS condition are presented in Table 6.

Table 6: Correlation Coefficients (*r*) between Typing Speed
and Score of after Discussion in the SNS Condition

Group size	After discussion
3-person groups	-0.20 *n.s.*
4-person groups	0.11 *n.s.*
5-person groups	-0.41 *n.s.*

4   Discussion

4.1   General Characteristics of the Collected Data

4.1.1   Typing Speed

This study deals with synchronous communication; thus, it is regarded not only as a CSCA study but also as a Synchronous Computer-Mediated Communication (SCMC) study. Lin noted that participants needed to be equipped with good/fast typing skill, because they needed to respond immediately to conversational partners in SCMC [23]. On the other hand, Sykes noted that variables such as typing speed could easily affect the outcome of a study [24]. In the results of the correlation analysis, no correlation was observed between the typing speed and the number of utterances, except for the 3-person groups in the second practice. Additionally, no correlation was observed between the typing speed and the score. In the 3-person groups, the number of times of turn-taking for the utterance is heavier than that of the other group sizes. Individuals with good typing skills can keep up with a conversation, whereas individuals with poor typing skills may be left out of a conversation and miss opportunities to remark.

Because the pace of turn-taking of utterances slows as the number of people increases, it is considered that 4-person group and 5-person group have a lower likelihood of losing the opportunity to remark than 3-person group. Based on these considerations, in terms of typing speed, 3-person group is not an appropriate group size for collaborative argumentation.

4.1.2   Transition Probabilities

In 3-person and 4-person groups, the number of transitions of conversations between two specific talkers tended to increase during the second practice. Moreover, in the second practice, we surmised that the participants may have attempted to promote a smooth discussion based on the discussion during the first practice. In both the SNS condition and the FTF condition, the tendency to talk more frequently after a specific person tended to be lower in 5-person groups in the second practice than in the first practice.

4.1.3   Mean of the Number of Utterances per Person

The mean of the number of utterances per person was significantly less in the SNS condition than in the FTF condition. This finding may be because typing was indispensable for the SNS condition, that is, participants assigned to the SNS condition used more time to input their thoughts. In addition, unlike a face-to-face discussion, it was not necessary to return a quick response to keep pace with others. This seemed to afford them enough time to form their opinions in their pace. Therefore, their remarks tended to be delayed [25].

4.1.4   Mean of the Score of Consensus Exercise per Person

Prior to the discussions during the first practice, participants ranked the ten items individually in the consensus exercise. We observed no significant difference in the mean scores of the consensus exercise between the SNS condition and the FTF condition. The mean scores of the consensus exercise, which were ranked based on the results of the discussions, had a significantly higher score in the SNS condition than in the FTF condition. In the rule of this consensus exercise, lower scores indicate better performance and greater consensus. In the SNS condition, the discussion progressed rather slowly. A few groups did not complete the exercise.

Therefore, to simply com-pare the SNS condition with the FTF condition was difficult. As further research, it is necessary to study the intergroup differences. In particular, it is important to determine whether consensus is obtained after a sufficient amount of time is provided for the discussion in the SNS condition.

4.2 Characteristics by Group Size

4.2.1 Characteristics of 3-person Groups

According to the transition probabilities, the tendency of members assigned to the FTF condition to talk with each other was high for both practices. In the SNS condition, participants who regarded themselves as poor communicators (coded as "C") tended to be isolated from the other members.

In the SNS condition, the number of utterances in 3-person groups tended to be higher than the number of utterances in 4-person and 5-person groups. This finding may be due to the number of persons in these groups, that is, members in the 3-person groups may have talked more than members of the other groups. The members of the 3-person groups talked more frequently in the second practice than in the first practice. This finding was true for both discussion types. Because the participants had become accustomed to the types of discussions and group members, the number of utterances may have increased.

The scores of the consensus exercise in the first practice in the SNS condition showed that the mean of the scores of 3-person groups after the discussion was significantly higher than the scores obtained by the 4-person groups. No significant difference was observed between the scores of 3-person groups and the scores of 5-person groups. In the first practice, lower scores indicated better group performance. The performance achievement of the 3-person groups and the 5-person groups were lower than that of the 4-person groups.

Collaborative argumentation in 3-person groups in the SNS condition may have been domi-nated by the two members with the highest discussion skill scores. In addition, because of poor group performance achievement, these groups had difficulty reaching consensus.

4.2.2 Characteristics of 4-person Groups

In the FTF condition, from the standpoint of the transition probabilities, many participants conversed with each other during both practices. No exchanges with any specific person were observed during the first practice in the SNS condition. We observed that the number of interactions with a specific person increased during the second practice (e.g., C talked continuously; C talked after B; A talked after D). From the utterances of A after D, it can be deduced that A, the leader, supported D who was not particularly good at discussions.

In the SNS condition, the number of utterances during the first practice in the 4-person groups was significantly less than the number of utterances in 3-person and 5-person groups. No significant difference was observed in the second practice. Many groups had an equal number of male and female participants—typically, two each. Daibo et al. reported that females were more active than males in participation and utterances towards problem solving in small groups of university students [26]. As was the case with Daibo et al. [26], in the SNS condition of this study, it seemed there was a possibility that the conversation did not advance because of gender differences at the first practice.

The number of utterances of 4-person groups was lower than those of 3-person groups and 5-person groups in the first practice. Nevertheless, the mean scores of the 4-person groups

were significantly lower than those of the 3-person and 5-person groups. This indicates that the perfor-mance achievement in 4-person groups was good. In the SNS condition, 4-person groups tended to reach a consensus more easily compared with the other groups.

4.2.3 Characteristics of 5-person Groups

In 5-person groups, members from the SNS condition and the FTF condition tended to not talk after someone in particular. In the SNS condition, the number of utterances in 5-person groups was less than the number of utterances in 3-person groups.

Based on the scores of the consensus exercise, discussions in the FTF condition led to consensus more easily compared with the other groups. By contrast, in the SNS condition, there was no significant difference between 3-person groups and 5-person groups. The difference was significantly higher between 4-person groups and 5-person groups. Therefore, we posit that reaching consensus in the SNS condition was difficult.

According to Kreijns et al., anonymity and nonparticipation increase as group size increases [27]. The social loafer differs from the free rider in that the former lacks the motivation to add to the group performance, whereas the latter attempts to profit from others while minimizing essential contributions [27]. Therefore, 5-person groups are not feasible for discussion on SNSs.

4.3 Optimal Group Size

Based on these considerations, we attempted to identify optimal group size for collaborative argumentation using SNS.

First, a 5-person group is not feasible, because of the possibility of a social loafer and because this number of individuals does not facilitate conversations or discussions. We did not observe many differences between 3-person and 4-person groups, especially regarding the transition probabilities between two talkers. The number of utterances in 4-person groups was less than the number of utterances in 3-person groups during the first practice. No significant difference was observed during the second practice. Regarding the ease of reaching consensus, statistically significant results were obtained for the 4-person groups. Therefore, 4-person groups were observed to be particularly apt for collaborative argumentation using SNS.

5 Conclusion

This study aimed to determine the optimal group size for collaborative argumentation using SNS for educational purposes. Based on the transition probabilities between two talkers, the number of utterances, and the task performance, the results showed that a 4-person group was the optimal group size for collaborative argumentation using SNS for educational purposes. Identifying the optimal group size for collaborative argumentation using SNS for educational purposes is critical to improve group performance and enhance the learning.

This study also demonstrated that the number of utterances per person in the SNS condition was significantly less than the number of utterances per person in the FTF condition. Therefore, a limitation of this study was that we could not obtain a sufficient number of utterances to calculate the transition probabilities. Another limitation was that we did not consider gender differences while assigning participants to groups.

We recommend that further research should consider the differences between the groups after providing a sufficient amount of time to finish the discussion in the SNS condition.

Additionally, researchers could ask students for their impressions and evaluations regarding collaborative argumentation using a questionnaire and examine the correlation between the answers and group size.

Acknowledgement

References

[1] J. Andriessen, M. Baker, "Arguing to Learn," in "The Cambridge handbook of the learning sciences, second edition," Cambridge University Press, 2014, pp. 439-460.

[2] Ministry of Education, Culture, Sports, Science and Technology, "the new Course of Study (Public notice in March, 2017)," from http://www.mext.go.jp/a_menu/shotou/new-cs/1383986.htm.

[3] C.A. Chinn and D.B. Clark, "Learning through collaborative argumentation," in "International Handbook of Collaborative Learning," New York: Taylor & Francis, 2013, pp. 314-332.

[4] T. Sumner and S.B. Shum, "Open peer review & argumentation: Loosening the paper chains on journals," Ariadne, 5, 1996.

[5] A.C. Jeong and J. Lee, "The effects of active versus reflective learning style on the processes of critical discourse in computer-supported collaborative argumentation," British Journal of Educational Technology, Vol. 39, 2008, pp. 651–665.

[6] Ministry of Internal Affairs and Communications, "White Paper on Information and Communication 2018" from http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/pdf/30honpen.pdf.

[7] Y. Nakahashi, "A Study on 'Reading Ability' in Social Media," Information science and technology, Vol. 68, No. 8, 2019, pp. 406-411.

[8] D. Tsovaltzi, T. Puhl, R. Judele and A. Weinberger, "Group awareness support and argumentation scripts for individual preparation of arguments in Facebook," Computers & Education, Vol. 76, 2014, pp. 108-118.

[9] C. Holland and L. Muilenburg, "Supporting student collaboration: Edmodo in the classroom," In Society for Information Technology & Teacher Education International Conference, Vol. 2011, No. 1, 2011, pp. 3232-3236.

[10] M. Sugai, T. Horita and Y. Wada, "Characteristics of High School Students' Collaborative Problem Solving using SNS for Educational Purposes," Information Processing Society of Japan Transactions, Education and Computer (TCE), Vol. 4, No. 1, 2018, pp. 43-56.

[11] P. Dillenbourg, "What do you mean by collaborative learning?," Collaborative-learning: Cognitive and Computational Approaches, Oxford: Elsevier, 1999, pp.1-19.

[12] D.W. Johnson, R.T. Johnson and E.J. Holubec, "Circles of learning: Cooperation in the class-room, 5th edition," Edina, MN: Interaction Book Company, 2002.

[13] S. Sugie, "Introduction to collaborative learning: understanding the basics and 51 ideas," Kyoto: Nakanishiya Publishing, 2011.

[14] J.W. Strijbos, R.L. Martens and W.M.G. Jochems, "Designing for interaction: Six steps to designing computer-supported group-based learning," Computers and Education, Vol. 42, No. 4, 2004, pp. 403–424.

[15] A. Valero Haro, O. Noroozi, H. Biemans, and M. Mulder, "First-and second-order scaffold-ing of argumentation competence and domain-specific knowledge acquisition: a systematic review," Technology, Pedagogy and Education, Vol.28, No. 3, 2019, pp. 1-17.

[16] O. Noroozi, A. Weinberger, H.J. Biemans, M. Mulder and M. Chizari, "Argumentation-based computer supported collaborative learning (ABCSCL): A synthesis of 15 years of re-search," Educational Research Review, Vol. 7, No.2, 2012, pp. 79-106.

[17] L.S. Fuchs, D. Fuchs, S, Kazdan, K. Karns, M.B. Calhoon and C.L. Hamlett, et al., "Effects of workgroup structure and size on student productivity during collaborative work on com-plex tasks," Elementary School Journal, Vol. 100, 2000, pp. 183–212.

[18] S. Yasunaga, K. Ejima, and M. Fujikawa, "Development of Discussion Skill Scale," Bulletin of the Faculty of Literature, Kurume University, Vol. 12, 1998, pp. 43-58.

[19] J.C. McCroskey, M.J. Beatty, P. Kearney and T.G. Plax, "The Content Validity of the PRCA-24 as a Measure of Communication Apprehension across Communication Contexts," Com-munication Quarterly, Vol. 33, No 3, 1985, pp. 165-173.

[20] J. Hall and W.H. Watson, W.H., The effects of a normative intervention on group decision-making performance. Human relations, Vol. 23, No. 4, 1970, pp. 299-317.

[21] R. Sakaita, M. Suwa and N. Noguchi, "Exploration of body intelligence to activate commu-nication space: Mr. Kenkoba's communication skill in variety show," The 11th Annual Con-ference on Artificial Intelligence (SIG-SKL), 2011.

[22] P. Bordia, (1997). "Face-to-face versus computer-mediated communication: A synthesis of the experimental literature," The Journal of Business Communication, Vol. 34, No.1, 1997, pp. 99-118.

[23] K.Y.M. Lin, "Chatting online to learn: repair sequences in text-based L1-L2 interaction in CMC," PhD thesis, School of Education, Communication and Language Sciences, Newcastle University, 2015.

[24] J.M. Sykes, "Synchronous CMC and pragmatic development: Effects of oral and written chat," CALICO journal, 2005, pp. 399-431.

[25] Y. Chiba and K. Nishimoto, "Investigation of activation of discussion and possibility of multi-viewpoint by iteration of different forms of communication," Information Processing Society of Japan Research Report, Vol. 2011-HCI-143, No. 5, 2011, pp. 1-7.

[26] I. Daibo, S. Matsuyama and K. Fujiwara, "The Communication behavior and interpersonal perception in task oriented small group -Social skills and interpersonal relationships-," IEICE Technical Report Human Communication Science (HCS), Vol. 111, No 393, 2012, pp. 21-26.

[27] K. Kreijns, P.A. Kirschner and W. Jochems, "Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research," Computers in human behavior, vol. 19, No.3, 2003, pp. 335-353.