Learning Environment for Undergraduates to Enhance
Spontaneous Growth of Communication Skills: Effects of group size*

Eiji TOMIDA*1, Etsuo MIZUKAMI*2, Ikuo MORIMOTO*3 and Hiroko OTSUKA*4

*1 Ehime University, 3, Bunkyo-cho, Matsuyama-shi, Ehime, 790-8577 Japan
*2 Keihanna Research Laboratories, National Institute of Information and Communications Technology, 3-5, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289 Japan
*3 Kansai Gwagkuin University, Uegahara, Nishinomiya, Hyogo, 662-8501 Japan
*4 The Institute of Behavioral Sciences, 2-9, Honmura-cho, Ichihaya, Shinjuku-ku, Tokyo, 162 Japan

Received for publication, April 30, 2010

This study examined the effects of participation order and group size on discussion training of the autonomous dialogue, which means communication skills for discussing spontaneously without third person’s help in a group of about 6 members. Measures for the effectiveness of discussion were self evaluation on rating scales. Participants were 89 undergraduates. They rated their own and others’ discussion in 7 dimensions: honesty, equity, activeness, diversity, depth, management, and cumulativeness. As a result, the activeness component was most facilitated when the students participated after observing others’ activity under the condition that their group size was kept in 6 members all through the sessions or increased gradually. The management component was most enhanced under the condition that their group size was increased gradually.

Key words : higher education, the autonomous dialogue, group size, discussion training

1. INTRODUCTION

1.1. Research Background

Recently, workshop-based learning has attracted a great deal of attention in college education (Otsuka et al. 2009). Workshop-based education is one of the many styles for interactive learning and creativity (Nakano 2003). It enables participants to spontaneously engage in class activities and to learn through social interaction in small groups. This is different from mono-directional knowledge transmission in lectures. The need for learning through interactive discussion in college education, including workshop-based learning, is related to the following current trends: teachers’ individual improvement, corresponding to the rising need for the assurance of quality college education; curriculum reform to be certified by accreditation programs such as the JABEE (an accreditation system for engineering education in Japan served by the Japan Accreditation Board for Engineering Education); government policies for educational reform such as the Good Practice projects operated by the Ministry of Education, Culture, Sports, Science and Technology (MEXT); legislation for the organizational implementation of Faculty Development in 2008; and educational reform for career development such as the ATT by the Ministry of Economy, Trade and Industry (METI).

Even those who do not engage in college education expect college education to include communication training courses. The need for the Literacy in Science and Technology is an example of such expectations (Otsuka et al. 2009). Modern society is becoming too complicated to solve its problems only through small numbers of specialists or public administrators (Kobayashi 2004). In a future society, specialists, administrators, and citizens should communicate with each other to form a collective consensus (Kobayashi 2004). The role of college education is to train citizens to be responsible in society. Developing the ability of students to participate in consensus-building processes is one responsibility of college education. Otsuka, et al. (2009) defined the Autonomous Dialogue as the citizens’ ability to

---

* This paper was originally published in *Jpn. J. Educ. Technol.*, Vol.33, No.4, pp.431–440 (2010)
communicate with each other on their own initiatives, without relying on third-party specialists. The Learning Sciences for Science Learning project (LSSL. 2009) has developed learning programs for higher education to facilitate learning the Autonomous Dialogue.

The trends shown above indicate that current college education has a need for instilling practical communication skills to create citizens who can actively engage in society. The present study examines how to facilitate students’ abilities to communicate effectively in a relatively large group of 5–6 members. Sugie, Sekita, and Yasunaga (2004) reported that although most undergraduates can proceed with discussion in a small group of 2 to 3 members, they find it difficult to do so if the number of members increases.

Despite the greater difficulty, it is necessary for members of modern society to discuss issues in larger groups. For example, in workshops for reaching a political consensus, discussions are held in a group of 5 to 6 members (Land, Infrastructure and Transportation Ministry 2005; Japan Construction Engineers’ Association 2006).

The size of these groups is derived from findings such as those in Foltch-Lyon and Trost (1981). Folch-Lyon and Trost indicated that 6 to 12 members is an adequate group size for a focused group interview. In the workshops in practice, this size is now applied to many meetings for regional development. Aiming to enhance discussion abilities in such relatively large groups, the present study tries to discover learning environments for the development of these discussion skills in a field experiment. In the following section, we review previous works on collaborative learning in higher education on a wider scope; subsequently, we will focus on the effects of different group sizes.

1.2. Previous Studies

Cooperative Learning

One of the major research areas in collaborative learning is cooperative learning. Cooperative learning is an instructional strategy in which two or more students learn through negotiation and shared intentions. CSCL (Computer Supported Collaborative Learning) is the most prominent research topic in cooperative learning. This study focuses on how to enhance collaborative learning on computer networks. Many CSCL systems have been developed, including: Knowledge Forum (Scardamalia and Bereiter 1994), WISE (Linn et al. 2004), Project Board (Nishimori et al. 2005), and ReCoNo (Miyake and Shirouzu 2005). The studies employed design-based research methods to examine the effectiveness of their systems. Although CSCL aims at facilitating learning through collaboration, fostering communication abilities, which is the objective of the present study, is hardly targeted. Even if such an aim was manifested in the paper, few studies have examined the effectiveness of an educational program.

As another strand of cooperative learning research, argumentative processes have attracted the attention of many researchers in recent years (Andriessen 2006, Tomida and Maruno 2004). Most studies of argumentative processes examined the learning process with the theoretical framework of argumentation or facilitative effects of argumentative discourse. Aside from Kun et al. (1997) and Reznitskaya et al. (2001), there have only been a few studies aiming at developing an intervention program that fosters communication abilities. In Kun and Reznitskaya’s studies, they only offered students the opportunity to have discussion with other students.

Know-How in Practice

Some systematic approaches to discussion-centered teaching methods are found in publications on teaching tips. The collection of teaching tips is designed for newly-appointed teachers. Nagoya University published a local version for the first time in Japan (Ikeda et al. 2001). Since then many universities have published their own local versions. Most publications on teaching tips include know-how for introducing discussion in a class and case examples.

More sophisticated know-how is explicated in Instructional Design (ID). ID is a highly systematic approach for designing learning courses that maximize their educational effect. In addition, it applies and considers the findings from educational psychology and other educational sciences. Although it does not specialize in designing collaborative learning, it contains information on how to design learning courses through discussion. For example, Gagne et al. (2007) summarized the basic frameworks for assessment of readiness, instructional flow, feedback for learners, and measurement and evaluation of educational effects.

Direct references to group size in discussion appear in the publications on collaborative teaching methods for college educators. In particular, Cooperative Learning (Johnson et al.
Learning Through Discussion (Rabow et al. 1994), Jacobs et al. (2002), and Sugie (2004) all articulated that a group of four or less members is preferable for group activity. Nakano (2003) noted that each different group size would create its own unique atmosphere in discussions. In a group of two, members can communicate smoothly. With three, synergetic effects tend to take place. In a group of four, a group would develop a cozy atmosphere and yield diversified ideas. In a group of five or more, some members begin to lose motivation.

**Evidence-based Findings:**

Most of the findings discussed above are derived from teaching experiences, not from empirical research. Empirical findings are very beneficial. However, evidence-based design of learning environments is also needed to examine the effectiveness of the teaching method proposed. Previous evidence-based studies on workshop-based teaching can be divided into two categories. The first one treats learning courses as systems and examines their teaching effectiveness and process. To the authors' knowledge, only Nakano and her colleagues' studies (2008, 2009) correspond to this category. They studied the effectiveness of curriculum developed by themselves. This three-year curriculum has the three components of communication (claim, refutation, and rebuttal) as the teaching core. This program prepares learning contents corresponding to students of each grade level. Such a longitudinal design-based study can offer ecologically and practically valid findings because it deals with an actual curriculum for more than one year. However, it is not suitable for examination of individual factors constituting the curriculum.

The other category of research examines the individual factors of teaching effects. For example, in the research area of educational technology, Ozawa et al. (2005) examined how students' reflection on their activity is facilitated by teaching assistants' evaluations and peer evaluation. In terms of group size, Tanaka (2003) concluded from a survey that four is the optimum number for group activity.

**Advancement from Previous Studies**

The previous studies discussed above indicated that the most adequate size of student groups is approximately four, according to the practical experiences of college educators and empirical studies. This conclusion proved true when they examined what size group is the most appropriate for students without any training in discussion. On the other hand, the present study focuses on how to train students to discuss in larger groups. No previous studies have examined this point to the authors' knowledge.

To hypothesize on effective group composition strategy, Nakano (2003) gives some suggestions as noted in the previous section. Deduced from Nakano’s suggestion, the following predictions can be made. When participants talk in dyads at the start, they can form relationships between themselves, if they are given enough time to talk to each other. This would be the basis for further discussion in larger groups. When they discuss in groups of three or four members, they encounter more diverse discussion than in a dyad conversation. This interpretation leads us to the effective strategy for group composition: the group size should increase from two to six gradually. However, Nakano’s suggestion has not yet been examined by experimental study. The present study aims at verifying the effect of group composition strategy to promote discussion ability in a relatively large group.

**1.3. Research Objectives**

The main objective of the present study is to find a group composition method that facilitates the development of discussion skills, especially in a group of approximately six members. As discussed above, a possible composition strategy is to gradually increase the size of group. To examine the effectiveness of this strategy, two more comparative studies were composed and are explained later.

Aside from the main objective, the progress of the learners’ participation in discussion was examined. In one condition, students observed before their own participation. In the other, they participated in the discussion before observation.

**2. PROCEDURES**

**2.1. Teaching Overview**

**Teaching Program**

The experimental program noted here was held as four classes of “Student Guidance,” which is a part of a teaching-training course of a regional national university in Japan. The program was performed in accordance with the Autonomous Dialogue program, which was developed for the purpose of fostering skills in science and technology (LSSL 2009). The core activity in the program is called the fishbowl in which the
participants are divided into the discussants and the observers. After the first discussion, they exchange their roles. Through this procedure, the observers are expected to find areas of improvement and to give feedback to the discussants. In the program, direct teaching is not recommended. Rather, the participants’ reflection prompts them to improve their approach to discussion. In addition, they share the findings of their activities in order to establish goals for the next session (Otsuka et al. 2009).

In this process, it is important to give learners an evaluative framework. The Autonomous Dialogue program adopts the seven evaluative dimensions, which were composed by Mizukami et al. (2008) and Suzuki et al. (2008). Mizukami et al. (2008) and Suzuki et al. (2008) recorded students’ discussions, then had third-party raters evaluate the recordings. After examining the results of factorial analysis on the ratings and the students’ discursive behaviors, the following seven factors were developed:

1. Honesty: To what extent the participants communicate their intentions and listen to others’ statements.
2. Equity: To what extent they discussed in an equitable manner.
3. Activeness: How actively they discussed.
5. Depth: To what extent each idea was examined and compared.
6. Management: To what extent they regulated the direction of the discussion.
7. Cumulativeness: To what extent they utilized their own arguments to reach a cumulative conclusion.

The present study adapted the seven factors for participants’ self-evaluation in the three discussion sessions.

Participants
The participants were 89 university students in their second year. Most of them have participated in introductory courses on presentation and discussion skills in college. Those who skipped one or more classes of the experimental sessions were excluded from analysis.

Experimental Design
Two–factorial between–participant design was employed. The first factor was the group composition strategy. The second was the engagement–observer order of the participants. The group composition factor has three levels: incremental, constant, and U-change. The order factor has two levels: engagement–first and observation–first. In the case of the incremental, they discussed in groups of two in the first session. In the second session, they did so in groups of three. Finally, in the third session, they discussed in groups of six. In the constant level, they discussed in groups of six. In the condition of the U–change, they discussed in a group of six in the first session, three in the second session, and six again in the final session. The two comparison groups were prepared as candidates for effective group composition strategies for the following reasons. The constant group had the longest training duration in six–party communication among all of the conditions and, therefore, had the greatest advantage. The U–change group would show the participants how difficult the large group discussion is in the first session. Then they might be able to learn more in the second small group discussion.

The discussion topics
The participants discussed the pros and cons of cell phone and Internet regulation among elementary school students. The specific instruction was as follows in the engagement–first condition: What does your group think about controlling the use of cell phones in elementary schools? Please discuss to reach a conclusion in accordance with the terms noted below. (1) Draw a conclusion after you examined both positive and negative sides. (2) Clarify your reasons and evidence to support or reject the proposition. (3) Clarify the conditions you support or reject in the proposition. In the observation–first condition, they were given the same question and came to one conclusion: we should control the access to the Internet among elementary school students.

Instructional Schedules
In the first class of the program, participants received an overview; filled out an agreement form; performed an information search activity; and had initial discussions. In the second class, the second discussion was held and the third discussion was held in the third session. All discussion sessions were held for the duration of 15 minutes.

2.2. Design for Discussion Sessions
The discussion session constituted mainly of fishbowl activities, in which the participants discussed alternating their roles as discussant and observer. After the session, discussants and observers rated their discussion on 5–point scales printed on the assessment sheet. Then they noted
the reasons for their ratings and merits and demerits of their discussion techniques. The mean scores of the ratings were calculated and plotted on the assessment chart as shown on Figure 1. On the assessment chart, individual participants’ rating scores and the mean rating scores of both discussants and observers were plotted in different colors. The color variation helped participants recognize the perception gaps between discussants and observers. Reflective talks on the graph’s findings were held with the goal to raise individuals’ awareness of their own discussion styles, as well as to highlight areas of improvement.

3. RESULTS

3.1. Analysis Procedures

Data for analysis were taken from the responses of the questionnaires, which were administered at the end of every session. Each questionnaire contains seven 5-point rating scales that correspond to the seven evaluation dimensions mentioned earlier. Participants rated how difficult those dimensions were after actual participation in the discussions. Specific items for difficulty evaluation are listed below: “it was difficult to listen and speak to others with sincerity”; “it was difficult to make all participants take part in the discussion on an equal basis”; “it was difficult to warm up the group”; “it was difficult to find diverse opinions in the group”; “it was difficult to examine each argument sufficiently”; “it was difficult to control the direction of discussion”; “it was difficult to accumulate the contents of discussion in a constructive manner.” At the end of the final session, the participants were requested to answer the two open-ended questions about the difficulty of proceeding with discussions.

The analysis section is comprised of three parts. Analysis I clarifies the exact difficulties of discussing in a group of six; the data was taken from the responses of the open-ended questions. Analysis II examines what group composition strategy resolves the difficulties that were extracted in Analysis I. Finally, analysis III notes the characteristic verbal behaviors, which were observed in the spontaneously improved groups.

In the present study, improvement in communication ability is evaluated by participants’ subjective ratings and reports. This is because the workshop-based learning is based on the idea that learners can learn through their own awareness. Analysis III, in terms of actual behavior, explores the discourse process in order to grasp the learning effect measured by subjective responses.

3.2. Analysis I: Differences in difficulty in different groups

To clarify the subjective difficulty, the participants felt that, in discussion, their responses in the open-ended questions, which were administered after the discussion sessions, were coded. The coding categories were newly constructed in accordance with the actual responses and based on the seven evaluative points adopted in our project as shown in Table 1.

Since the original evaluative points of honesty and activeness have no corresponding results in this analysis, Table 1 does not contain those categories. Besides the existing evaluative categories, responses concerning lack of time and difficulty in making remarks were often found. Subsequently, the two categories were added.

The instruction for coding procedure was written by the first author (who? Be specific), as well as the initial coding of all responses. The categories in Table 1 are not exclusive to each other. One response can correspond to one or more categories. Consequently, a response was given a binary code for each category. After the initial coding, a third person coded a part of the responses to examine inter-rater reliability. Before actual coding, the additional coder performed trial coding, which clarified the coding instructions’ categories for the authors. Then all responses were coded by the third coder. The reliability indices of Cohen’s k coefficients were .84 for lack of time; .79 for difficulty in making remarks; .93 for equity; .84 for
Table 1. Categories to classify the difficulty the participants experienced in discussion

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Mention of lack of time for discussion.</td>
<td>&quot;I did not have enough time to discuss carefully.&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;There was enough time to hear all members' opinions.&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;It was difficult to find the right time to state my opinion.&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;It was difficult to say anything.&quot;</td>
</tr>
<tr>
<td>Utterance</td>
<td>Mention of difficulty in speaking.</td>
<td>&quot;It was difficult to create equal opportunities to speak.&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;Specific individuals spoke more often.&quot;</td>
</tr>
<tr>
<td>Equality</td>
<td>Mention of the difficulty in keeping equal relationships among group members and opportunity to make remarks.</td>
<td>&quot;There were not very many opinions.&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;We were swept away by the major opinion.&quot;</td>
</tr>
<tr>
<td>Diversity</td>
<td>Mention of difficulty in having diversity and avoiding conformity of opinions.</td>
<td>&quot;The course of discussion easily veered off.&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;Proceeding with the discussion was difficult.&quot;</td>
</tr>
<tr>
<td>Management</td>
<td>Mention of difficulty in giving a direction to the discussion without digressions or in facilitating the development of the discussion.</td>
<td>&quot;The discussion tended to be superficial.&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;Talking in depth was difficult.&quot;</td>
</tr>
<tr>
<td>Depth</td>
<td>Mention of difficulty in deepening the content of the discussion.</td>
<td>&quot;Opinions were too different to put together.&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;It was difficult to consider all opinions.&quot;</td>
</tr>
<tr>
<td>Cumulativeness</td>
<td>Mention of difficulty in cumulating or summarizing the discussion.</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Difficulties Perceived in Different Group Sizes

<table>
<thead>
<tr>
<th></th>
<th>Time</th>
<th>Utterance</th>
<th>Equity</th>
<th>Diversity</th>
<th>Management</th>
<th>Depth</th>
<th>Cumulativeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>2–3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>25</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>members</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>10</td>
<td>25</td>
<td>8</td>
<td>15</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>members</td>
<td>7.7%</td>
<td>9.6%</td>
<td>24.0%</td>
<td>7.7%</td>
<td>14.4%</td>
<td>2.9%</td>
<td>38.5%</td>
</tr>
<tr>
<td>p</td>
<td>0.049</td>
<td>n.s.</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>0.048</td>
<td>n.s.</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

diversity; .71 for management; .79 for depth; and .88 for cumulativeness. Those values indicate sufficient inter-coder reliabilities. The analyses in the following sections are based on the coding results by the first author.

The average number of codes given to a single participant's response was 1.41 (SD = .73). Among all 104 responses, 2 responses were given 4 codes, 6 responses were given 3 codes, 28 responses were given 2 codes, 65 responses were given 1 code, and the remaining 3 responses were not given any code. The content of those non-coded responses are the following: "I got into a jam when none of the members were talkative. When a group had only one person who was the opposite gender, he/she was isolated"; "It was difficult to see my own opinions objectively and required to state my opinions any time in the discussion"; "The discussions became small talk."

Aiming to clarify the difficulty of discussions in groups of six compared to discussions of two or three, The Fisher's Exact Test was performed. Table 2 shows the frequencies of each code and the results of our statistical test.

The categories that have significantly high scores in six-member groups, compared to groups of two or three, were the lack of time, equity, management, and cumulativeness. The result suggests which part of discussion ability should be trained in the lesson. On the other hand, the participants felt ensuring variability of opinion in discussion was more difficult in a dyad than in a larger group, which means that the lack of time, equity, management, and cumulativeness are especially difficult in groups of six.

Although this result was meaningful, the relative degree of difficulty cannot be shown. In addition, some of those perceived difficulties might be resolved only by repeating discussion training. In order to examine those aspects, rating scores on the difficulties were analyzed chronologically, as shown in Figure 2.
Among the seven dimensions, the scores of honesty and depth changed significantly through the repetition of discussion. The former score decreased ($F = 6.866$, $df = 2$, $N = 48$, $p = .002$, partial $\eta^2 = .217$), the latter one increased ($F = 3.375$, $df = 2$, $N = 48$, $p = .038$, partial $\eta^2 = .067$)

The effect sizes of honesty and depth are relatively large and moderate, but the latter score changed the most. This result indicates that rudimentary improvement of honest participation can be brought up by only repetitive participation in discussion. However, the difficulty of the depth of discussion becomes worse with repetition. One can see the score of equity, in relation to other dimensions, was lower.

Following the results described above, from the perspective of participants, it is indicated that management, cumulativeness, and depth of discussion were especially important tasks in 6-member discussion.

3.3. Analysis II: Improvement of perceived difficulties in different group sizes

Analysis I showed the difficult dimensions that needed improvement. In what group size and with what kind of participation method could those difficulties be solved?

As mentioned earlier, perceived difficulty during discussion was measured with the seven evaluative dimensions (honesty, equity, activeness, diversity, depth, management, and cumulativeness). On each score measured with the seven dimensions, the degree of change in difficulty was calculated by subtracting the score measured in the first session from the score in the third session. Therefore, positive change in score meant that perceived difficulty was increased, and vice versa. The effects of the group composition and the participation order were examined using the scores as dependent variables.

The two-way between-participant ANOVA (the group composition vs. the participation order) was performed on each evaluative dimension. The sample size of each condition consists of the following: U-change and engagement-first, 8; U-change and observation-first, 10; incremental and engagement-first, 9; incremental and observation-first, 5; constant and engagement-first, 9; constant and engagement-first, 7. In the activeness and management scores, significant effects and interactions were found.

Activeness has significant main effects on the group composition ($F = 3.466$, $p = .039$, partial $\eta^2 = .124$) and the participation order ($F = 9.050$, $p = .004$, partial $\eta^2 = .156$) but not of their interaction ($F = .328$, n.s., partial $\eta^2 = .013$). Although the sample sizes were small, the effect sizes vary enough. The average score of each level is shown in Figure 3. Multiple comparison by Tukey's HSD method ($\alpha = 5\%$) indicated a significant decline of perceived difficulty in the condition of constant and observation-first and of incremental and observation-first. The difficulty score of the condition of constant and observation-first was significantly higher than any other condition. No other significant difference was found.

The results indicated that the activeness in discussion was encouraged when the participants
discussed in a group of six members all through the sessions or when the number of members increased gradually. In addition, observing other groups’ discussions helped encourage activeness in their own discussion.

The group composition factor has significant effects on the management score ($F = 6.152, p = .004$, partial $\eta^2 = .201$). However, the effect of the participation order and the interaction were not significant ($F = 2.815$, n.s., partial $\eta^2 = .054$; $F' = 1.180$, n.s., partial $\eta^2 = .046$). The effect size of the group composition was the largest in this analysis. Even though the sample size was rather small, there were enough data to gather significant findings. Average scores in each condition are shown in Figure 4. Multiple comparisons by Tukey’s HSD method ($\alpha = 5\%$) indicated that in the conditions of incremental and engagement–first and of incremental and observation–first, the difficulty scores fell significantly more than others. The difficulty score in U-change and engagement–first rose significantly more than others.

The results indicate that management ability in discussion is promoted when the number of members increases gradually.

3.4. Analysis III: Exploring improvement of discussion methods

In Analysis II, the rating scores showed that the difficulties in activeness and management declined in specific conditions. If the decline of the scores means an improvement in discussion manner, what kind of verbal behaviors can be observed in the actual discourse of the improved groups? In only 15-minute discussions, there is not enough time for all participants to be evaluated properly.

However, it is important that the relationship between the discourse process and the difficulty scores is discovered for not only demonstrating the effectiveness of the training program, but for also constructing a concrete instructional guideline. The ability to recognize specific good verbal behaviors would be important for instructors to intervene and encourage in students’ discussions in real classrooms.

In Analysis III, which aims at demonstrating verbal behaviors that relate to the decline of perceived difficulties, we explored discourse processes in a few groups. The groups where the difficulty score of management largely decreased and increased were selected for comparison of their discourse processes. Those groups were selected to maximize the differences in the discourse process.

As the first step in exploration, verbal behaviors corresponding to management and thought to be helpful for facilitation of discussion, were selected. The selection was performed based on the concept of management of discussion which is defined by Mizukami et al. (2008) and Suzuki et al. (2008). The extracted behaviors are listed in Table 3. As shown in Table 3, the specific verbal behaviors were “speaker designation,” “course planning,” “opinion seeking,” and “opinion surveying.”

The utterance corresponding to those categories was identified by the first author and the third coder. The coding was performed after the coders listened to all the recordings of the discussions. The coding results are shown in Table 4. Table 4 contains both coders’ results, since each category’s frequency was too low to calculate $\kappa$ coefficient that establishes coding reliability. Because the maximum frequency was low and many cells indicated zero, the result was not included for statistical examination. As a general tendency, the group of decreased difficulty apparently showed more management-related utterances than the counterpart.

This tendency suggests that verbal behaviors listed in Table 3 relate to the participants’ perception that they have improved their management of discussion. Direct instruction of how to use phrases shown in Table 3 would improve a group who has difficulty varying and changing their discussion.

3.5. Summary

Analysis I revealed that management and cumulativeness are important factors that need improvement, as shown by the contents analysis of
free responses and the time-line analysis on the difficulty scores. However, it is found that only the difficulty score of management decreased through spontaneous reflection.

Analysis II shows the effects of group composition and participation order, which had significant scores in activeness and management. This suggests that activeness in discussion can be promoted under the condition that group size is kept at six from the outset or that the size increases gradually. Management of discussion processes can be improved under the condition that the size increases gradually. Participation order in discussion was not related with the management of discussion.

Analysis III shows some verbal behaviors that might help discussion management in a group that had the lowest difficulty score to a large extent. However, the reliability of the result should be verified by further examinations.

In previous studies, the group composition strategy has been mentioned only as informal tips for teachers and was rarely thought to be an element for a proactive solution for development of discussion skills. In the present article, it is suggested that certain group composition strategies can promote participants’ improvement of discussion skills through their spontaneous reflections. Workshop-based learning is originally a method in which learning takes place by the promotion of a learner’s reflection. The findings of the present study are meaningful in the sense that we actually induced the learners to reflect on their own discussion methods based on the original concept of workshop-based learning. In addition, the finding that management skill can be facilitated by increasing the size of the group gradually is meaningful also in the sense that it can be applied to teaching practices.

4. ISSUES IN THE FUTURE

In Analysis I, management, cumulativeness, and depth were found to be especially difficult elements in a discussion of six-member groups. Analysis II showed that only the management element improved under the specific group composition to some extent. However, the present study did not find any strategies to improve the elements of the cumulativeness and the depth of the discussions. Those remain to be examined in the future.

Analysis III examined behaviors corresponding to management strategy and compared their frequencies between the groups whose difficulty score increased and decreased. As a result, the discourse of the group that experienced decreased difficulty contained some utterances that would facilitate the development of a discussion such as speaker designation, course planning, opinion seek, and opinion survey. However, the results were derived from an explorative analysis based on a small number of cases. Especially in the analysis of utterance category, the frequencies of each

---

**Table 3. Verbal behaviors corresponding to the management category**

<table>
<thead>
<tr>
<th>Category</th>
<th>Definitions and Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker designation</td>
<td>Designating someone who should make a remark, like a chairperson.</td>
</tr>
<tr>
<td></td>
<td>Example: So, to begin, would you tell us your opinion, Mr./Ms. A?</td>
</tr>
<tr>
<td>Course planning</td>
<td>Deciding how a discussion should proceed.</td>
</tr>
<tr>
<td></td>
<td>Example: First of all, how about hearing the negative opinions after the affirmative opinions?</td>
</tr>
<tr>
<td>Opinion seek</td>
<td>Asking other members to state their own opinions.</td>
</tr>
<tr>
<td></td>
<td>Example: Are there any people who have other opinions?</td>
</tr>
<tr>
<td>Opinion survey</td>
<td>Collecting information on members’ attitudes.</td>
</tr>
<tr>
<td></td>
<td>Example: Can you put your hand up if you agree?</td>
</tr>
</tbody>
</table>

**Table 4. Frequencies of utterance categories corresponding to the management**

<table>
<thead>
<tr>
<th></th>
<th>The groups the difficulty score increased</th>
<th>The groups the difficulty score decreased</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 5-1</td>
<td>Group 6-1</td>
</tr>
<tr>
<td>Speaker designation</td>
<td>Coder A</td>
<td>Coder B</td>
</tr>
<tr>
<td>Course planning</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Opinion seeking</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Opinion survey</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
category were low. Longer discussions should be examined. In the future, more data are required for more rigorous inquiry.

ACKNOWLEDGEMENTS

This research was held as a part of the project “Learning Science for Science Learning Science” (FY2005-2009), funded by the RISTEX of Japan Science and Technology Agency (JST). Here we express our gratitude to Mr. Masaki Ishimaru, Ms. Haruna Oono, Ms. Asuka Oguro, and Ms. Azusa Waniwaka for their assistance.

REFERENCES


