Analysis of Students' Peer Assessment Processes

Izumi HORIKOSHI [a*] & Yasuhisa TAMURA [b]
[a] *Graduate School of Information Science,*
Sophia University, JAPAN
[b] *Dept. Information and Communication Sciences,*
Sophia University, JAPAN
*izumihorikoshi@eagle.sophia.ac.jp

1. Introduction

In recent years, there has been an increasing trend for students to present their work during classes and to mutually evaluate each other. We have introduced presentations and peer assessments to classes in the courses of various grades in charge. Our main objectives to adopt peer assessments are as follows:
• Purpose A: To encourage students to better understand the evaluation criteria and to promote the student's presentation ability to improve
• Purpose B: To use it for grades when teachers are unable to see all presentations in large classes
      Meanwhile peer assessments by students have problems of reliability and validity. Fukazawa (2010) organized related researches on the reliability and validity of peer assessments by students. According to Fukazawa, the related researches are roughly divided into following two: peer assessments by students is reliable as evaluation by teachers, and there is doubt about reliability and validity. As an example of the former, Fukazawa cited three studies that showed a high correlation by analyzing the correlation between evaluation by teachers and peer assessments by students (Miller (1996), Hughes (1993), and Stefani (1994)). On the other hand, Fukazawa cited two studies as examples of the latter (Stefani (1994) and Freeman (1995)).
      However, all of these related researches are basically discussing reliability and validity based on the score of peer assessments. In this paper, we focus on this fact. In many previous studies, the reliability of peer assessment is basically determined by whether students' evaluations are consistent with each other, and the validity is basically judged based on whether the evaluation by teacher and by students are consistent. We thought that there might be evaluations in which the evaluated timing was totally different, even though the evaluations are regarded as being "consistent with each other" in previous method of the score-based judgment.
      For this reason, we set the purpose of this research as to discuss the necessity of analyzing evaluation timing and temporal behavior in students' peer assessment processes. In this research, we focused on students' peer assessment of oral presentation.

2. Materials and Methods

2.1 Data Acquisition

In order to acquire evaluation timing in peer assessment, we developed a peer assessment tool with the function to detect students' temporal behavior. This tool is a kind of web-based form implemented in HTML, JavaScript and PHP. This form consists of list of evaluation items and radio buttons corresponding each score (1-5). When reviewers evaluate presenters, they select score of each item by clicking these radio buttons. Reviewers can change their scores until they click submit button. Assessment process logs and other information are sent into Learning Record Store (LRS), when each reviewer clicked the radio buttons or submit button. In order to distinguish between the evaluation log of the intermediate evaluation sent by pressing the radio button and the log of the final evaluation sent by pressing the send button, we defined "submit type": "onClick" is the former and "submit" is the latter. The items stored in LRS are as below: Reviewer's student ID, presenter's student ID, date and time, evaluation item ID, score, and "submit type".

2.2 *Target Course, Unit, and the Number of Participants*

We set one target course for the experiment. This course was held in Sophia University, Japan, and one of the authors was in charge. The target course was mainly for students of the third year, and the number of participants was 72 students. The students were divided into six groups, and one student selected from each group made a presentation. Each group was given a total of 15 minutes, consisting of a 10-minute presentation and a 4-minute question and answer session. Six groups presented the presentation in one class. In the first week, group A to F made presentations, and in the second week group G to H made presentations. Of these two classes, we targeted the first week as an experiment. In the experiment, we requested the students to conduct peer assessment using the developed tools and got a log of the evaluation behavior.

3. Results and Discussion

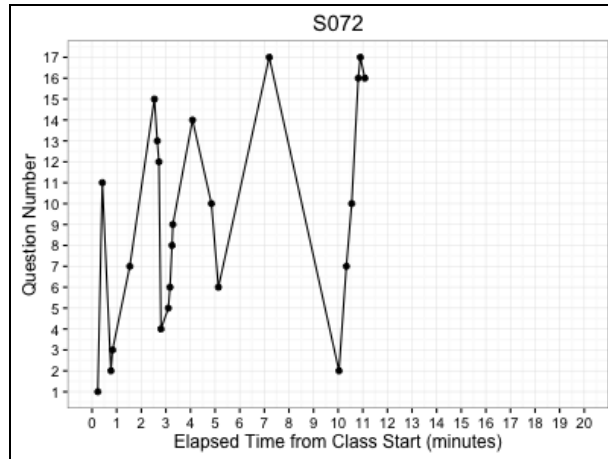### 3.1 *Examples of Evaluation Behavior during Evaluation for One Group*

Examples of plotting log of the acquired evaluation behavior are as shown in Figure 1. The horizontal axis of the figure is the elapsed time from the start of the presentation (maximum 20 minutes), and the vertical axis is the evaluation item number (question number: 1-17). The three examples in Figure 1 are the logs of the evaluation for the group A.
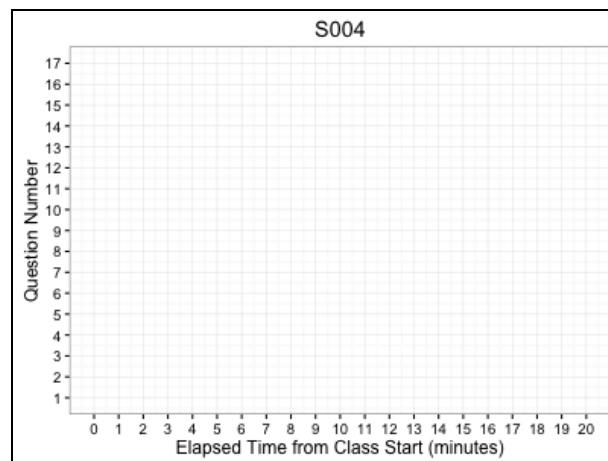
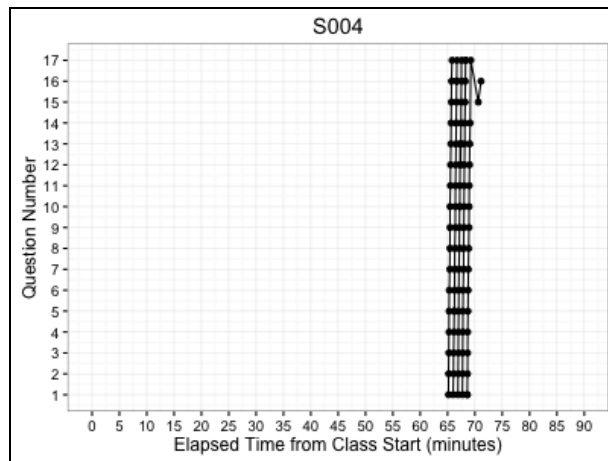(a) Log of student S004
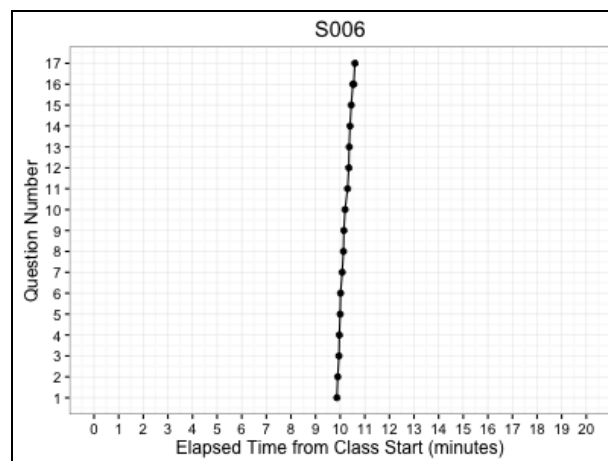
(b) Log of student S006

(c) Log of student S072

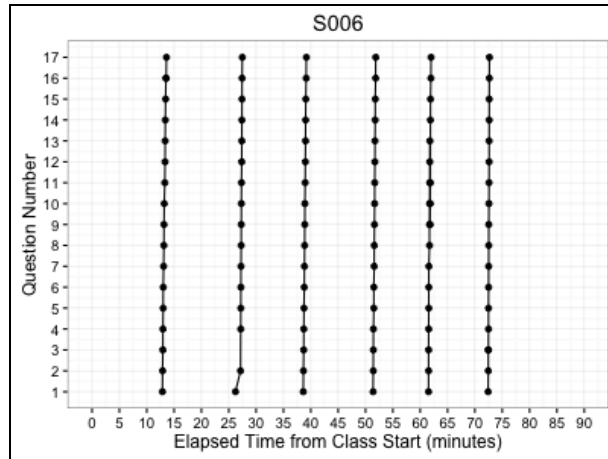Figure 1. Examples of Evaluation Behavior during Evaluation for One Group

As it can be seen from the Figure 1-(b) and (c), while some students, like student S006, evaluated in a short time, some students, like student S072, came back and forth, or changed the score once evaluated. Also, as in the graph of the student S004 (Figure 1-(a)), there were some students whose graphs were empty. Such students are students who did not make an evaluation during the presentation.

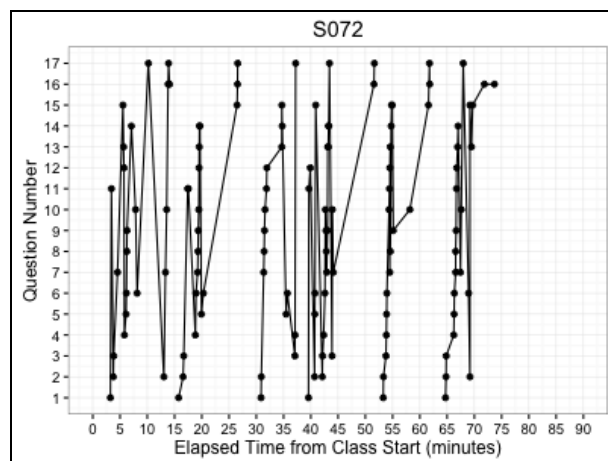## 3.2 Example of Evaluation Behavior during Evaluation for All Group

Figure 2 shows logs of evaluation behaviors of evaluation for all groups on a single graph. The horizontal axis of the figure is the elapsed time from the start of the class (maximum 90 minutes).

(a) Log of student S004

(b) Log of student S006

(c) Log of student S072

Figure 2. Examples of Evaluation Behavior during Evaluation for All Group

As it can be seen from the Figure 2, the evaluation behavior is greatly different for each student. Student S004 evaluated all the groups at the end of the class regardless of which group is presenting the presentation at the moment. Student S006 is similar to student S004 in that he evaluated in a short time, but he or she evaluated during each group's presentation. Student S072 seems to have evaluated by seeing and listening to the presentation in all groups.

3.3 All Students' Evaluation Behavior during Evaluation for All Group

Figure 3 shows all students' logs of evaluation behaviors of evaluation for all groups on a single graph. The horizontal axis of the figure is the elapsed time from the start of the class (maximum 90 minutes).

As it can be seen from the figure 3, there are other students like the student S004 who evaluates after a long time since the presentation was over. Furthermore, in the evaluation for group F, there are students who end the evaluation before the presentation starts. As a worse case of that, the student represented by the line at the right end of the figure (S004) has erroneously ended the evaluation of group G, as he or she completed evaluation of many groups in a short time. Despite this group's presentation scheduled for the next week.
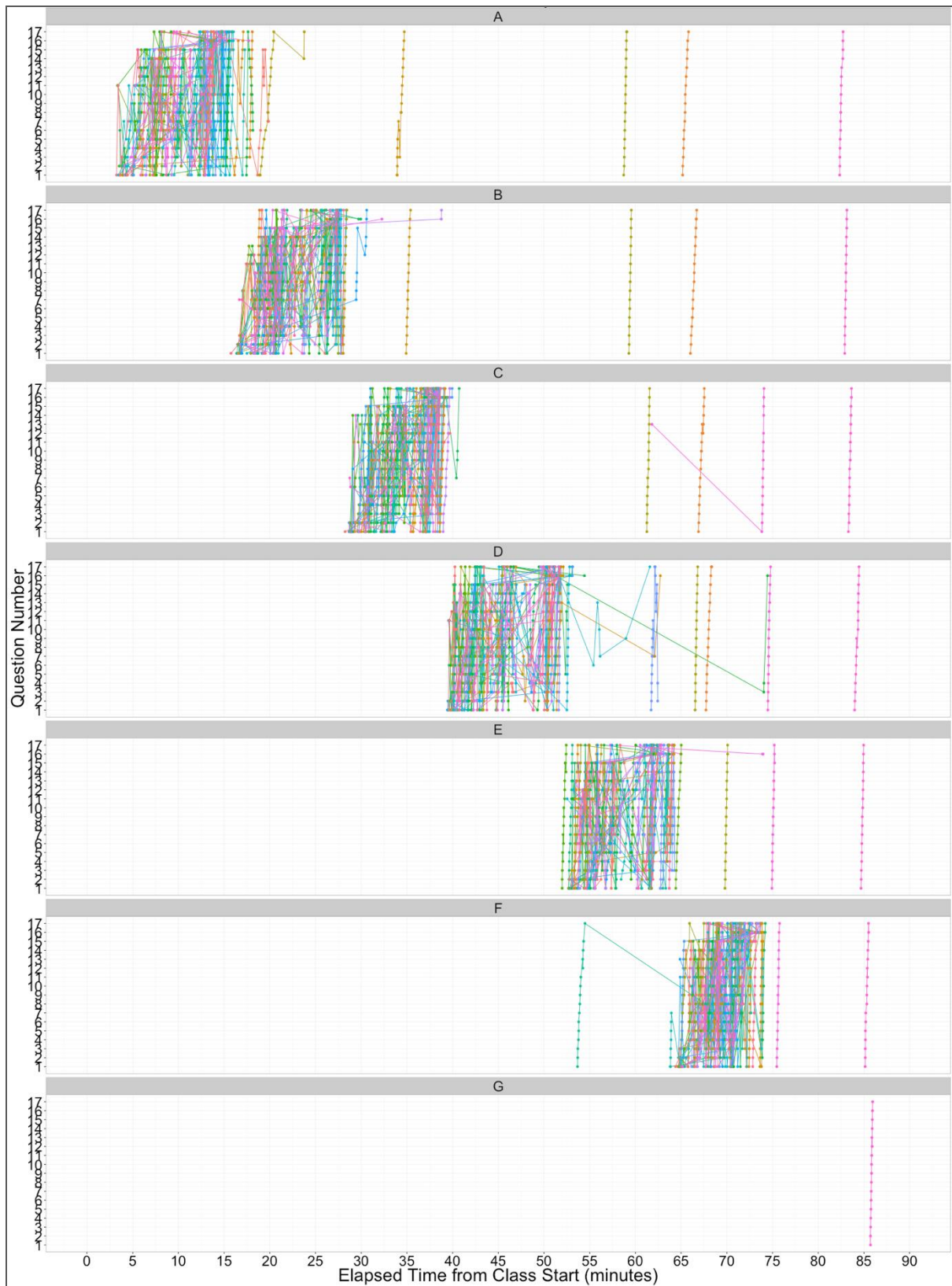
Figure 3. All Students' Evaluation Behavior during Evaluation for All Group

4. Conclusion and Future Works

The purpose of this research was to discuss the necessity of analyzing evaluation timing and temporal behavior in students' peer assessment processes. As the results show, we found a characteristic behavior that students evaluated after a long time since the presentation was over, or on the contrary, before the presentation starts. Other students evaluated many items in a short time.

Conventionally, reliability and validity of peer assessment were often judged by the degree of consistent of assessment scores. However, by visualizing the evaluation timing, we found that students' peer assessments include those evaluated during time periods not during presentation. As suggested this time, if the students did not seriously evaluate, no matter how much the evaluation results are consistent, it might be mere coincidence.

We only touched on the outline of evaluation timing in this paper, for analysis on more detailed evaluation behavior remains as future works. In the future, we would like to clarify the followings: (a) whether students understand the evaluation timing suitable for each evaluation item, (b) relationship between degree of consistency of score and evaluation timing, comparison between teacher's evaluation behavior and student's evaluation behavior, (c)conditions that makes it difficult to implement appropriate mutual evaluation.

Acknowledgements

References

Freeman, M. (1995). Peer assessment by groups of group work. *Assessment & Evaluation in Higher Education*, *20*(3), 289-300.

Fukazawa, M. (2010). Validity of Peer Assessment of Speech Performance. *ARELE: annual review of English language education in Japan*, 21, 181-190. (in Japanese)

Hughes, I. E., & Large, B. J. (1993). Staff and peer-group assessment of oral communication skills. *Studies in Higher Education*, *18*(3), 379-385.

Miller, L., & Ng, R. (1996). Autonomy in the classroom: Peer assessment. *Taking control: Autonomy in language learning*, 133-146.

Stefani, L. A. (1994). Peer, self and tutor assessment: relative reliabilities. *Studies in Higher Education*, *19*(1), 69-75.