So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Effects of Using Rubric Forms on Evaluation Behavior in Student Peer Assessment

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

Abstract: This study aimed to clarify the effect of using rubric forms on evaluation behavior by analyzing the influence of differences in peer assessment form design on student evaluation behavior. We prepared two types of evaluation form, a rubric form and a non-rubric form and compared the evaluation times. The results revealed that the behavior of evaluating multiple items in a short period of time and in the numerical order of evaluation items was more often observed when using the non-rubric form. In addition, the medians of evaluation time were higher among students who used the rubric form.

Keywords: Peer Assessment, Rubric, Evaluation Behavior, Learning Analytics

1. Introduction

Along with the spread of active learning classes, peer assessment has become popular. While peer assessment has various advantages, its reliability and validity are a major concern (Fukazawa, 2010). While conventional studies have mostly utilized evaluation scores to address this issue (Hughes & Large, 1993; Stefani, 1994; Freeman, 1995), research that analyzes peer assessment from the perspective of evaluation behavior is still scarce (Horikoshi & Tamura, 2018).

As the results in previous papers, we have revealed that each student has different evaluation behaviors. For example, some students took a long time to evaluate while others evaluated in a short time. Some students evaluated in the order of evaluation items while others did not. However, the causes or interpretations of these characteristic evaluation behaviors have not been established.

The objective of this paper is to clarify the effect of using rubric forms on evaluation behavior. In order to achieve this objective, we set the following hypotheses:

Hypothesis A: The behavior of evaluating multiple items in a short period of time in the numerical order of evaluation items is more often observed when using a non-rubric form than when using a rubric form.

Hypothesis B: A longer evaluation time is observed when using a rubric form than when using a non-rubric form.

2. Methods

2.1 Data Acquisition

In this study, a Web evaluation form was implemented and used to obtain the evaluation time for each evaluation item in a student peer assessment. This online form records the evaluation time and score for each assessment item before and after the "submit" button is clicked. This form was used in previous studies by the authors of this paper. However, in order to verify Hypotheses A and B, two types of forms were created for this study: a rubric form and a non-rubric form (Figure 1).

Evaluation Item	Score (1-5)				
Category 1: Structure/content **Q1: Was the title of the presentation appropriate?**	○ There was NO title.	○ The title DID NOT match the presentation.	○ The title contained words UNRELATED to the presentation.	○ The title was CONSISTENT with the content of the presentation.	○ The title was attractive and motivate audience to hear the presentation.
Q2: Did the presenter give an appropriate introduction before moving to the main topic, and was the introduction useful for the audience in understanding the presentation?	○ There was no introduction and suddenly entered the main subject.	○ The introduction did not consider the prior knowledge of the audience.	○ The introduction took into consideration the prior knowledge of the audience.	○ The introduction helped the audience understand before the main subject.	○ The introduction was attractive and the audience wanted to hear ahead.
Q3: Was the main point of the presentation clear?	○ I didn't know what the presenter wanted to say.	○ I could partially understand the point of the presentation.	○ I could almost understand the point of the presentation.	○ The point of the presentation was clear.	○ The point of the presentation was expressed in easy-to-understand words.
Q4: Was the presentation well-structured and organized?	○ Introduction, main thesis, concrete examples, conclusions, etc. were not organized.	○ Introductions, main themes, examples, conclusions, etc. were not logically consistent with each other.	○ The presentation was partially structured and organized logically.	○ The presentation was well-structured and organized logically.	○ The relationships among the introduction, the main thesis, concrete examples, conclusions, etc. were carefully explained.
Category 2: Presentation technique	○	○	○	○	○

(a) Rubric form

Evaluation Item	Score (1-5)
<Category 1: Structure/content> Q1: Was the title of the presentation appropriate?	1 ○ ○ ○ ○ ○ 5
Q2: Did the presenter give an appropriate introduction before moving to the main topic, and was the introduction useful for the audience in understanding the presentation?	1 ○ ○ ○ ○ ○ 5
Q3: Was the main point of the presentation clear?	1 ○ ○ ○ ○ ○ 5
Q4: Was the presentation well-structured and organized?	1 ○ ○ ○ ○ ○ 5
<Category 2: Presentation technique>	

(b) Non-rubric form

Figure 1. Two types of peer assessment forms used in this research (17 items in total, excerpts from the screen)

2.2 Experimental Design

Participants in this study were students of the "Information Literacy" course at Sophia University, Japan. In order to verify Hypotheses A and B, we conducted experiments with two conditions and collected logs in the peer assessment. The target class was divided into two parts, half of which used the non-rubric form in the first week of the experiment and the rubric form in the second week. The other half of the class used the forms in reverse order. Students in this course were divided into 12 groups, of which six gave presentations each week and carried out peer assessment.

3. Results and Discussion

3.1 Acquired Data

Table 1 shows the number of reviews for each presenter group. It should be noted that though the groups which made their presentations in the target week of this paper were only groups A to F, there was reviewer who erroneously evaluated group G, and therefore group G also appears in Table 1. Section 3.2. verifies Hypotheses A based on Group A, while Section 3.3 verifies Hypotheses B based on groups

A to F.

Table 1. *Number of reviews for each presenter group*

Presenter Group	**Number of Reviews**
Group A	53
Group B	53
Group C	56
Group D	51
Group E	53
Group F	49
Group G	1

3.2 Hypothesis A

To verify Hypothesis A, the evaluation behavior for each type of peer assessment form for group A was analyzed (Figure 2). In Figure 2, each graph shows the evaluation behavior of one reviewer. The vertical axis of the graphs shows the number of the evaluation item (Q1-Q17), while the horizontal axis shows the elapsed time from the start of the presentation (0 to 20 minutes). The dark shaded area of the graphs shows the presentation time, while the light shaded area shows the Q&A time. The reviewers who evaluated before the presentation started or after 20 minutes are not included in Figure 2.

To verify Hypothesis A, the evaluation behavior graphs shown in Figure 2 were classified based on whether or not they showed the behavior of evaluating multiple items in a short period of time in the numerical order of evaluation items. The graphs were classified by visual confirmation according to the following criteria.

- Evaluation in a short period of time: Includes the reviewers who completed the evaluation within one or two minutes. Reviewers who changed only one item after a certain time like S035 can be included in this category.
- Evaluation in the numerical order of evaluation items: Includes the reviewers who evaluated in one direction, in descending or ascending order. Reviewers who changed only one item like S026 can be included in this category.

Table 2 shows the results of the classification. Both types of forms were assigned to participants in approximately the same number; however, because of absence and agreement/disagreement with the study, there was a difference in numbers. Despite this difference, as shown in Table 2, it appears that the target behavior was more frequent in the non-rubric form, while other behaviors were more frequent in the rubric form. In other words, it is suggested that the type of form may have an influence on whether or not students show the behavior of evaluating in a short period of time in the numerical order of evaluation items. In order to statistically verify this, a $\chi 2$ test was performed and a significant difference was found ($\chi 2(1)= 4.1346$, $p<.05$).

Table 2. *Comparison of target behaviors (Group A)*

	Evaluation in a short period of time and in the numerical order of evaluation items	Other evaluation behaviors	Total
Non-rubric form	15	12	27
Rubric form	5	17	22
Total	19	30	49

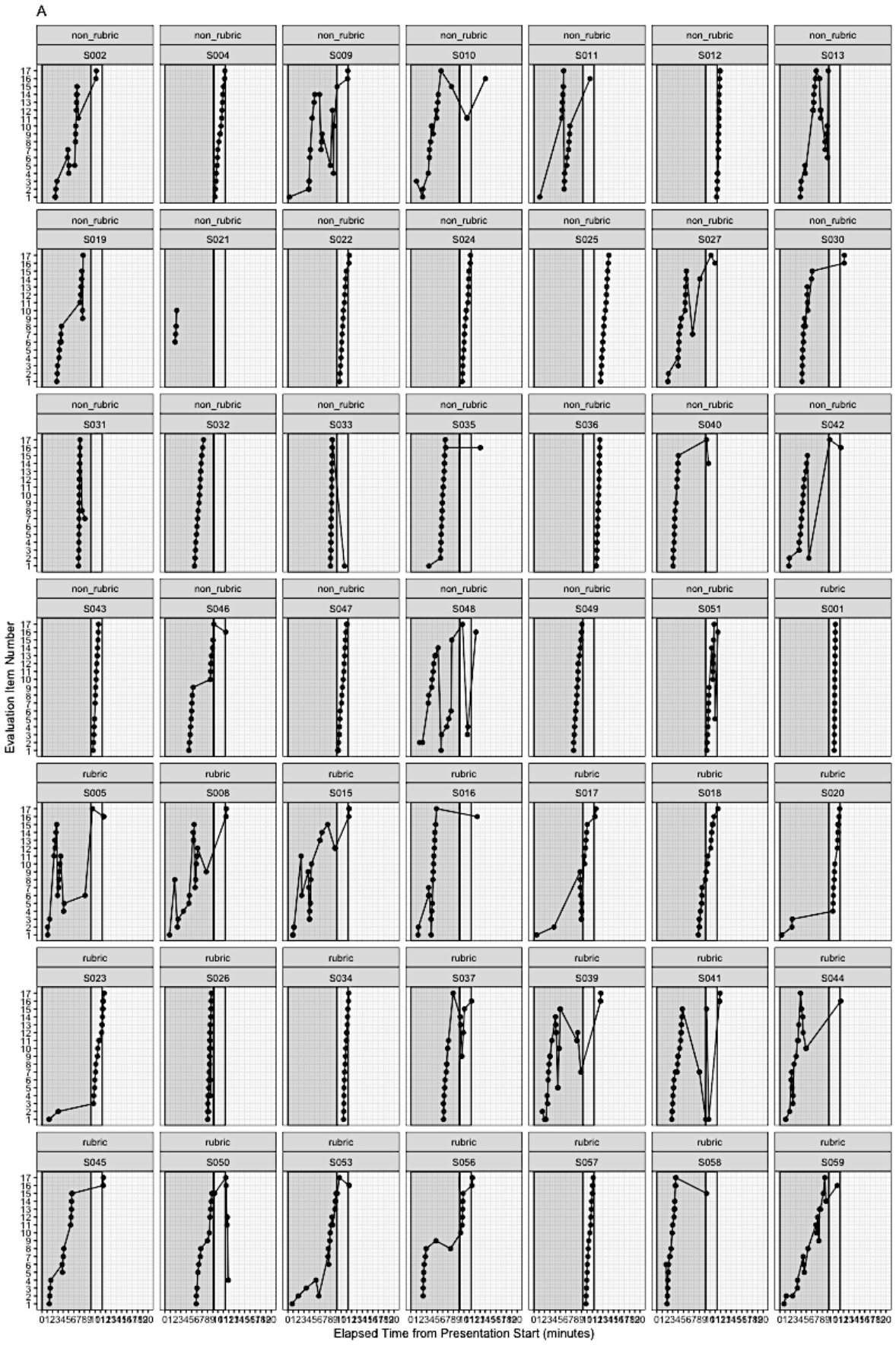
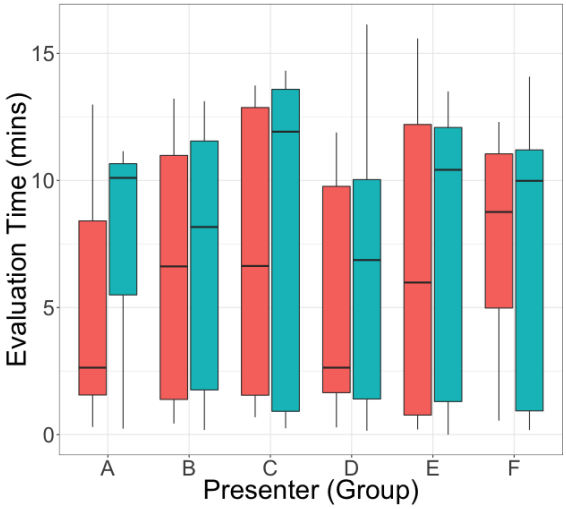


Figure 2. Evaluation behavior for each type of form (Group A)

3.3 Hypothesis B

To verify Hypothesis B, the evaluation time (ET) was compared between the two types of evaluation form. In this study, ET refers to the time difference between clicking on the radio button of the first evaluation item and clicking on the radio button of the last evaluation item.



left bars: non-rubric form users
right bars: rubric form users

Figure 3. Comparison of"Evaluation Time (ET)" (Groups A-F)

Figure 3 shows the distribution of ETs. The left bars in each group refer to the non-rubric form users, and the right bars refer to the rubric form users. As for the distribution, there is a large variation; there are reviewers in the non-rubric form user group (left bars) who took considerable time to evaluate and there are also reviewers in the rubric form user group (right bars) who evaluated in a short period of time. On the other hand, all ET medians were higher in the rubric form users than in the non-rubric form users in all evaluation groups. In sum, ET varies depending on the reviewers. However, the median shows that the evaluation using the rubric form takes longer.

4. Conclusion

This study aimed to clarify the effect of using rubric forms on evaluation behavior. The results of the experiment clarified that the behavior of evaluating multiple items in a short period of time and in the numerical order of evaluation items was more often observed when using the non-rubric form. In addition, the median of students' evaluation time was higher in the condition of using the rubric form.

As a future task, we would like to clarify the reasons why students who used the rubric form took longer to complete the evaluation. This may be because such students tried to make a serious evaluation or took longer to read the form because the rubric form had more characters.

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. *Annual review of English language education in Japan*, 21, 181-190.

Horikoshi, I., & Tamura, Y. (2018) Feature Extraction of Learners' Motivation from Peer Assessment Process Logs. In *Proceedings of the 26th International Conference on Computers in Education*, 352-354.

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

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