

# Feature Extraction of Learners' Motivation from Peer Assessment Process Logs

Izumi HORIKOSHI<sup>a\*</sup> & Yasuhisa TAMURA<sup>b</sup>

<sup>a</sup>*Graduate School of Science and Technology, Sophia University, Japan*

<sup>b</sup>*Department Information and Communication Sciences, Sophia University, Japan*

\*izumihorikoshi@eagle.sophia.ac.jp

**Abstract:** In this paper, we attempted to analyze learners' motivation based on peer assessment process logs. In many previous studies, the reliability or validity of peer assessment was discussed based on the students' peer assessment score. In contrast, we have discussed the necessity of analyzing evaluation-timing and temporal behavior in the assessment processes. We conducted an experiment and the following feature variables, such as "Evaluation Time" or "Standard deviation of the scores of the evaluation items", were calculated from the collected log data. We analyzed the relationships among these feature variables. As a result, we found some positive and negative correlations among them. Furthermore, when conducting continuous evaluations during a single lesson, these trends were more noticeable in the evaluation at the end of the lesson than at the beginning of the lesson. We expect that the characteristics of the evaluation behavior would be useful as indicators of the lack of motivation in terms of serious evaluation in peer assessment evaluations.

**Keywords:** Peer assessment, Web survey, Paradata, Learning analytics

## 1. Introduction

With the spread of active learning style classes in recent years, there have been increasing cases of making presentations in classes and evaluating them with peer assessment. The purpose of peer assessments by students vary depending on the teacher or the lesson. However, according to Topping (2009), the assessment purposes can be roughly divided into the following two types: (1) cognitive or metacognitive gains and (2) time saving.

Peer assessments have such benefits. However, peer assessments conducted by students can also have serious problems in terms of reliability and validity. In many previous studies, the reliability of the peer assessment was determined based on the consistency between the students' evaluations, and its validity was judged based on the consistency between the teacher's evaluation and the students' evaluation. We thought that there might be evaluations in which the evaluated timing would be entirely different, even though the evaluations were regarded as being "consistent with each other" in the previous methods of the score-based judgment. For this reason, we have been conducting research by acquiring logs of peer assessment processes and analyzing the timings of the evaluations.

We have not been able to find any researches similar to ours in the field of educational engineering. Meanwhile, it became clear that many preceding social science research studies had focused on this kind of behavior. With the spread of computers and the Internet, their use of surveys that employ computers and the Internet has also expanded. They enable the collection of the data generated automatically by the data collection process, and these "data from survey data acquisition processes" are called "paradata." (Couper & Kreuter, 2013). In the various "paradata" research studies, the "Response Time" has frequently been the main focus, and it is also similar to our target data for the peer assessment behavior study. According to works of literature, it is said that "Shorter response times" are associated with "Lack of motivation to answer accurately as the survey continues" (Bassili, 1996; Yan & Tourangeau, 2008)

Based on these findings in the analysis of survey response process logs, we hypothesized that learners' motivation can be detected even from peer assessment process logs. The expected

benefit of this research is to find characteristics of evaluation behavior that can be used to indicate lack of motivation in terms of serious evaluation during peer assessments.

## 2. Materials and Methods

In order to acquire evaluation-timing data during peer assessment, we have used a web-based form as the peer assessment tool (Horikoshi, 2017). This tool has a function that can detect students' temporal behavior, and it is implemented in HTML, JavaScript, and PHP. The items stored in the server were as follows: the reviewer's student ID, the presenter's student ID, date and time, evaluation item ID, and score.

The target course was held at Sophia University, Japan. The number of participants was 72, and they were divided into 12 groups. Each group was given a total of 14 minutes, which consisted of a 10-minute presentation and a 4-minute question and answer session. Six groups presented the presentation in one class (groups A-F and G-L). This paper utilizes data from this first week. During the experiment, we requested the students to conduct peer assessment using the developed tools and acquired a log of their evaluation behavior.

The following variables were calculated from the collected log data of the peer assessment process. We will visualize these variables and analyze the relationships between them.

- **Evaluation Time (ET):** Diff-time between clicking the radio button of the first evaluation item and the last evaluation item
- **Click Count (CC):** Total number of times the radio button of the evaluation item was clicked
- **Mean of the score (sM):** Average score of all the evaluation items
- **Standard deviation of the score (sSD):** Standard deviation of the scores of all the evaluation items
- **Mean of the evaluation timestamp (tM):** Average of elapsed time since the start of the presentation
- **Standard deviation of the evaluation time stamp (tSD):** Standard deviation of the timestamps for all the evaluations

## 3. Results and Discussion

Figure 1 shows correlations among the six feature variables. In Figure 1, the green lines indicate positive correlations, and the red lines indicate negative correlations. In addition, the thicker and darker lines show that the absolute values of the correlation coefficient are larger, and the thinner lines show that the absolute values of the correlation coefficient are smaller. Figure 1-(a) shows the correlations of Group A (where the reviewer conducted the evaluation at the beginning of the class) and Figure 1-(b) shows that of Group F (where the reviewer conducted the evaluation at the end of the class).

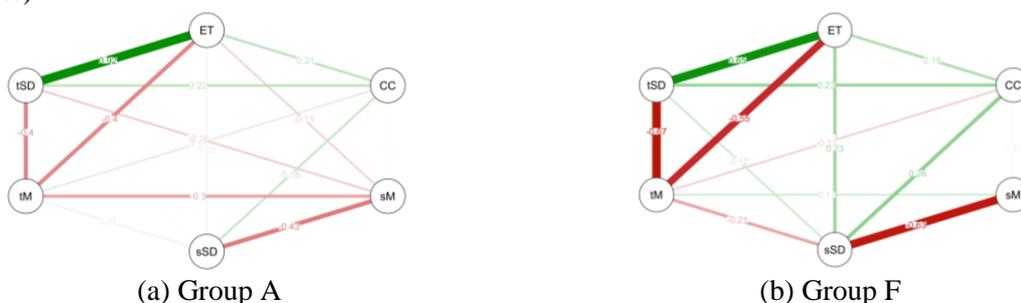


Figure 1. Correlation coefficient between feature variables (Group A and Group F)

As shown in Figure 1, there are some positive and negative correlations among the six feature variables. In this paper, we only focus on three negative correlations (between tM-ET, tSD-tM and sSD-sM) and one positive correlation (between sSD-ET) because of space limitations.

- Negative correlation between tM-ET and between tM-sSD

The negative correlation between tM-ET indicates that, the reviewer who spent less time on the evaluation (= ET is smaller) had a larger mean of the evaluation time stamp (= tM is larger). This is because it would have been after the passage of a long time since the presentation was over when the reviewer, who did not spend time on the evaluation, performed his or her evaluation. The negative correlation between tM-sSD came from the positive correlation between ET-sSD. In other words, these two correlations represent the same trend.

- Negative correlation between sSD-sM

This negative correlation indicates that the reviewer who did not give the same score to many items (= sSD is larger) had a small average score (= sM is smaller). This is because the reviewer who gave the same score for many items tended to give high scores such as 4 or 5.

- Positive correlation between sSD-ET

This positive correlation indicates that the reviewer who spent a longer period of time on the evaluation (= ET is larger) had a larger variation in the score (= sSD is larger). It is very interesting that a correlation appeared between feature variable of the score (sSD) and the feature variable of the evaluation time stamp (ET). It is assumed that, as the reviewer's seriousness in performing the evaluation increases, the lengths of the ET and the sSD increase as well. This tendency did not appear in the group that the reviewer evaluated at the beginning of the class (Group A), but it appeared in the group that the reviewer evaluated at the end of the class (Group F). This result suggests that the continuous evaluations of peer assessments may affect the student's evaluation behavior.

#### 4. Conclusion and Future Works

The objective of this paper was to verify the hypothesis whether “learners’ motivation can be detected from peer assessment process logs” or not. As a result, we found the characteristic evaluation behavior that could be used as indicators of the lack of motivation to conduct a serious evaluation. Specifically, the following trends can be used as indicators of “not conducting serious evaluation”: Short Evaluation Time (ET), large Mean of the score (sM), small Standard deviation of the score (sSD), large Mean of the evaluation time stamp (tM), and small Standard deviation of the evaluation time stamp (tSD). Furthermore, when conducting continuous evaluations during a single lesson, we found that these trends were more noticeable in the evaluation at the end of the lesson than at the beginning of the lesson. We expect that the characteristics of the evaluation behavior would be useful as indicators of the lack of motivation in terms of serious evaluation in peer assessment evaluations.

With regard to future research works, the following two tasks remain: (1) confirm the reproducibility of this research (determine whether this trend will appear in a similar way during peer reviews in other contexts) and (2) verify the correlation between the indicators we found this time and other highly reliable indicators.

#### References

- Bassili, J. N. (1996). The how and the why of response latency measurement in telephone surveys. In N. Schwarz and S. Sudman (Eds.), *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, 319-346. San Francisco: Jossey-Bass.
- Couper, M. P., & Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *In Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1), 271-286.
- Horikoshi, I., & Tamura, Y. (2017). Analysis of Students' Peer Assessment Processes. In Yang, J. C. et al. (Eds.), (2017). *Extended Summary Proceedings of the 25th International Conference on Computers in Education. (ICCE 2017)*, 25-30.
- Topping, K. J. (2009). Peer assessment. *Theory into practice*, 48(1), 20-27.
- Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *In Applied Cognitive Psychology*, 22(1), 51-68.