Classification and Analysis of Learners' Proficiency Level in Marker Use based on Learning Logs

Taito KANO[a*], Izumi HORIKOSHI[b], & Hiroaki OGATA[b]
[a]*Graduate School of Informatics, Kyoto University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*kano.taito.87a@st.kyoto-u.ac.jp

Abstract: As the use of Learning Analytics has become widespread, real data has started to accumulate, and evaluation using such data has been practiced and shown to be effective globally. When data are used for evaluation, each teacher is required to combine and evaluate appropriate data according to the situation of each learner in their charge. However, teachers are not yet familiar with analyzing data; therefore, effective evaluation methods must be proposed and shared from Learning Analytics to encourage them to take the initiative in using data for evaluation. In this study, as an example of an evaluation method, we defined an evaluation index, "*proficiency level*," for the use of markers on e-books using Zimmerman's Multilevel model. The data used were three months of study log data of 112 Japanese second-year junior high school students stored in a learning analysis platform called Learning and Evidence Analytics Framework. Subsequently, we focus on Active Reading (AR) English language learning which was held in the last week of that period. By comparing the reading speed of English text, Word Per Minutes, measured before and after Active Reading, we were able to clarify the differences in learning effects according to "*proficiency level*". This allowed us to confirm the validity of "*proficiency level*" as an evaluation index. In the future, we will work on the creation of a higher-order evaluation index from both data and theoretical perspectives, and it will enable teachers to support evaluation that responds to each learner.

Keywords: Learning Analytics, formative assessment, multilevel model

1. Introduction

With the recent digitalization of education, Learning Analytics has been put to practical use in real educational environments, and many types of learning log data have been accumulated (Sung, Chang, & Liu, 2016). "Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" (Ferguson, 2012, p.305).

Recently, these learning log data have begun to be used as learning evaluations (Faber, Luyten, & Visscher, 2017). This enables immediate evaluation that reflects each individual's situation in detail, which is difficult to achieve through test scores or peer evaluation (Sung et al., 2016). When log data are used for evaluation, each teacher is required to combine and evaluate appropriate data according to the situation of each learner in their charge. However, most evaluations in Japan today are still conducted on paper. Therefore, it is not easy for teachers to analyze simple features in existing log data, such as learning time or number of markers, and generate higher-order evaluation indices that reflect the learning situation. Consequently, it is necessary to share effective evaluation methods that utilize data among teachers. Learning Analytics will need to create and propose highly interpretable evaluation measures, correlated with learning outcomes, to support teachers.

Many of the features that have been focused on in Learning Analytics up to now were simple features that were obtained by summing raw log data. When these features were used for evaluation, it was necessary to correlate them with grades or regress them on external indicators, and then select those that were likely to be relevant to the evaluation. However, these features are not appropriate for teachers to use as evaluation indicators because of their low interpretability. In this study, we aimed to create a theoretically interpretable evaluation index.As a highly interpretive assessment, Zimmerman (2000) used a multilevel model to describe the proficiency of knowledge and skills acquired by learners through self-regulated learning (SRL). In this model, four levels of proficiency in SRL skills are assessed. We consider that the model could be applied similarly to evaluate SRL skills on e-book readers, such as the use of markers for highlighting. Formative assessment methods based on this theory have been tested in field settings and found to be effective in practice (Granberg, Palm & Palmberg, 2021).

To create a new higher-order assessment instrument, we focus on Active Reading (AR) English language learning, which requires the use of markers on the e-book reader. The students' use of markers in this class could be divided into multiple levels. We analyzed the logs of markers, which are simple features, to determine the *proficiency level* of marker use for each student based on four calculated student characteristics: *Performability*, *Applicability*, *Extensibility*, and *Continuity*. These were created according to a multilevel model. To examine the relationship between the obtained *proficiency levels* and learning effectiveness, we examine whether there are significant differences in reading speed in English (Word Per Minute, WPM) between the different *proficiency levels*. Although we focus on the use of markers in this study, in the future, we expect to generate higher-order evaluation indices for other functions to support evaluations that correspond to each student's learning. The research questions addressed in this study are as follows:

RQ1: For the four characteristics of Zimmerman's multilevel model, what value would be the criterion for determining whether it is high or low?
RQ2: Is marker *proficiency level* related to learning effectiveness?
RQ3: Compared to the number of markers, is *proficiency level* more indicative of a student's learning effectiveness?

2. Methods

2.1 Learning Evidence Analysis Framework

A Learning Evidence Analysis Framework (LEAF) system is a platform for using, managing, and collecting educational big data to support teachers' and students' learning in response to information and communication technologies (Ogata, Majumdar, Yang, & Warriem, 2022). Learning Management System (LMS) and e-book reader (BookRoll) logs are stored in the Learning Record Store (LRS) and used in the Learning Analytics Dashboard (LogPalette) to support teaching and learning.

Specifically, BookRoll allows students to view materials shared by teachers on the device. The student can highlight lines with the marker function and remove the highlighted lines. These interactions are saved in the log as 'ADD MARKER' and 'DELETE MARKER' in the LRS.

In the LEAF system, features (indicators) representing characteristics of learning, as shown in Table 1, are extracted from the data stored in the LRS. These indicators are used in LogPalette to support teaching and learning.

In this study, we focus on marker counts, obtained by summing the counts of 'ADD MARKER' and 'DELETE MARKER,' as an example of the simple features described in Section 1.

Table 1. *Example of Indicators*

Navigation Indicators	Annotating Indicators	Testing Indicators
Browsing time	Memo counts	Answered questions counts
Completion rate	Marker counts	Quiz completion rate
Bookmarks counts	Marker text	Accuracy rate at the first attempt
		Accuracy rate on recent attempts

2.2 Proposed Method

In this study, we extracted higher-order evaluation indices from marker counts, which are simple features. Based on the four levels in Zimmerman's (2000) multilevel model, we defined four characteristics that could be related to marker *proficiency level*, as shown in Table 2. The four levels in the multilevel model are classified by the teacher, who checks the students' efforts and determines which level they belong to. Therefore, to evaluate the levels using data, it is necessary to quantify the factors that are usually used by teachers when determining a student's level. The four characteristics in Table 2 are indicators that can be judged as being at the level of emulation if performability is high, self-control if applicability is high, self-regulation if extensibility and continuity are high, and self-regulation if extensibility and continuity are high. Observation is not included in the four characteristics because there is no corresponding data in the existing log data.

Based on the below definitions, the following characteristic variables were provided for each characteristic and quantified using log data.

- *Performability*: mean and standard deviation per minute of markers in classes with AR
- *Applicability*: mean per day and standard deviation of markers in classes other than AR
- *Extensibility*: mean and standard deviation per day of markers at home
- *Continuity*: percentage of daily change for all markers

Table 2. *Multilevel Model and Four Characteristics of the Proposed Proficiency Level*

Multilevel model (Zimmerman, 2000)		Four features proposed in this study	
1. Observation	Vicarious induction of skill from a proficient model	Not applicable	
2. Emulation	Imitative performance of the general pattern of the style of a model's skill with social assistance	Performability	Characteristics that perform when directed in a well-organized environment
3. Self-control	Independent display of the model's skill under structured conditions	Applicability	Characteristics of performing without being directed to do so in a well-organized environment
4. Self-regulation	Adaptive use of skill across changing personal and environmental conditions	Extensibility	Characteristics to be performed in an independent study setting
		Continuity	Sustained execution characteristics

For each of the quantified characteristics, the criteria for classifying each student into two groups, High and Low, were determined and classified. Based on the multilevel model, from the student groups, the *proficiency level* of the marker was classified into five levels as shown in Table 3. "-" denotes Low or High. We have divided self-regulation level in the multilevel model into two levels since it is given two characteristics.

Table 3. *Classification of Proficiency Levels by Four Characteristics*

Level	Performability	Applicability	Extensibility	Continuity
0	Low	Low	-	-
1	High	Low	-	-
2	-	High	Low	-
3	-	High	High	Low
4	-	High	High	High

2.3 Context

This study uses data from a Japanese junior high school that implements LEAF. The data included 13to 14-year-old students (n=112) who belonged to a class in which English AR was conducted using the BookRoll's marker function. In AR, students are asked to reread the English text they read when their WPM was first measured. Students highlight unfamiliar words in yellow and the words they consider important or as keywords in red. This teaches students how to use the markers as a kind of task strategy and helps them deepen their understanding of the text. The experimental class was held for three days (on June 27, 29, and 30). On the first day, AR was conducted, and WPM records were measured. On the second day, students deepened their understanding of the content through group activities and prepared for the following day's presentation. On the third day, the groups gave presentations on the contents of the previous day's group activities. The data were the markers' learning log data from April 1 to June 30, 2022, and the WPM records were measured before (June 27) and after (July 1) the AR. For the markers' log data, we used the markers' log data automatically stored in the LEAF system as described in Section 2.1 WPM was calculated from the number of words in the text and the reading time by measuring the time each student spent reading the English text by themselves. We then asked students to register their WPM on LogPallet and used only this recorded WPM for the analysis.

2.4 Measuring Four Features and Proficiency Level

The four characteristics in this study were generated for formative assessment. Since formative assessment must return a real-time evaluation of the student's current situation and learning behavior, it is necessary to calculate them using the student's latest data. For this reason, they were generated using the AR-recent data from June 24 to June 30. However, the criteria for determining the highs and lows of the four characteristics must be standardized and not variable from period to period or person to person. Therefore, we used data from April 1–June 30 to calculate the criteria. Moreover, we divided each student into two groups, high and low, based on the established criteria for the four AR-recent characteristics and determined the *proficiency level* of the marker for each student according to Table 3.

Since students are taught how to use markers in English reading comprehension exercises in AR, there should be a change in the use of markers in the AR-recent period. Therefore, the *proficiency level* of the AR-recent period should reflect the differences in the use of markers for each student.

2.5 Data Analysis

Three analyses were conducted in this study to investigate each research question:

- Analysis 1. Analysis for RQ1: "For each of the four characteristics of Zimmerman's multilevel model, what value is the criterion for determining whether it is high or low?"

The standard criteria of the four characteristics were determined to divide the markers into two groups, High and Low.

- Analysis 2. Analysis for RQ2: "Is marker *proficiency level* related to learning effectiveness?" There was a difference in WPM measured before and after AR, and the change in value is thought to be related to learning before and after AR. Therefore, we used markers' *proficiency level* as a value to quantify differences in learning by each student, and students were divided into two groups according to their *proficiency level*. We visualized the differences in the distribution of WPM obtained before and after AR between the groups. To confirm whether there is a significant difference in WPM between the two *proficiency levels*, a Mann-Whitney u-test between the two groups of High and Low students was conducted.

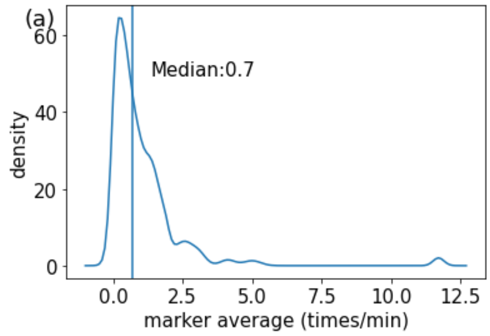
- Analysis 3. Analysis for RQ3: “Compared to the number of markers, is the *proficiency level* more indicative of a student’s learning effectiveness?”

For each *proficiency level*, we visualized the distribution of WPM scores obtained before and after AR. We also visualized the distribution of WPM concerning the number of markers during the entire period for comparison with the distribution by *proficiency level*. We used analysis of variance to confirm whether WPM before and after AR changes with proficiency level or with the number of markers, respectively.

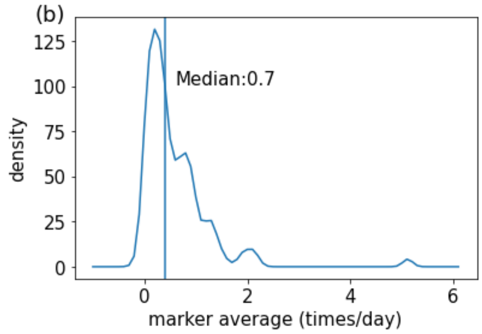
3. Result and Discussion

3.1 Results of RQ1

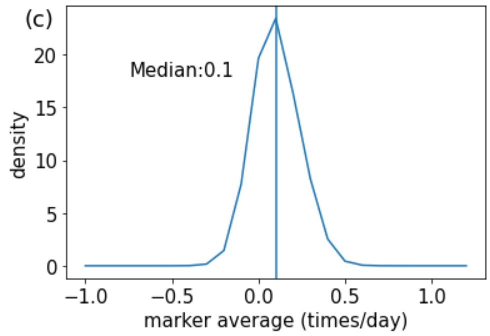
From the log data of students’ markers in the three classes in which AR was conducted, we extracted features related to the four characteristics (Figure 1).



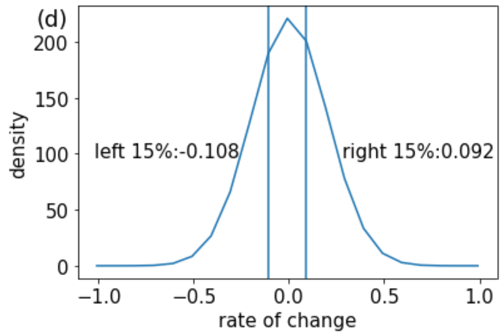
(a) Mean distribution of markers per minute in Active Reading (*Performativity*).



(b) Average distribution of markers per day in classes other than Active Reading (*Applicability*).



(c) Mean distribution of markers per day at home (*Extensibility*).



(d) Distribution of the percentage of daily changes for all markers (*Continuity*).

Figure 1. Distribution of the four characteristics and criteria for marker use.

We defined *performability*, *applicability*, and *extensibility* as the mean distribution of markers by the superposition of the normal distribution created from each student’s mean and standard deviation. *Continuity* was defined as the distribution of the percentage of change by superposition of the normal distribution created from each student's percentage of change and standard deviation 0.1.

Because graphs (a), (b), and (c) in Figure 1 are all skewed to the right and outliers exist, we used the median as the standard for *performability*, *applicability*, and *extensibility*. However, for (d), we classified High and Low based on the 15% interval from the median. Regarding the number of people classified, we determined the standard. Thus, *performativity* is considered high if it is used approximately 0.7 times per minute during AR, *applicability* is considered high if it is used approximately 0.7 times per day at school, *extensibility* is considered high if it is used more than 0.1 times per day at home, and *continuity* is considered high if the rate of change is between -0.108 and 0.092 in the last week.

3.2 Results of RQ2

Figure 2 shows the distribution of WPM calculated before and after AR for each group after dividing the *proficiency level* into two groups, High and Low, from the 24th to 30th of June.

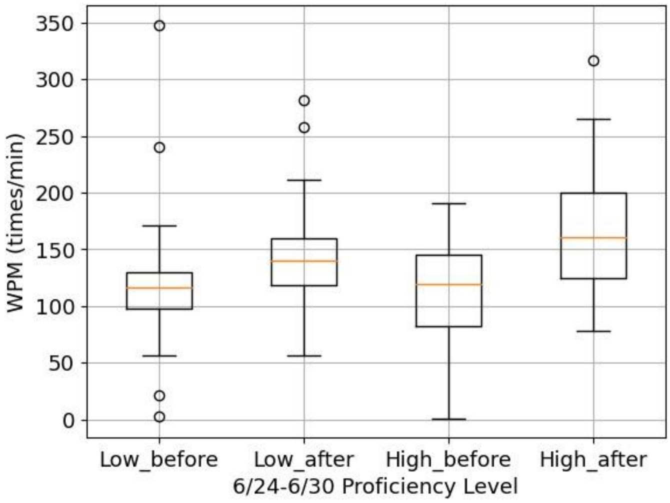


Figure 2. Distribution of words per minute before and after Active Reading for the two groups regarding *proficiency level* from June 24th to 30th.

Next, to confirm that the High group had higher WPM than the Low group, the WPM scores were compared between the two groups. First, the normality of the data for each group was checked using the Shapiro-Wilk test. Since normality was not found, we performed a Mann-Whitney u-test on the two groups with no correspondence between Low_before and High_before and Low_after and High_after to see if they indicated a link between WPM before and after AR, respectively. The results showed that $p=.62$ for WPM before AR and $p=.03$ for WPM after AR. Assuming a significance level of 5%, there was a significant difference in medians between the groups after AR, while there was no significant difference between the groups before AR. Therefore, it appears that students with higher *proficiency levels* improved their WPM more than those with lower *proficiency levels*. From this, we infer that displaying marker proficiency level to students will have the effect of encouraging marker use and further improving WPM.

3.3 Results of RQ3

Two multiple regression analyses were conducted to show that *proficiency level* is more indicative of a student's learning effectiveness than the number of markers. The first was a regression including the WPM values on the number of markers for each student, a dummy variable indicating whether the value was before or after WPM, and their intersection terms as explanatory variables. Figure 3 shows the regression line and the distributions of WPM calculated before and after AR concerning the number of markers. The second was a regression including the WPM values on the *proficiency level* for each student, a dummy variable indicating whether the value was before or after WPM, and their intersection terms as explanatory variables. Figure 4 shows the regression line and distributions of WPM calculated before and after AR for each *proficiency level*.

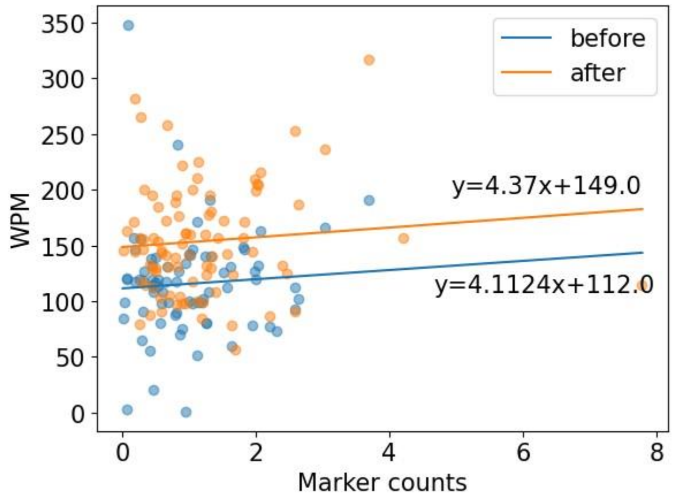


Figure 3. Distribution of words per minute before and after Active Reading concerning the number of markers from April 1 to June 30.

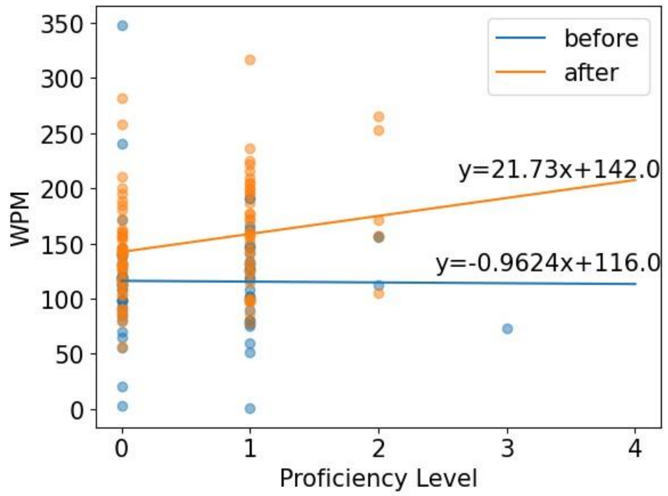


Figure 4. Distribution of words per minute before and after Active Reading concerning *proficiency level* from June 24 to 30.

For the two factors used in the regression in Figure 3 and their interactions, we performed an analysis of variance with these as factors: for WPM, there was no significant difference in the effect of the number of markers (p=.75). Similarly, no significant difference was found in the effect of the interaction between the dummy variable indicating pre- and post-WPM and the marker (p=.82). From this, it appears that there is no association between the number of markers and WPM. Conversely, for the two factors used in the regression in Figure 4 and their interactions, we performed an analysis of variance with these as factors: for WPM, there was no significant difference in the effect of proficiency level (p=.91). However, the effect of the interaction of the marker with the dummy variable indicating that it is the WPM before and after is p=.06, which means that a significant difference is found when the significance level is set at 10%. From this, it appears that proficiency level is not directly related to WPM values, but that proficiency level is related to WPM growth. This suggests that *proficiency level* captures the learning effect of students in more detail than the simple indicator of number of markers. Therefore, assessment by marker proficiency level would reflect the diversity of student efforts and would allow for accurate feedback tailored to the student's situation.

4. Conclusion

In this study, we quantified the degree of *proficiency levels* in marker use from Zimmerman's multilevel model as an example of how to create a higher-order evaluation index. To examine whether the *proficiency level* of the created markers can express learning effects better than the simple feature of the number of markers, we visualized and clarified the difference between the distribution by *proficiency level* and the distribution by the number of markers for WPM. Subsequently, analysis of variance was conducted on WPM obtained before and after AR in terms of proficiency level or number of markers, and no significant difference was shown by number of markers before and after AR, but there was a significant difference in proficiency level. The results showed that the *proficiency level* reflected the learning effect better than the number of markers. In addition, we examined whether differences in *proficiency levels* were indeed related to learning effectiveness and found that students with higher *proficiency levels* showed higher growth in their WPM. From this result, we confirm that the use of *proficiency levels* as an evaluation index can enhance the interpretability of the features and enable evaluation that is more responsive to the student's learning.

However, this study has some limitations. In this study, the *proficiency level* of markers was determined using Zimmerman's multilevel model, which could be applied to a class that teaches the use of AR markers. However, it is not always possible to determine whether the teacher encouraged the use of other functions in the class, other than markers, and it is difficult to calculate the *proficiency level* in the same way. Additionally, the *proficiency levels* of the markers alone are not sufficient to represent the diversity of learning. Therefore, it is necessary to calculate higher-order features for simple features extracted from the logs of other functions in BookRoll and LMS (Moodle), as shown in Table 1. However, as mentioned above, we do not know if the multilevel model can be applied. Therefore, it is necessary to devise a new theory to apply the multilevel model and to consider other models in future research.

As evaluation indicators are expanded in the future, teachers will be able to select from a wider range of evaluation indicators and benefit from a form of evaluation that responds to a variety of learning styles.

Acknowledgements

This study is supported by NEDO JPNP18013, JPNP20006 and JSPS KAKENHI JP22K20246.

References

Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education*, *106*, 83–96. https://doi.org/10.1016/j.compedu.2016.12.001

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304–317. https://doi.org/10.1504/ijtel.2012.051816

Granberg, C., Palm, T., & Palmberg, B. (2021). A case study of a formative assessment practice and the effects on students' self-regulated learning. *Studies in Educational Evaluation*, *68*(3), 100955. https://doi.org/10.1016/j.stueduc.2020.100955

Ogata, H., Majumdar, R., Yang, S. J. H., & Warriem, J. M. (2022). LEAF (Learning & Evidence Analytics Framework): Research and practice in international collaboration. *Information and Technology in Education and Learning*, *19*(41). https://www.let.media.kyoto-u.ac.jp/wp-content/uploads/2021/11/EUHK-1.pdf

Sung, Y. T., Chang, K. E., & Liu, T. C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, *94*, 252–275. https://doi.org/10.1016/j.compedu.2015.11.008

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13–39). Academic press. https://doi.org/10.1016/B978-012109890-2/50031-7