Analysis of Learning Behavior Logs based on a Fill-in Workbook System - Toward Real-time Learning Analytics -

Kousuke Abe, Tetsuo Tanaka, Kazunori Matsumoto

***Abstract*—**It is difficult for teachers to assess overall class performance in lecture situations where there are more than 100 students per class. In such situations, it is difficult to decide when to take appropriate remedial action. In an attempt to assess the overall class attitude and degree of understanding based on the learning behavior shown by all the students in a large class, including inattentive or inactive students, we analyzed data logs generated by a fill-in workbook system developed by the authors. This time series data made it possible to identify whether students were effectively following the class learning schedule. Analysis of the results confirmed that the degree of understanding shown by individual students could also be usefully assessed by employing this approach.

***Keywords*—**Lecture support, Filling in blanks, Real-time learning analytics, Learning behavior log, Learning attitude, Degree of understanding

I. Introduction

TEACHING faculties need to keep an eye on each individual student's performance during class and to flexibly adjust class progress according to each student's attitude, mental condition (i.e., concentrating, being sleepy, unable to follow course content, giving up, etc.) and degree of understanding. For example, if the number of students who have lost concentration increases, faculty members could then let students discuss the topic with each other or work on exercises instead. However, it is difficult for teachers to accurately assess the overall situation in a large class with more than 100 students. As a result, it is difficult to know when to take appropriate remedial action in keeping with the current classroom situation.

The application of learning analytics is attracting increasing attention as one possible approach to helping solve these problems. Learning analytics is a research field that contributes to improving learning efficiency by collecting and analyzing learning behavior logs in e-learning systems. This makes it possible to predict learners' future learning outcomes and identify hidden problems [1][2]. In the past, conventional learning analytics mainly relied on off-line analysis, but more recent developments have included the introduction of a real-time learning analytics environment that can support teachers and students in the classroom – made possible by the spread of wireless LAN in the classroom and the adoption of BYOD (Bring Your Own Device) where students bring their own laptops to classes. Some research results have already been reported in the literature [3][4][5].

The learning behavior logs used in conventional learning analytics generally collect data on parameters such as the page turning times associated with the learning material being used, the amount of underlined content, and the number and type of questions asked. As described in Ogata et al. [6], the learning status of so-called active learners who actively underline, take notes or ask questions can be usefully assessed using these types of data. However, it is difficult to similarly assess the status of many students who are attending classes but who may not be actively learning.

Therefore, the purpose of this study was to examine whether lectures could be improved through timely and appropriate actions based on a better understanding of class attitudes and the level of understanding of all students, including inattentive or inactive students, in order to help them take in lectures more efficiently.

In this study, we used the fill-in workbook system developed by the authors [7] (hereinafter referred to as "this system") in order to collect logs of learning behavior from all students, including those students who are not paying attention. Using this system, all students in class are asked to fill in the blanks in their workbooks, so that data on the learning activities of even the most inattentive students can be collected.

In Section 2, we describe the topic of targeted lectures and their associated problems, and then discuss possible approaches to solving these problems in Section 3. Section 4 outlines the lecture support system used when applying the fill-in textbook approach. In Section 4, we also describe the results obtained by analyzing the learning behavior logs collected using this system in actual lectures.

II. Targeted Lectures and Associated Problems

In this study, we focused on the type of face-to-face lecture classes that are commonly held in private Japanese universities.

Kousuke Abe is with Course of Information and Computer Sciences, Graduate School of Engineering, Kanagawa Institute of Technology, Atsugi, Kanagawa, Japan (e-mail: s1985027@cce.kanagawa-it.ac.jp).

Tetsuo Tanaka is with Department of Information and Computer Sciences, Kanagawa Institute of Technology, Atsugi, Kanagawa, Japan (e-mail: t-tanaka@ic.kanagawa-it.ac.jp).

Kazunori Matsumoto is with Department of Information and Computer Sciences, Kanagawa Institute of Technology, Atsugi, Kanagawa, Japan (e-mail: matumoto@ic.kanagawa-it.ac.jp).

Classes are generally held in classrooms that accommodate from 100 to 200 people and these classrooms often have tiered seating as shown in Fig.1.

Especially in such large classes, students cannot always maintain a high level of concentration over time. In order to keep students focused, many faculty members use the "Monta Method" [8] that leaves important words blank, as shown in Fig.2, and displays them together with the relevant explanations in the students' workbooks. Having to fill in the blank content displayed for each explanation shown in print can help students continue to concentrate on the lesson.

Fig. 1 Lecture in classroom with tiered seating

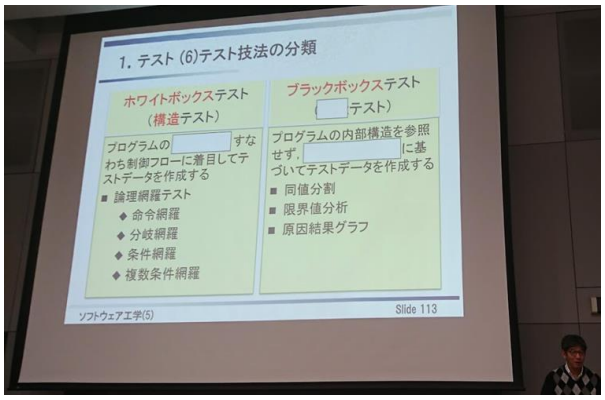

Fig. 2 Lecture using the Monta method

However, in a class comprising more than 100 students, especially in a classroom with tiered seating, it is difficult to ascertain the status of all students - how many students are actually filling in the blanks or working on exercises, for example. Therefore, it is difficult to know when to take appropriate remedial actions such as letting neighbors discuss topics with each other or presenting them with problem-solving exercises when the number of inattentive students increases.

In addition, it is difficult to assess a student's level of understanding in real time during class. Exercises and quizzes are effective in confirming whether students understand the teacher's explanation, but these take time to score and count. As a result, it is not always possible to take actions in a timely manner, such as changing the details of an explanation according to the degree of understanding exhibited by the students. And if some students do not understand what is being said, this may prompt them to whisper to each other and cause other students to lose concentration and focus as well.

It is difficult for students to judge for themselves whether they understand the content of a lecture, and how much they understand or do not understand. In addition, some students may think that they can understand the content, but they may not actually understand it at all, or have an incorrect understanding of it. This does not improve learning efficiency.

III. Approach

A. Extraction of Learning Attitude and Understanding Data from the Learning Behavior Log

In this study, we used a fill-in workbook system developed by the authors to collect learning behavior logs from all students, including inattentive ones. In a class using this system, the teacher makes the students fill in the blanks in a digital textbook generated from a PDF document. As a result, students are forced to interact with the system, and the learning behavior logs of inactive students (students who do not actively take part in learning actions such as underlining and writing notes) can be acquired.

In addition, we analyzed the relationship between the learning behavior log and the learning attitude / understanding of students in the class, and assessed the learning attitude and degree of understanding based on the data obtained from these learning behavior logs. In classroom situations, the results can then be visualized and presented to the teacher in a timely manner. Here, the learning attitude is defined as the level of student activity and involvement, and the degree of understanding is defined as the degree to which the content and meaning of the teacher's explanation is understood by the student.

If teachers can effectively assess the student's learning attitude in real time during class, they can then take actions that encourage them to concentrate whenever the number of students with a poor learning attitude is seen to increase. In addition, if the degree of understanding can be assessed, it is then possible to focus on improved explanations for topics that many students may not understand. Alternatively, teachers can take other actions such as giving hints during an exercise.

B. Learning Behavior Log

The overall structure of the system described above is shown in Fig. 3. This system includes four modes: the authoring mode, the browsing mode, the presentation mode, and the analytics mode. Teachers can create teaching materials by uploading PDF documents in the authoring mode and then adding blanks to important sentences. Students fill in the blanks while referring to these materials in the browsing mode. The operation is recorded in the database as part of a learning behavior log. In the presentation mode, the learning behavior log is presented to the teacher in real time. The analytic mode is used for retrospecting, and the teacher can refer to the learning behavior log in three ways: "all students", "individual students", or "pages".

The items in the learning behavior logs collected using this

system are as follows:

- Fill-in blanks: Content and time of each answer written in the blank, and whether answers are correct or incorrect
- Underlining: Underline position
- Notes: Note position and contents
- Operation logs: Page focus time, number of mouse operations, number of keyboard operations per minute
- Page turning details: Time and destination page numbers
- Question details: Contents of questions asked by students during class

In addition, a teacher's behavior log is also compiled and can be used to compare course progress with the student's learning behavior log. The teacher behavior log includes the following items:

- Page turning details: Time and destination page numbers
- Deleted blanks (displaying correct answer): Deletion times and target blank numbers

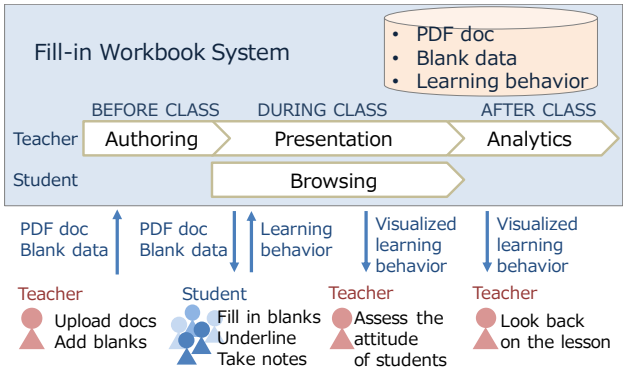

Fig. 3 System overview

IV. Learning Behavior Log Analysis

The analysis was based on data collected during 90 minutes' of actual lectures conducted by one faculty member for about 90 students from the Department of Information and Computer Sciences.

The 4 lectures used for the experiment were taken from a 15-part "Software Engineering" course on software development methods. A faculty member created the fill-in material in advance, using the authoring mode, and then gave the lectures using the presentation mode. At the beginning of the lecture, students were given a 5-minute explanation on how to use the system. They then participated in the lecture using the browsing mode. When using browsing mode, students filled in the blanks on the screen, as shown in Fig. 4. (The blanks are deleted once they have been filled in.) The students then completed a questionnaire survey when the lecture was over.

In analyzing the results of this study, we focused on the degree of synchronization between the teacher's slide transition, along with the student's slide transition, the use of the note and mark functions, how questions were asked, and off-screen time data – as collected by the learning behavior log as an indicator of class attitude.

We also examined the timing associated with the filling in of blanks during the exercise (i.e., before or after the teacher displayed the correct answer), whether an incorrect answer was given before answering correctly, and the reference page being used at the time of the exercise (i.e., whether or not the page being referred to actually related to the exercise). These were all used as an indicator of the student's level of understanding.

The following section describes how the class attitude and level of understanding were extracted from the data collected in the learning behavior log.

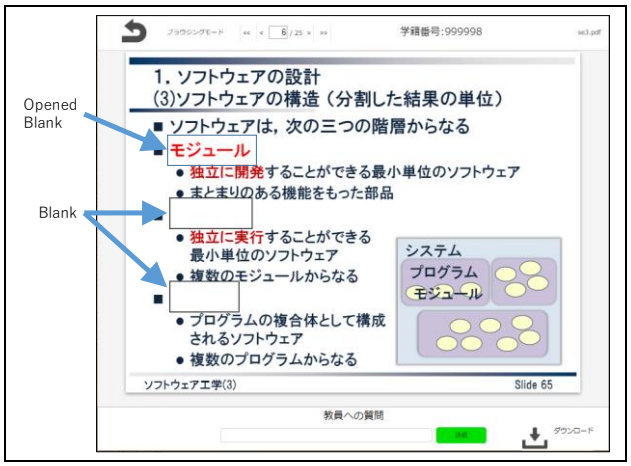

Fig. 4 Screen showing browsing mode in use

A. Extracting Learning Attitude from the Learning Behavior Log

1) Reference Page Synchronization

In order to determine whether or not a specific student was following the lecture being presented to the class, we analyzed the synchronization between the page of the teaching material being referenced by the student and the page being explained by the teacher.

Fig. 5 shows the degree of synchronization recorded for each student in the class. Each row represents the data recorded for an individual student, along with the page that the teacher was explaining and the actual time. The bar graph underneath shows the number of students that were not in synchronization with the teacher at any given time throughout the lecture. The students' data shown in the figure are sorted by synchronization rate, showing how closely they were synchronized with the teacher, on a minute-by-minute basis. In the figure, any student cells that are not synchronized for more than 30 seconds in any given minute are shown filled in.

As shown in Fig. 5 (a), there was a clear division between those students who were lagging behind the teacher, in terms of page-turning timing, and those who were ahead. Also, as shown in (b), there were many students who were not synchronized with the teacher at all. Furthermore, as shown in (c), some of these students remained unsynchronized for a relatively long time.

Possible reasons for the delays observed could include students taking notes on the previous page and being late turning the page, or sleeping or doing other things and not listening to the teacher's explanation. Possible reasons for going

forward could include moving on to the next page without listening to the explanations given on pages that have few blanks (or none at all) or simply proceeding onwards after understanding the contents of the page being explained by the teacher.

Therefore, synchronous data such as that described above could be used to keep teachers informed, in real time, about the number of students who have not understood explanations, or have not caught up, or are moving forward ahead of the lecturer. Teachers could then encourage students to concentrate more at just the right time, or simplify their explanations, or adjust the pace of their lessons accordingly.

However, these data, alone, do not provide enough information to determine whether students are just lagging behind or not listening, and in such cases it may be necessary to consider other actions taken such as students "referring to the previous slide" and perhaps taking note of the "keystrokes made" by each student. These are issues that may need to be considered in the future.

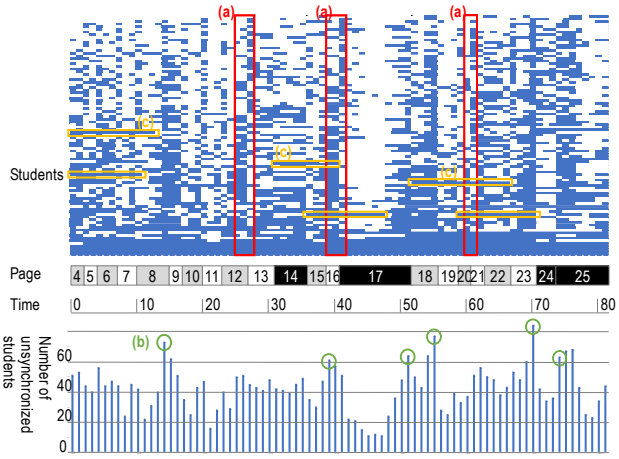

Fig. 5 Reference page synchronization

2) Unfocused Time

Another indicator that can help determine whether a student is paying attention in class is the amount of time during which the student is not focused on the system (i.e., when the system is closed, when the system is open but the student is using other applications, and so on).

Fig. 6 shows the amount of "unfocused time" recorded for each student. The upper part shows individual student data and the lower part shows a bar graph for the total number of students. At the top, students' cells that are off the screen for more than 30 seconds in any given minute are shown filled in. (The horizontal axis shows the time from the start of class.)

As shown in Fig. 6 (a), some students remained off the screen for a relatively long time (3-4 pages of teaching material). These students could have been doing other things, or even sleeping, without listening to the teacher's explanation. Using this sort of data in the classroom, alerts could be issued from the system to encourage concentration.

As shown in Fig. 6 (b), there were also times when the number of off-screen students gradually increased. This could represent pages that have no blanks to fill in, or it might suggest that teachers spent too long explaining the relevant course content. In such cases, around 3 minutes after moving to the next page, the number of off-screen students starts to increase. Based on this data, if teachers are told that there are more students who have not understood the explanations, teachers could take the opportunity to encourage students to concentrate more.

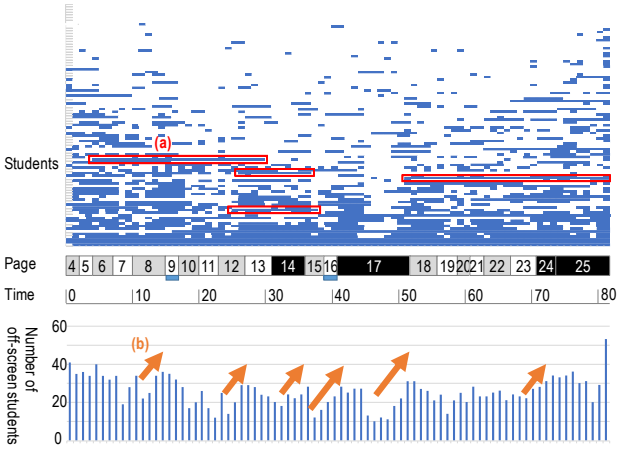

Fig. 6 Unfocused time

3) Summary of Learning Attitude Extraction

As described above, it was possible to determine whether students were listening to the teacher's explanation closely and were following the lecture, based on the synchronization between the page being explained by the teacher and the page being referenced by the student, and by the amount of off-screen time recorded. In addition, we confirmed that when the explanation of each page by the teacher was taking too long, students did not listen to the teacher's explanation but referred instead to other pages or left the system.

Therefore, by using this sort of data in a classroom situation, it would be possible to notify the teacher accordingly or to alert each student directly from the system.

In addition, as a means of after-class reflection, teachers could use this data to identify places where students are not following the lecture properly and where more students begin to do other things. This information could then be used to improve future classes. It could also be used to allow students to reflect on their own attitude in class. How best to implement these support functions will be a topic for further study in the future.

B. Extracting Degree of Understanding from the Learning Behavior Log

This section describes the analysis carried out on the answer times for the exercise page in order to determine the student's level of understanding.

In class, all students had to complete 3 exercises, with the teacher only displaying the correct answer after the students had finished each exercise by themselves. In addition, a multi-choice problem was described involving the selection of

the correct formula from a range of 4 possible choices.

Students who understood the lecture material could answer correctly, without having to refer to the previous slide, before the teacher displayed the correct answer. Students who had some understanding of the lecture material could answer by referring to the previous slide, and while students who did not understand the lecture material at all could not answer the question, they could still copy in the right answer once it was displayed by the teacher.

In this system, when an answer was entered in the blank space, the blank was deleted only when the answer was correct, and not when it was invalid. Therefore, it was immediately known whether the entered answer was correct or incorrect. In the multiple-choice question described, even if the students didn't understand the topic they could still answer correctly within 4 attempts by simply entering each possible answer in order. Students who understood the topic could answer correctly with a single entry, while students who possessed only partial understanding usually needed to try several times before answering correctly.

In order to analyze these results, we focused on the times taken to fill in the blanks correctly during the exercise (both before and after the teacher displayed the correct answer), whether there was an incorrect answer given before answering correctly, and the reference page in use at the time of the exercise (i.e., whether or not a page actually relating to the exercise was being referred to) and used these parameters as an indicator of each student's overall level of understanding.

1) Exercise answer times

In the experimental class, three test exercises were conducted. The first was a descriptive exercise with 3 parts, the second and third exercises involved answering multiple-choice questions

Fig.7 shows whether each student answered before the teacher displayed the correct answer. The horizontal axis lists each of the 3 exercises while each row shown above represents the data for one student. The individual cells are filled in wherever a student filled in the blank after the teacher had displayed the correct answer.

Students who only filled in the blanks after the teacher had displayed the correct answers had either not understood enough of the course material to solve the problem by themselves, or were not motivated enough to answer on their own. In such cases, if it turns out that there are many students in class who do not understand the course content properly, their teachers need to explain the questions in more detail. If there are many students who are not motivated, it may be necessary to change the format to include questions that attract students' interest more. This will help to improve classes.

In order to distinguish between a lack of understanding and a lack of motivation, it may also be necessary to analyze whether any additional keystrokes are made before displaying the final correct answer.

2) Existence of incorrect exercise answers

Fig. 8 shows whether each student entered an incorrect answer during the exercise. The horizontal axis lists each of the 3 exercises, while each row of data shown above represents the data for one student. The individual cells are filled in wherever the student entered an incorrect answer.

The presence of an incorrect answer suggests either that students lacked understanding, or that they simply guessed (incorrectly).

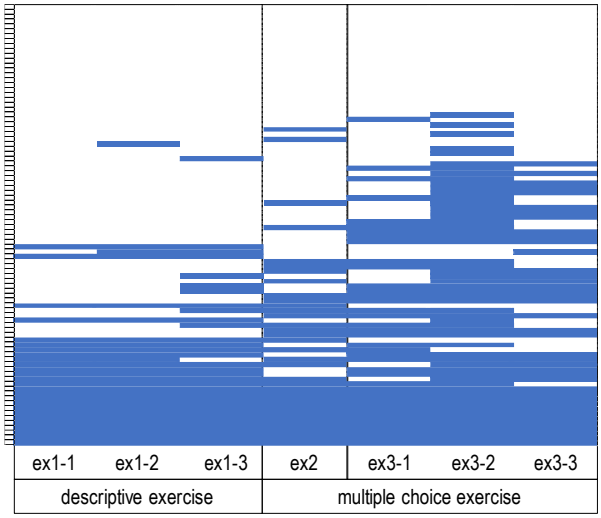

Fig. 7 Exercise answer times

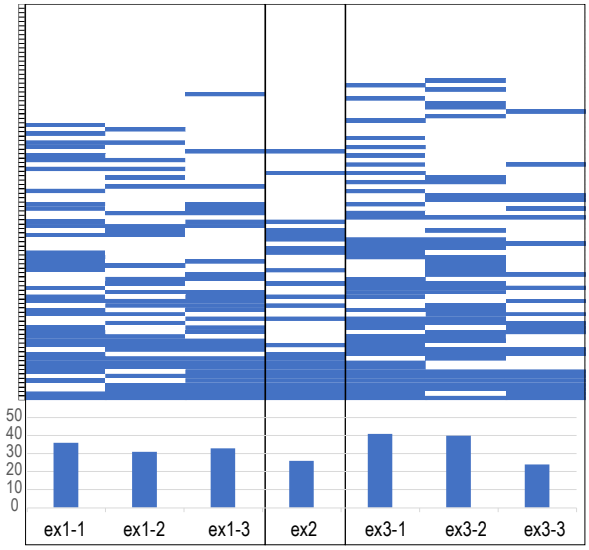

Fig. 8 Number of incorrect exercise answers

3) Reference pages used for the exercise

Fig. 9 shows whether other pages were being referenced when completing the three exercises. Each row represents the data for one student, and filled-in cells show when other pages were being referred to.

Although there were many references to other pages in Exercise 1, Exercises 2 and 3 could be answered correctly by just entering all possible choices in order, without the need to refer to any other pages (showing that a multi-choice exercise is not really suitable for such experiments).

However, referencing other pages during Exercise 1, as shown above, could be interpreted as signifying a lack of understanding and that the student felt that the exercise could

not be completed without checking the previous page. It could also signify willingness on the part of the student to solve the problem without waiting for the teacher to display the correct answer.

Those who entered the correct answer without referring to other pages either answered correctly by themselves or simply copied the teacher's correct answer. In order to distinguish between the two possibilities, it was necessary to combine this data with the answer times described in 4.1.1. If this could be done in a classroom situation, teachers could give a useful hint to students giving up without referring to other pages.

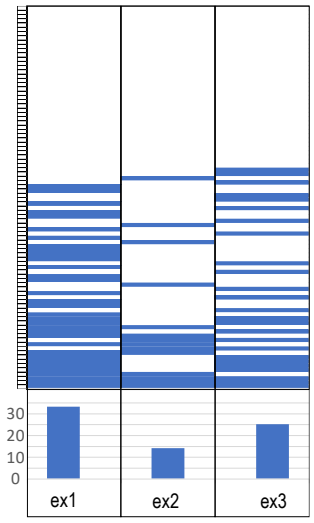

Fig. 9 Reference pages used for each exercise

V. Conclusion

In order to obtain a useful indication of the overall class attitude and degree of understanding of all the students in a class, even the inattentive or inactive ones, we analyzed their learning behavior logs – using a fill-in workbook system developed by the authors.

Specifically, we tried to discriminate between different learning attitudes based on the use of reference page syncronization and unfocused time data. Also, we tried to determine the degree of student understanding based on answer times, the existence of incorrect answers, and the reference pages used for a series of exercises.

The results showed that it is possible to use this type of data to identify whether or not students are paying attention to and understanding class content at any given time. It was also confirmed that the degree of understanding exhibited by individual students could be usefully assessed using this approach.

In future, we will examine the target data and the analysis results further, and improve the method used for discriminating between different learning attitudes and levels of understanding. In addition, we will try to develop an effective presentation method for feeding back the analysis results to teachers and students in real time during class.

References

[1] Olga Viberg, Mathias Hatakka, Olof Balter, Anna Mvroudi, The current landscape of learning analytics in higher education, Computers in Human Behavior 89, pp.98-110, 2018.

[2] Diana M. Naranjo, Jose R. Prieto, German Molto, Amanda Calatrava, A Visual Dashboard to Track Learning Analytics for Educational Cloud Computing,

[3] Atsushi Shimada and Shin'ichi Konomi, Hiroaki Ogata, Real-time learning analytics system for improvement of on-site lectures, Interactive Technology and Smart Education, Vol.15, No. 4, pp.314-331, 2018

[4] Atsushi Shimada and Shin'ichi Konomi, A Lecture Supporting System Based on Real-time Learning Analytics, Proc. of CELDA2017, pp.197-204, 2017.

[5] Xinyu Fu, Atsushi Shimada, Hiroaki Ogata, Yuta Taniguchi, Daiki Suehiro, Real-time learning analytics for C programming language courses, Proc. of the Seventh International Learning Analytics & Knowledge Conference, pp.280-288, 2017.

[6] Hiroaki Ogata, Yuta Taniguchi, Daiki Suehiro, Atsushi Shimada, Misato Oi, Fumiya Okubo, Masanori Yamada, Kentaro Kojima, M2B System : A Digital Learning Platform for Traditional Classrooms in University, Proc. Of The 7th International Conference on Learning Analytics & Knowledge Understanding, pp.155-162 (2017).

[7] Kousuke Abe, Tetsuo Tanaka, Kazunori Matsumoto, Lecture Support System using Digital Textbook for Filling in Blanks to Visualize Student Learning Behavior, International Journal of Education and Learning Systems, Vol.3, pp.138-144 (2018)

[8] Garr Reynolds, The "Monta Method" – Presentaion Zen, https://www.presentationzen.com/presentationzen/2005/10/the_monta_metho.html (2005)