Rodrigo, M. M. T. et al. (Eds.) (2021). Proceedings of the 29th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Mining Students' Engagement Pattern in Summer Vacation Assignment

Hiroyuki KUROMIYA[a*], Rwitajit MAJUMDAR[b] & Hiroaki OGATA[b]
[a]*Graduate School of Informatics, Kyoto University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*khiroyuki1993@gmail.com

Abstract: Learning Analytics (LA) is an emergent field which aims at a better understanding of students and providing intelligence to learners, teachers, and administrators using learning log data. Although the use of technology in class is increasing in the K-12 sector as well as territory education, cases of effective implementation of LA in secondary schools were rarely reported, especially in Japan. In this paper, we offer an example where LA is implemented at a junior-high Math class in Japan. We introduce our LA platform, LEAF - LMS and e-book integrated learning analytics dashboard - and its usage during summer vacation period in the target class. We analyzed 121 students' question answering logs and their exam performance after the vacation by K-means clustering method. As a result, we found that students' progress patterns were able to be categorized as four types: early engagement, late engagement, high engagement, and low engagement and the early and high engagement group got significantly higher scores than the low engagement group. It implies the importance of the engagement at the beginning of the vacation. Moreover, by comparing the previous studies in MOOCs, we concluded that self-regulation skills are an important factor for student success in a long vacation period, too. Finally, we introduce a monitoring tool which aims to detect and send messages to at-risk students at an early stage in the next summer vacation period. Our case will become the first model case of how to implement LA in secondary school in Japan.

Keywords: Learning Analytics, Secondary Education, Long Vacation Period, Pattern Mining

1. Introduction

Learning Analytics (LA) is an emergent field which aims at a better understanding of students and providing intelligence to learners, teachers, and administrators using learning log data (Law & Liang, 2020). Higher Education (HE) has been using LA for improving the services and students' retention rate (Bienkowski et al., 2012). However, adopting LA in school is not an easy task. It is estimated that it will take two or three years to adopt LA within primary and secondary education (Freeman et al., 2017). There are many barriers for the adoption of LA in K-12 context: the privacy issues are more sensitive (Gunawardena, 2017), resources for supporting analytics implementation more constrained, and expertise in data analytics is very limited in school context (Kovanović et al., 2020). As a result, it is reported that the number of studies conducted in schools is much fewer than that of in higher institutes (Li et al., 2015); 82.9 % of studies focus on higher education while 17.1 % of secondary school. As a matter of course, cases from Japanese junior-high schools that report LA implementations in school were very limited in the current state.

In this paper, we offer an example where LA is implemented at a junior-high Math class in Japan. In particular, we focused on a summer vacation period in the target school. We addressed the following problems in this paper: 1) What kind of learning patterns were extracted from students' log data during a long vacation period? and 2) How were the extracted patterns related to their performance? Our paper is structured as follows. Related works introduces research on learning analytics implementations at school level education and the prior studies about students' engagement patterns. In the Methodology section, we describe our implementations of LA in the target school and the study settings where we conducted in a junior-high Math class in Japan. Results section shows typical students' learning patterns in summer vacation period extracted by an unsupervised clustering analysis and the relationship with their academic performance after the vacation. Discussion section summarizes our findings and proposes a development of a system which enables timely-intervention in the next summer vacation period for us.

2. Literature Review

2.1 Learning Analytics Implementation at School

As we have mentioned above, there are not so many cases from secondary education context in the learning analytics field. Although the use of technology in class is increasing in the K-12 sector (Horn & Staker, 2011), just using technology in classrooms is not enough for what we call LA. According to Clow (2012), LA is defined as a cycle which consists of four phases - learners, data, metrics, and intervention. Closing the loop is a crucial factor for successful implementation of LA (Corrin et al., 2020). In that sense, the cases about the implementation of LA in schools are very limited in the current state.

Here, we introduce some examples and trends about the adoption of LA in school level education. In Spanish, a research project called PILARES (Smart Learning Analytics Platform to enhance Performance in Secondary Education) was developed for blended learning in secondary school (Sancho et al., 2015) financed by the Spanish government and the collaboration of the Catalan Ministry of Education. It includes a large Moodle based LMS called AGORA, that is used by more than 1,500 schools in Catalonia, and it aims at building a LA platform to allow better insight of the learning process through the LMS. In Uruguay, a countrywide LA tool was introduced for secondary education (Macarini et al., 2019). Although they shared several challenges and constraints they faced during its conception and development, they pointed out the feasibility of finding meaningful patterns using the data obtained from the database, and proposed a prototype for tracking the students' scholar trajectory. Although the substantial growth of the learning analytics field itself provided more possibilities to use LA in primary and secondary education (Ochoa et al., 2017), the actual implementations are very limited in K-12 context.

2.2 Student Engagement Patterns

Students' engagement patterns have been investigated mainly using psychological questionnaires to students. Schnitzler et al. (2020) analyzed 397 high school students' profiles by latent profile analysis (LPA) based on three indicators - participation, cognitive engagement, and emotional engagement. Although the first indicator was assessed by the number of hand-raisings in the classroom, the others were measured with survey items. Finally, they discovered five engagement patterns - disengaged, compliant, silent, engaged, and busy - and the significant differences among the learning patterns. The similar approach was taken in the U.S., targeting 1125 middle school students in a science course to identify engagement profiles and their relationship with science achievement (Bae & DeBusk-Lane, 2019) . By applying LPA, they discovered five engagement types - Moderately engaged/disengaged, Behaviorally engaged/disengaged, and Disengaged - from the survey of students.

Some studies used learning log data to categorize students' engagement patterns. Ebook reading logs were used to categorize students' study patterns at a university in Japan (Akçapinar et al., 2020). They constructed study sequences based on the timestamp they opened the material from the click-stream data and adopted hierarchical cluster analysis to the dataset. As a result, they found three different study patterns from the dataset. MOOCs' interaction logs with lectures and assignments were also used to identify learners' study patterns (Boroujeni & Dillenbourg, 2018). They extracted the action sequences from learners' log data and transformed them into probability distribution matrices for distance computing. By the analysis, they classified students into mainly two types - fixed approach and changing approach - based on their strategy of learning. In the analysis, both hypothesis-driven and data-driven approaches were taken in this study.

2.3 Research Questions

As we have seen before, the LA implementations in secondary schools were very limited; at the same time, student engagement study during a long vacation period in schools were very few. In this paper, we investigate following two questions regarding the LA utilization at school during a long vacation period. In Japanese secondary school context, long vacation period is not only a break from study, but also home study period for students. Usually, students are assumed to engage in their home working assignments given by the teacher before the vacation period.

RQ1: What kind of learning patterns were extracted from students' log data during a long vacation period?

Here, we investigate the study patterns of students in a junior-high school during a long vacation period using the LA platform implemented at the school. Standard clustering techniques based on the timestamp of the engagement from students' interaction logs with an ebook reader will be used to answer this question.

RQ2: How were the extracted patterns related to their academic performance?

To answer this question, we explore the relationship between students' study patterns extracted in the previous question and their academic performance after the vacation. Statistical testing of average exam score by each cluster will be conducted to check if there are any significant differences among clusters.

In the next section, we described a case of a junior-high school in Japan during a summer vacation period. We introduce a specific learning analytics platform called LEAF and offer a use case of it during a summer vacation period in the target school. Then, we analyzed the data retrieved from the platform and discussed the possibility of it.

3. LA Implementations for School in Our Context

3.1 LEAF Platform

Since April 2019, we have been offering an LMS-integrated learning analytics platform called LEAF (Learning Evidence Analytics Framework) to a junior-high school in Japan (see Figure 1). LEAF has three online learning tools - Moodle, BookRoll, and LAViEW (Flanagan & Ogata, 2018). In LEAF, school teachers upload learning materials to an ebook reader BookRoll, and students view the materials as they need. Teachers are also able to post quizzes or recommendations (external link) to their materials and students can comment, highlight text, and write handwritten memos on that material. BookRoll is accessed from Moodle LMS by LTI (Learning Tools Interoperability) authentication method, so users can log in to BookRoll without creating an account for it. It gives a benefit to researchers as well because we can retrieve students' information by their moodle ids. The last component, LAViEW is a dashboard that visualizes the learner-content interactions in BookRoll. Teachers can see students' highlights, memos, and time spent on each page.

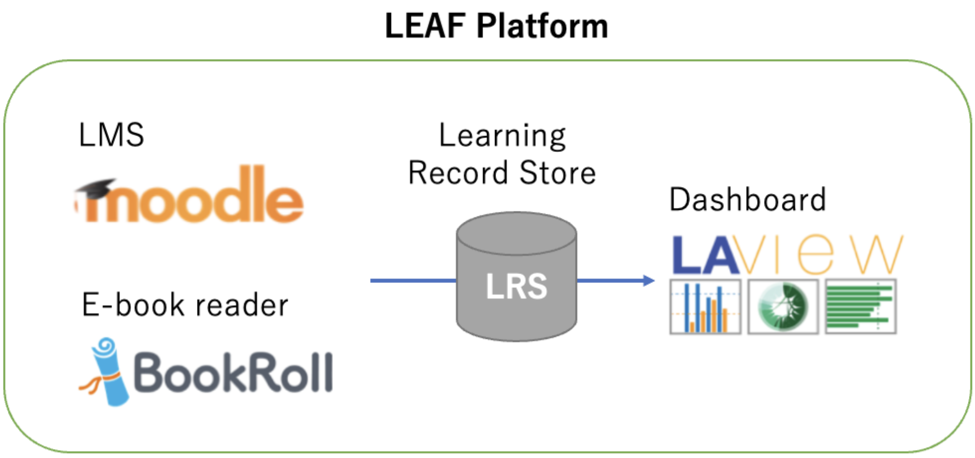

Figure 1. Three Learning Tools in the LEAF Platform in the Target School.

3.2 *Use Case Scenario: Monitoring Summer Vacation Assignment with LEAF*

In this paper, we focus on a specific use case scenario of the LEAF platform in summer vacation period in Math class. We targeted three classes containing 121 students in a junior-high third grade Math course. Before the summer vacation, a teacher uploaded an assignment containing forty-nine Math questions to BookRoll (see Figure 2). The assignment consists of one question per page, and one simple questionnaire is implemented per page. The questionnaire has options which represent three different understanding levels - perfect, understood, or not well understood - as you can see from Figure 2 on the right hand side. In the summer vacation period, students are required to answer the questionnaire each and every time students finish solving a problem as well as submit a paper which contains the answers and working formulas to all the questions. The duration of the summer vacation was from 22 July to 21 August 2019. To measure students' performance, the examination was conducted on 23 August after the summer vacation. We used this score to investigate the relationship between students' behavior and performance.

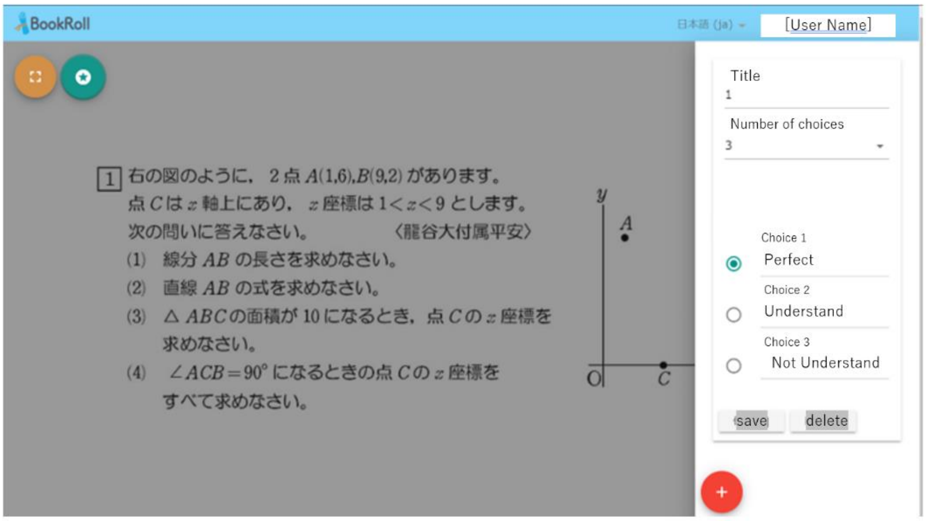

Figure 2. Summer Vacation Assignment Broadcasted via Bookroll.

4. Analysis Procedure

4.1 *Preprocessing of the E-book Logs*

To measure students' engagement during the summer vacation period, we used the students' answer logs to the questionnaire on the assignment. By analyzing students' answer logs, we can get which students solved which question on what day. For the log analysis, we targeted ebook logs from 17 July to 22 August. Overall, we extracted 6,250 answers from the students. we excluded 1,172 answers from page no.1 where the content was just a description of the assignment. Then, we excluded 132 answers after the exam. As a result, we analyzed 4,936 answers by students. Throughout this process, sixteen students were excluded because they had no answer data. Moreover, three students were excluded because they had no score data. In total, we got 102 students.

4.2 *Categorization of the Learning Patterns*

As the summer vacation period was over thirty-seven days, we separated the answers into two periods: logs before 4 August were from 'first-half' and after were from 'second-half.' Duplicated answers to the same question in the same period were excluded in this process. Finally, standard K-means was conducted based on the number of interactions during the first-half and the second-half period for each student. The number of clusters was determined by the Elbow plot based on several cluster indicators (AIC, BIC, and WSS).

4.3 Relationship with Performance

After the categorization step, we compared the average exam score after the vacation period among each cluster. ANOVA and Post-Hoc testing were adopted using statistical testing software, JASP (Love et al., 2019).

5. Results

5.1 Clustering Students' Behavior

Figure 3 shows the Elbow plot which represents the transition of three cluster information over nine cluster numbers (from two to ten). Three cluster information is AIC (Akaike Information Criteria), BIC (Bayesian Information Criteria), and WSS (Within Sum of Squares). Based on the value of BIC, we decided the optimal cluster number as four because it was the lowest BIC value in the plot.

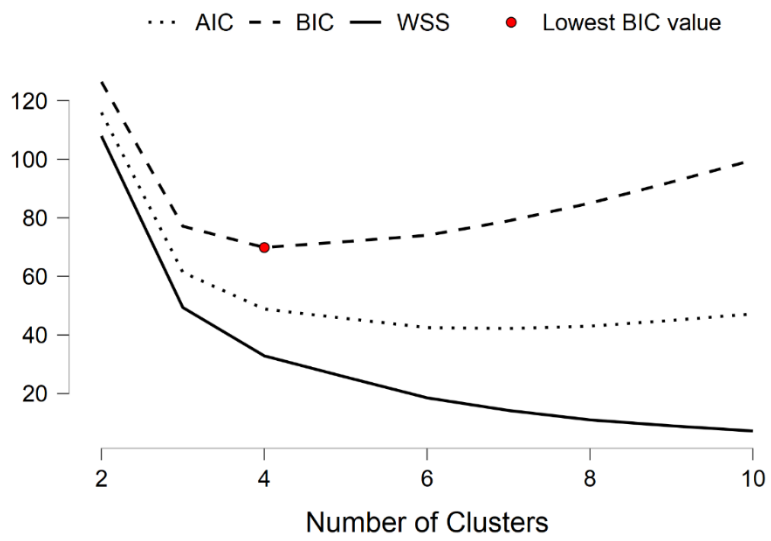

Figure 3. Elbow Plot for Determining Optimal Cluster Number.

Figure 4 shows clustering results when the number of clusters was four. We labeled the cluster based on the location of the scatter plot. Nineteen students were categorized as the early engagement pattern, thirty-two students as high engagement pattern, fourteen students as late engagement pattern, and thirty-seven students as low engagement pattern.

Figure 5 shows actual students' progress patterns for each cluster. Horizontal axis represents the day of the vacation and the vertical axis represents the page of the question they answered. As you can see, students in the early engagement group finished the assignments within the first-half period. On the other hand, students in the late engagement group made little progress by the half of the vacation. Students in the high engagement group can be categorized as two sub groups: some students finished the assignment once by the half of the vacation and solved the questions again at the end of the vacation and others continuously were working throughout the vacation. On the other hand, students in the low engagement group were not able to finish the assignment.

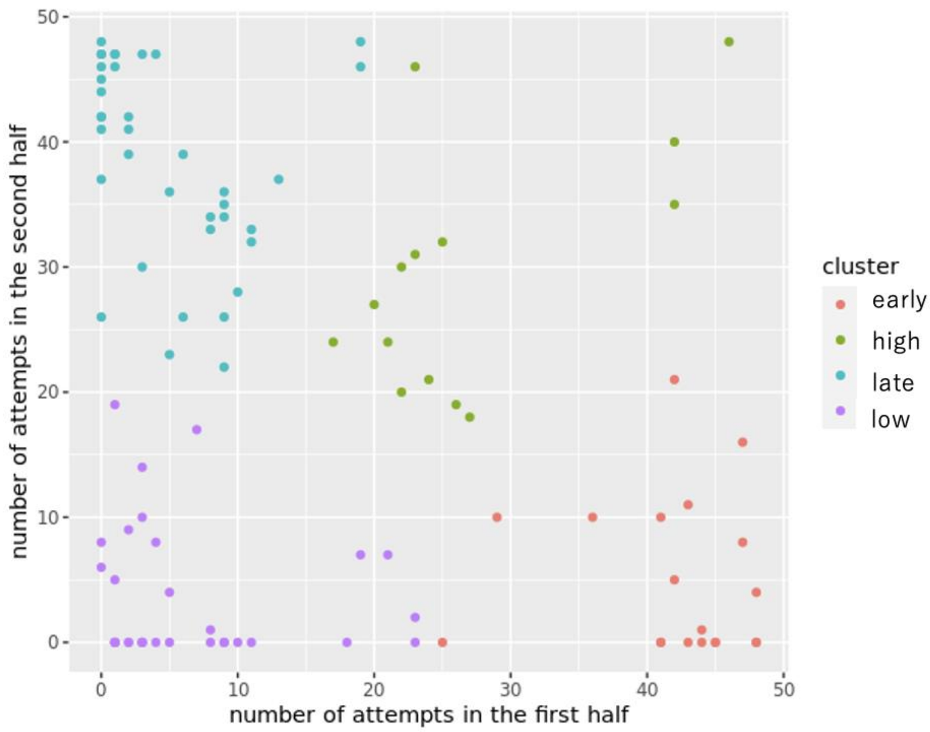

Figure 4. Clustering Results based on the Optimal Cluster Number (K=4).

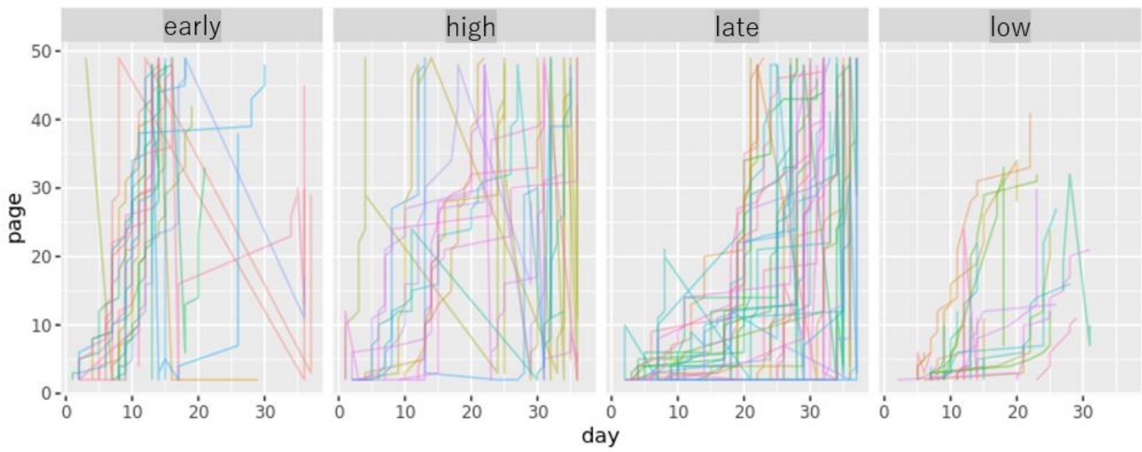

Figure 5. Students' Progress Patterns for each Cluster.

5.2 Relationship with Performance

Finally, we investigated the relationship between their solving patterns and performance. Figure 6 shows the descriptive plot of students' exam scores among four clusters. The error bar stands for standard error of each cluster. As you can see, students in the early engagement group got the highest score (M: 80.8, SD: 14.5), and students in the high engagement group got the second highest score (M: 80.7, SD: 15.0). Students in the late engagement group were the third rank (M: 73.4, SD: 19.9) and students in the low engagement got the lowest score among all (M: 61.9, SD: 21.0). Finally, we conducted ANOVA to test the difference in exam score by clusters. Before adopting ANOVA, we checked the homogeneity of the data and the result was not significant ($p = .18$). The result of ANOVA was significant ($p = .001$), so we conducted a Post-hoc comparison between clusters (see Table 1). As a result, we got significant differences between cluster 1 and 4; and cluster 2 and 4. The p-values were adjusted for multiple comparisons.

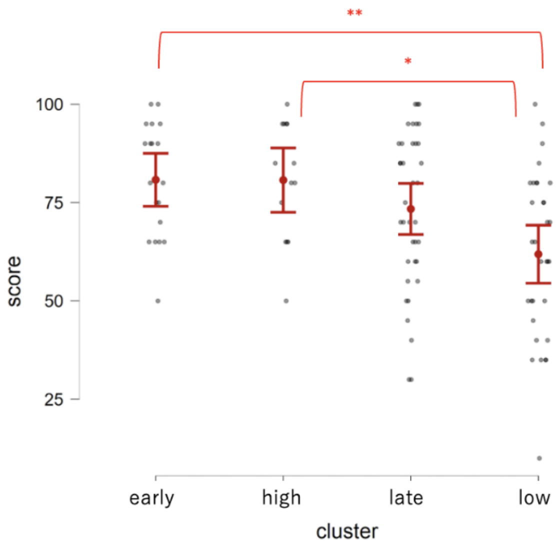

Figure 6. Average Exam Score after the Vacation for each Cluster.

Table 1. *Post-Hoc Comparison Results between Every Cluster Combination*

		Mean Difference	SE	T	P
Early	High	0.075	6.618	0.011	1.000
	Late	7.411	5.303	1.398	0.504
	Low	18.914	5.442	3.476	0.004**
High	Late	7.336	5.896	1.244	0.600
	Low	18.839	6.021	3.129	0.012*
Late	Low	11.503	4.536	2.536	0.061

*Note. p-value adjusted for comparing a family of four (*p<.05, **p<.01)*

6. Discussions and Conclusions

6.1 Summary of Our Findings

By the K-means clustering of the students' answer logs, we found that students' progress patterns were able to be categorized as four types:

1. Early engagement: Students who engaged at the beginning of the vacation
2. High engagement: Students who were continuously working throughout the vacation
3. Late engagement: Students who did little until the end of the vacation
4. Low engagement: Students who didn't finish the assignment

Before our analysis, the teachers in the target school had an assumption that students can be categorized as four types: 1) constant group, 2) early engagement group, 3) late engagement group, and 4) early + review group. Constant engagement group refers to students who continuously work the assignment (e.g. one question per one day) and early + review group refers to students who finished early and review the questions at the end of vacation, too. They didn't anticipate the low engagement pattern because the submission of the assignment was compulsory for students. Compared to teachers' prior expectation, our analysis revealed that 1) there were some students who didn't complete the assignment, and 2) there were not so many students who were continuously working on the assignment: most students focused on the assignment on specific days. Plus, by comparing their performance on the exam after the vacation, we found significant differences between early engagement group and low engagement group, and high engagement group and low engagement group. It implies the importance of the engagement at the beginning of the vacation.

6.2 Theoretical Interpretations

It is often said that online engagement on an assignment affects students' academic performance. For instance, a study investigated predictors of students' weekly achievement and found that time spent on homework and labs were more strongly related to their performance than time spent on discussion boards or books (DeBoer & Breslow, 2014). Another research insisted that the features regarding assignment features such as average submission lead time and total quiz submission plays important roles in the dropout prediction model in MOOCs (Gardner & Brooks, 2018). However, unlike our study, the timing of the engagement was not often considered in the performance prediction before. It is not well known whether the time to finish the assignment is related to the academic performance of students.

In this paper, we make a hypothesis that early engagement reflects high self-regulation skills of students. Until now, many studies indicated that students' self-regulation skills contributed to high performance in MOOCs. For example, an online survey was conducted to identify the main cause of the high drop-out rate in MOOCs (Nawrot & Doucet, 2014). They revealed that the bad time management was the biggest reason among students. (Im & Kang, 2019) conducted path analysis between self-regulated learning and learning achievement from a large dataset from Korean Cyber University. They revealed that self-regulated learning was positively correlated to the participation and the participation was positively correlated to the learning achievement. There, the self-direction skill indirectly contributes to the learning performance. In our context, students who finished the assignments early can set a goal and deadline and make a plan to get there in time. It would be proof of the high self-regulation skills of the student, which lead to a high score of examination after the vacation.

6.3 System Development for Closing the Loop

In this study, we extracted the typical students' engagement patterns during the summer vacation period and explored the relationship between engagement patterns and their performance. However, just analyzing data is not enough for effective learning analytics implementation. As we mentioned in the introduction, effective learning analytics study should make a loop which includes some interventions to students based on the analysis results (Clow, 2012). Therefore, we plan an intervention study in the next summer vacation period at the target school. In particular, we plan to send messages to low engagement students by the first half of the vacation and encourage students to finish the assignment early. To do this, we prepared a real time monitoring tool for students' engagement in the next summer vacation period (see Figure 7). Through this dashboard, teachers can see the progress of the summer vacation assignment and send messages for each student. To check the effectiveness of the message, first we randomly select the students as an experimental group and send messages only to the half of the students at risk. Then, we will monitor if there are some behavioral changes and performance differences between the experimental and control group. It will contribute to closing loops for effective LA implementations at the target school.

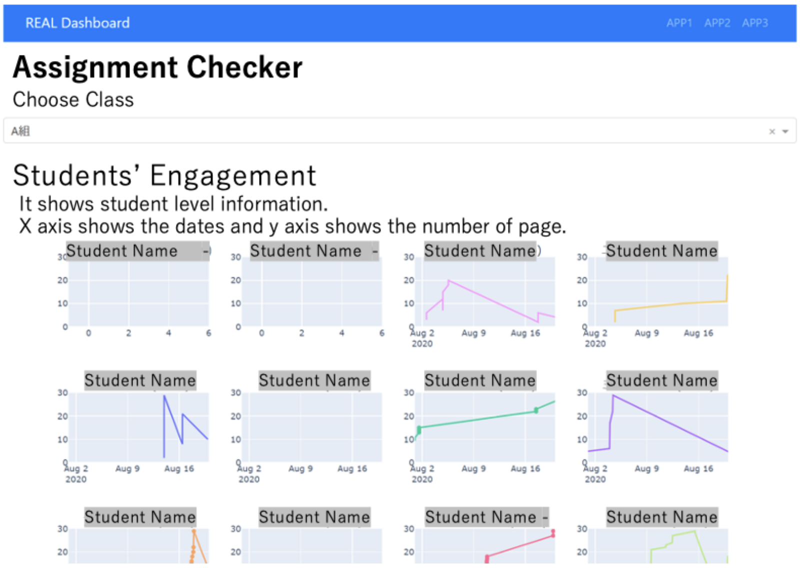

Figure 7. System Mock-up for Timely Intervention for the next Summer Vacation Period.

6.4 Limitations and Conclusion

Broadly speaking, there are two limitations in this study. First is the validity of our categorization. In this study, we used the number of answers in the first and second half of the vacation as the clustering features. These features give us advantages when we visualize the results of the clustering. However, in that context, we were not able to consider the rich time-series information for pattern clustering because they are aggregated features of daily engagement of students. Advanced clustering methods such as time-series clustering (Hung et al., 2015) may enable us to treat the rich time-series information of students’ answering data. Second is the lack of causality. In this study, we just took correlation between the extracted patterns and their performance, so we couldn’t say exactly if engagement patterns affected students’ learning performance. We planned an experimental plan above in order to check the causal relationship between them.

In this paper, we offered an example of a learning analytics platform implemented at actual junior-high Math class in Japan. In particular, we introduce a case in a summer vacation assignment. Summer vacation period is a temporal remote learning period in face-to-face classrooms, so the use of the technology is easier than normal face-to-face learning period. The analysis of log data proposed that 1) students’ progress patterns were categorized as four types based on the number of the engagement in the first and second half of the vacation, and 2) engagement in the beginning of the vacation was important for their academic performance after the vacation. Finally, we showed a real-time monitoring tool for closing the loop of learning analytics implementation at the target school. It will be used in the next summer vacation period in the target school. Our case will become the first model case of how to implement LA in secondary school in Japan and we hope the use of LA will increase in the secondary education sector as well as higher education institutes.

Acknowledgements

This study is supported by JST JPM-JAX20AA, JSPS 21J14514, SPIRITS 2020 of Kyoto University, JSPS 20K20131, JSPS 20H01722, JSPS 16H06304, NEDO JPNP18013, and NEDO JPNP20006. We thank all the teachers in the target school who kindly cooperated with this study.

References

Akçapinar, G., Hasnine, N. M., Majumdar, R., Chen, A. M.-R., Flaganan, B., & Ogata, H. (2020). Exploring Temporal Study Patterns in eBook-based Learning. *Proceedings of the 28th International Conference on Computers in Education*, 342–247.

Bae, C. L., & DeBusk-Lane, M. (2019). Middle school engagement profiles: Implications for motivation and achievement in science. *Learning and Individual Differences*, 74, 101753.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *Office of Educational Technology, US Department of Education.* http://files.eric.ed.gov/fulltext/ED611199.pdf

Boroujeni, M. S., & Dillenbourg, P. (2018). Discovery and temporal analysis of latent study patterns in MOOC interaction sequences. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 206–215.

Clow, D. (2012). The learning analytics cycle: closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, 5.

Corrin, L., Scheffel, M., & Gašević, D. (2020). Learning Analytics: Pathways to Impact. *Australasian Journal of Educational Technology*, 36(6), 1–6.

DeBoer, J., & Breslow, L. (2014). Tracking progress: predictors of students' weekly achievement during a circuits and electronics MOOC. *Proceedings of the First ACM Conference on Learning @ Scale Conference*, 169–170.

Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning: An International Journal*, 10(4), 469–484.

Freeman, A., Becker, S. A., & Cummins, M. (2017). NMC/CoSN horizon report: 2017 K. *The New Media Consortium.* https://www.learntechlib.org/p/182003/

Gardner, J., & Brooks, C. (2018). Dropout Model Evaluation in MOOCs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1). https://ojs.aaai.org/index.php/AAAI/article/view/11392

Gunawardena, A. (2017). Brief survey of analytics in K12 and higher education. *International Journal on Innovations in Online Education*, 1(1). http://onlineinnovationsjournal.com/streams/analytics/07ea372322ccb0f5.html

Horn, M. B., & Staker, H. (2011). The rise of K-12 blended learning. *Innosight Institute*, 5. https://aurora-institute.org/wp-content/uploads/The-Rise-of-K-12-Blended-Learning.pdf

Hung, J.-L., Wang, M. C., Wang, S., Abdelrasoul, M., Li, Y., & He, W. (2015). Identifying at-risk students for early interventions—A time-series clustering approach. *IEEE Transactions on Emerging Topics in Computing*, 5(1), 45–55.

Im, T., & Kang, M. (2019). Structural relationships of factors which impact on learner achievement in online learning environment. *The International Review of Research in Open and Distributed Learning*, 20(1). https://doi.org/10.19173/irrodl.v20i1.4012

Kovanović, V., Mazziotti, C., & Lodge, J. (2020, March 31). *Special Section on Learning Analytics for Primary and Secondary Schools-Call for Papers*. https://learning-analytics.info/index.php/JLA/announcement/view/161

Law, N., & Liang, L. (2020). A Multilevel Framework and Method for Learning Analytics Integrated Learning Design. *Journal of Learning Analytics*, 7(3), 98–117.

Li, K. C., Lam, H. K., & Lam, S. S. Y. (2015). A Review of Learning Analytics in Educational Research. Technology in Education. *Technology-Mediated Proactive Learning*, 173–184.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). JASP: Graphical Statistical Software for Common Statistical Designs. *Journal of Statistical Software*, Articles, 88(2), 1–17.

Macarini, L. A., Cechinel, C., Santos, H. L. dos, Ochoa, X., Rodés, V., Alonso, G. E., Casas, A. P., & Díaz, P. (2019). Challenges on implementing Learning Analytics over countrywide K-12 data. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 441–445.

Nawrot, I., & Doucet, A. (2014). Building engagement for MOOC students. *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion. the 23rd International Conference, Seoul, Korea*. https://doi.org/10.1145/2567948.2580054

Ochoa, X., Lang, A. C., & Siemens, G. (2017). Multimodal learning analytics. *The Handbook of Learning Analytics*, 1, 129–141.

Sancho, M., Cañabate, A., & Sabate, F. (2015). Contextualizing learning analytics for secondary schools at micro level. *2015 International Conference on Interactive Collaborative and Blended Learning (ICBL)*, 70–75.

Schnitzler, K., Holzberger, D., & Seidel, T. (2020). All better than being disengaged: Student engagement patterns and their relations to academic self-concept and achievement. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-020-00500-6