Analysis of Student Learning Behavior based on Moodle Log Data

Xiangfeng Tan
Graduate School of Engineering
Tokyo University of Technology
1404-1 Katakuramachi, Hachioji
Tokyo 192-0982, Japan
g512002511@edu.teu.ac.jp

Jinhua She
Graduate School of Engineering
Tokyo University of Technology
1404-1 Katakuramachi, Hachioji
Tokyo 192-0982, Japan
she@stf.teu.ac.jp

Shumei Chen
Department of Liberal Arts
Tokyo University of Technology
1404-1 Katakuramachi, Hachioji
Tokyo 192-0982, Japan
chin@stf.teu.ac.jp

Sumio Ohno
Graduate School of Computer Science
Tokyo University of Technology
1404-1 Katakuramachi, Hachioji
Tokyo 192-0982, Japan
ohno@stf.teu.ac.jp

Hiroyuki Kameda
Graduate School of Computer Science
Tokyo University of Technology
1404-1 Katakuramachi, Hachioji
Tokyo 192-0982, Japan
kameda@stf.teu.ac.jp

2021 14th International Conference on Human System Interaction (HSI) | 978-1-6654-4112-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/HSI52170.2021.9538680

***Abstract*—The importance of online learning grows rapidly in higher education nowadays because it allows students to learn without time and space constraints, which are required by traditional classroom-based instructional models. Moodle is one of the most used learning management systems (LMSs) for online learning. This study investigated factors that influence student academic performance. We statistically analyzed Moodle log files of a beginner's Chinese course at Tokyo University of Technology and found that the most important factors that affected student academic performance were four: learning time zone, access times for optional learning material, completeness of course material, and assignment-submission time.**

Keywords—e-learning, higher education, learning behavior, learning performance, learning management system (LMS), Moodle.

I. Introduction

Electronic learning (e-learning) has been developing with the advancement of science and technology. COVID-19 pandemic speeded up the use of online learning at universities in Japan. learning management system (LMS) provides us a teaching platform for managing online education. In the 2000s, the functions of LMSs were very simple. The basic function is to display text contents on Web pages [1]. The functions of LMSs have been gradually enhanced and new functions have been developed to power up the LMSs. Moodle is one of the most widely used open-source e-learning platforms. It is easy to use Moodle to edit courses, to distribute video tutorials, to conduct online tests, and to collect answers [2]. It provides students and teachers a platform to communicate in a real-time fashion and allows a teacher to complete an entire teaching process from registration to a final examination in an online style [3].

According to the statistics [4], about 40% of universities in Japan use LMS for teaching and learning in 2019. Moodle Association of Japan decided to offer 3-month free academic licenses to accelerate the use of Moodle in June 2020 [5]. Nowadays, it is a hot topic how to analyze online teaching quality. Those analysis results help us to find and overcome drawbacks in training courses, to adjust the level of curriculum design to suit students' level, and to improve academic performance [6]. An important issue for online education is the analysis of the effectiveness of online education. For example, how to use the features of Moodle to analyze the overall course performance. This can be carried out by studying students' online behavior [7]. In the past few years, data mining and statistical analysis have made great progress in capturing and analyzing student online behavior. The analysis of the log data for a Moodle system was used to classify students to find students with similar characteristics, to predict students' scores, to evaluate the correlation between online learning behavior and final scores, and to find factors that had a big impact on scores, and finally to improve online courses [8].

In this study, we analyzed students' learning behavior in an online learning environment. We collected the log data for a course at Tokyo University of Technology. Then, we investigated factors that affect students' scores. As a result, we finally found four factors that influence student academic performance.

II. Previous Work

Many studies have been carried out to investigate the use of Moodle. For example, Data mining is used to analyze the log data of Moodle to classify students' learning styles [9]. It is also used to analyze students' online learning behavior, to understand the relation quantitatively, and to provide visualization of Moodle disciplined student activities [10] [11].

Linear regression, clustering, and other commonly used statistic methods are also effective for this kind of analysis [12]. By analyzing LMS usage through log data, it is easy to understand students' learning status and to predict learning performance [13]. A study at Central Queensland University shows that there was a relationship between page views and final scores [14]. Fourteen variables among 29 variables were found to have a big impact on final scores [15].

Exploring the log data of Moodle shows that we need to pay close attention to the number of clusters. A suitable number is a key to suitable analysis in the later steps [16]. While SPSS and other professional programs can be used for this purpose, Microsoft Excel is also useful to perform basic data analysis [17].

978-1-6654-4112-4/21/$31.00 ©2021 IEEE

TABLE I. SIX VARIABLES USED TO EVALUATE STUDENT ACADEMIC PERFORMANCE

Name	Meaning
V_1	Learning time zone
V_2	Access times for optional learning materials
V_3	Times of late submission
V_4	Times of on-time submission
V_5	Completeness of course material
V_6	Student's major

III. METHODOLOGY

The purpose of this study is to find factors related to online-learning behavior that affect student academic performance by statistical analysis to help students improve their course performance and give suggestions to teachers to improve the design of online courses.

Base on preparatory analysis, we extracted six factors (Table I) to evaluate student academic performance. The study issue considered here is *how the six factors affect student academic performance for a beginner's Chinese course*.

When a student uses his personal account to access Moodle, Moodle records the student's behavior, including time spent clicking on courses, length of reading, login IP, study time, etc., on a Web server. We use Microsoft Excel to perform data pre-processing of raw data. This study used correlation analysis, regression analysis, analysis of variance data analysis, and other methods for statistical analysis, and used SPSSAU for data visualization.

After carrying out the first analysis on all log data, we extracted six variables (Table I) to further analyze the relationship between them and student academic performance.

IV. INSTRUMENTATION

This study analyzed student academic performance in online learning by examining students' behavior recorded in the Moodle system. The log data of a total of 172 students in a beginner's Chinese course, which is offered to second-year students, in the spring semester, 2020, were analyzed. Students were from four schools of the university (School of Engineering, School of Computer Sciences, School of Media Science, and School of Bioscience and Biotechnology) enrolled to participate. It was an elective course. The course was taught in a fully online mode due to the COVID-19 pandemic.

Students' profiles were exported from the university administration system. They were slightly reformatted, saved in the Excel spreadsheet format, and were then imported into SPSSAU for further analysis.

We analyzed data related to the beginner's Chinese course in 2020. Course materials include instructional videos, reports, tests, and optional learning materials. There is an assignment at the end of each class. They are uploaded to Moodle construct lessons. The course has 6 parts that are explained below.

A. Learn points provided by video lessons

The first part of the course is a video tutorial that consists of 3 sections and each has 5 minutes. This part is a teaching content pre-recorded by the teacher. Fig. 1 shows a typical video lesson content.

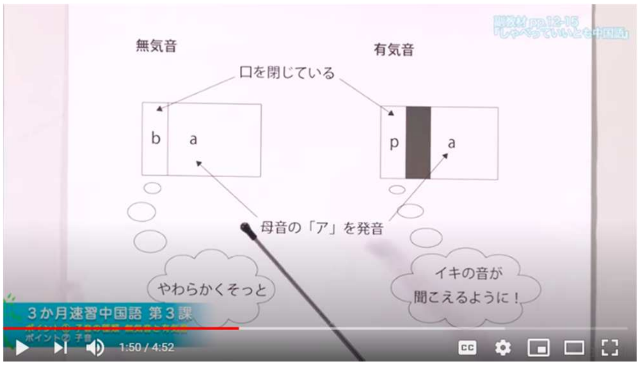

Fig. 1. Screenshot of video lesson

B. Write a learning report

After the completion of the first part of the video content, students need to write a brief impression of learning, learning points listed in this lesson, and difficulties they felt.

C. Take a test

In this part, students need to complete an examination within a specified time. The content of the examination is about the understanding of the video learning content. There are a total of 4 questions that have a multiple-choice type in the examination.

D. Work on assignment

The content of this part is Chinese comprehension. It is the most important part of the learning course. We prepared different types of questions to correctly collet the levels of students' understanding. All the assignments are handwritten and corrected by the teacher.

E. Access to optional learning material

The content of this part is not compulsory for learning. The teacher designed this part to provide learning content for highly motivated students.

V. RESULTS

Figure 2 shows the attendance completion status of the course. A course completion rate of more than 95% indicates a good level of participation in the online course.

Figure 3 shows the students' learning time zones. Most students concentrate on finishing their studies in the afternoon or evening.

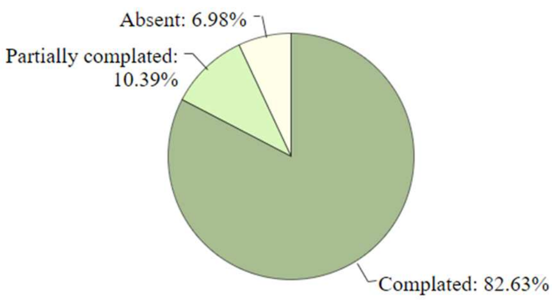

Fig. 2. Completion status of the course

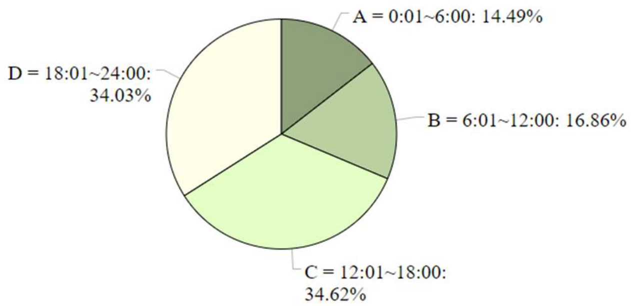

Fig. 3. Learning time zones

TABLE II. SUMMARY TABLE OF ONE-WAY ANOVA BETWEEN SCORES AND STUDENTS' MAJORS

Items	Category	*n*	Mean	Std.	*F*	*p*
Scores	SE	35	79.43	9.68	0.706	0.550
	SC	67	75.07	21.35		
	SB	50	75.60	9.29		
	SM	20	77.50	9.67		
	Total	172	76.40	15.25		

SE: School of Engineering, SC: School of Computer Science, SB: School of Applied Biology, SM: School of Media.
*: $p < 0.05$, **: $p < 0.01$

TABLE III. PEARSON'S CORRELATION COEFFICIENT BETWEEN SCORES AND LEARNING TIME ZONES

Learning time zone	Score
Learning time-zone A (00:01-06:00)	−0.399**
Learning time-zone B (06:01-12:00)	0.066
Learning time-zone C (12:01-18:00)	0.227**
Learning time-zone D (18:01-24:00)	0.120

*: $p < 0.05$, **: $p < 0.01$

Analysis of variance (ANOVA) is used to determine whether any of the differences between two or more population means are statistically significant. A significance level of 0.05 ($p = 0.05$) is usually used as a criterion. Table II represents the result of the ANOVA that investigated the relationship between scores and students' majors. The value of F is 0.706, which reaches significance with a p-value of 0.550 (which is higher than the 0.05 level). This means that majors had little connection with scores.

Pearson's correlation coefficient is used to investigate the relationship between two variables. The coefficient is in the range [−1, 1]. A value of 1 means that the two variables have a perfectly positive linear relationship; a value of −1, a perfectly negative linear relationship; and a value of 0, no linear correlation.

Pearson's correlation coefficient between scores and access time for optional learning materials is 0.304. It shows significance at the 0.01 level. It indicates a significant positive correlation between scores and access time for optional learning material.

Pearson's correlation coefficient between scores and learning time zones (Table III) shows that, while Learning time-zones B and D have no correlation with scores, Learning time-zone A has a significant negative correlation with scores (the coefficient is −0.399) and Learning time-zone C has a significant positive correlation with scores (the coefficient is 0.227).

TABLE IV. LINEAR REGRESSION ANALYSIS RESULTS [EQ. (1)]

Item	Coefficient	95% CI	VIF
Intercept term	7.290 (1.475)	−2.400 ~ 16.979	-
V_3	7.358** (6.615)	5.178 ~ 9.538	2.839
V_4	7.253** (9.267)	5.719 ~ 8.787	3.634
V_5	2.365** (4.072)	1.226 ~ 3.503	1.542
n (sample number)	172		
R^2	0.578		
Adjusted R^2	0.570		
F test	$F(3,168) = 76.609$, $p = 0.000$		
D-W test	2.003		

Dependent variable: Scores
*: $p < 0.05$, **: $p < 0.01$. The results of t test are shown in parentheses.

The analysis of ANOVA and Pearson's correlation coefficient shows that V_6 has litter relationship with student scores, V_2 has strong relationship with student scores, and it depends on the time zone whether V_1 have relationship with student scores or not.

Next, we carried out linear regression to verify the above result and to further derive a model that describes the relationship between the six variables in Table I and student scores. Stepwise regression analysis that used V_1-V_6 as independent variables shows that only V_3, V_4, and V_5 has a linear relationship with the scores. The linear regression model is

$$\text{Score} = 7.290 + 7.358V_3 + 7.235V_4 + 2.365V_5. \quad (1)$$

Table IV shows the linear regression analysis results. $R^2 = 0.578$. It means that V_3, V_4, and V_5 can explain 57.8% changes in scores. Since the model passed the F test ($F = 76.609$, $p = 0.000 < 0.05$), all VIF (variance inflation factor) values in the model are less than 5. This means that there is no collinearity problem. *D-W* test is near 2. It means that the model does not have autocorrelation. The regression coefficient value of V_3 (the times of late submission) is 7.358 ($t = 6.615$, $p = 0.000 < 0.01$). The regression coefficient value of V_4 (the times of on-time submission) is 7.253 ($t = 9.267$, $p = 0.000 < 0.01$). And the regression coefficient value of V_5 (completeness of course material) is 2.365 ($t = 4.072$, $p = 0.000 < 0.01$). The results show that V_3, V_4, and V_5 have a significant positive linear relationship with scores.

VI. CONCLUSION

This study examined the relationship between six factors and student academic performance for a beginner's Chinese course for second-year students in the spring semester, 2020.

ANOVA showed that majors had little connection with student academic performance. This is easy to understand because a beginner's Chinese course is a course of liberal arts that has connection with students' interest in languages rather than in specialties.

Pearson's correlation coefficient showed two facts: First, the scores have a strong relationship with access time for optional learning material. If a student spends time to accesses optional learning material, it means that he is very active in learning. No doubt, he will obtain a high score. Second, students learn during 00:01-06:00 may not help them to achieve a good score. Thus, good study habits are important.

Linear regression analysis shows that the completeness of course material, the times of on-time submission, and the times of late submission have significant positive effects on the scores. This reveals the importance of completing the course and assignment.

This study provided us some hints to improve course design. For example, optional learning material may stimulate students' interest in language learning. However, the examination of log data of optional learning material shows that most students accessed them at least one time, but only a very small number of students accessed them multiple times. Thus, it is desirable to prepare optional learning material at different levels so that all students can enjoy it. Note that it is important to complete assignments. We may need to build a help desk to push students to finish their assignments on time.

A synthetic analysis of log data is of great importance in improving teaching and learning and will be carried out in the future.

ACKNOWLEDGMENT

This work was supported by JSPS Grants-in-Aid for Scientific Research under Grant 17K02947.

REFERENCES

[1] H. Coates, R. James and G. Baldwin, "A Critical Examination of the Effects of Learning Management Systems on University Teaching and Learning", *Tertiary Education and Management*, vol. 11, pp. 19-36, September 2005.

[2] J. Cole, and H. Foster. Using Moodle: Teaching with the Popular Open Source Course Management System, 2007.

[3] P. Legoinha, J. Pais and J. Fernandes, "O Moodle e as comunidades virtuais de aprendizagem," *VII Congresso Nacional De Geologia*, 2006.

[4] Academic eXchange for Information Environment and Strategy, "Research report on the use of ICT in higher education institutions," *Academic eXchange for Information Environment and Strategy*, 2020. [Online]. Available: https://axies.jp/_media/2020/03/2019_axies_ict_survey_v2.1.pdf]. [Accessed: June. 4, 2021] (in Japanese).

[5] A. Kuwata, "Moodle provided free of charge to high schools and universities nationwide," June 2020. [Online]. Available: https://reseed.resemom.jp/article/2020/06/22/378.html. [Accessed June. 4, 2021] (in Japanese).

[6] A. Anzer, H. A. Tabaza, and J. Ali, "Predicting Academic Performance of Students in UAE Using Data Mining Techniques," *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2018.

[7] Y. Zhang, A. Ghandour, and V. Shestak, "Using Learning Analytics to Predict Students Performance in Moodle LMS." *International Journal of Emerging Technologies in Learning (iJET)*, 2020, 15(20):102.

[8] C. Romero, S. Ventura, E. García, "Data mining in course management systems: Moodle case study and tutorial," *Computers & Education*, 2008, vol. 51, pp. 368–384.

[9] P. Madura, K. S. Lg, and H. Masahito, "Detecting Learning Styles in Learning Management Systems Using Data Mining," *Journal of Information Processing, 2016, Vol. 24, pp. 740-749.*

[10] Estacio, R. R., and R. Jr, "Analyzing students online learning behavior in blended courses using Moodle," *Asian Association of Open Universities Journal,* 2017, pp.52–68.

[11] W. Villegas-Ch and S. Luján-Mora, "Analysis of data mining techniques applied to LMS for personalized education," *2017 IEEE World Engineering Education Conference (EDUNINE)*, 2017.

[12] A. Charitopoulos, M. Rangoussi, and D. Koulouriotis, "Educational data mining and data analysis for optimal learning content management: Applied in moodle for undergraduate engineering studies," *2017 IEEE Global Engineering Education Conference (EDUCON)*, 2017.

[13] T. Yu, and IH. Jo, "Educational technology approach toward learning analytics: relationship between student online behavior and learning performance in higher education, " *Proceedings of the fourth international conference on learning analytics and knowledge*, 2014.

[14] C. Beer, K. Clark, and D. Jones, "Indicators of engagement," *Curriculum, technology & transformation for an unknown future. Proceedings ASCILITE Sydney 2010*, 2010.

[15] Á. F. Agudo-Peregrina, S. Iglesias-Pradas, M. Á. Conde-González, and Á. Hernández-García, "Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning," *Computers in human behavior*, 2014*, Vol. 31, pp. 542-550.*

[16] A. Bovo, S. Sanchez, O. Héguy, and Y.Duthen, "Yves Analysis of students clustering results based on Moodle log data," *6th International Conference on Educational Data Mining - EDM 2013*, 2013.

[17] A. Konstantinidis, and C. Grafton, "Using Excel macros to analyse Moodle logs," *2nd Moodle Research Conference,* 2013.