Journal of Education and Social Science Vol.1, No.1 (2020), pp.15-22
http//dx.doi.org/10.46410/jess.2020.1.1.03
Received: March 18, 2020 Reviewed: May 7, 2020 Accepted: June 10, 2020

The Effect of LMS Data on Total Score of Learner's in Lifelong Distance Education Center: A Learning Analytical Approach

Jong-Teak Seo [1]*, Young-gi Kim [2] and Ra-chel Ju [3]

[1] Kookmin University 1; suho1793@hanmail.net
[2] Kyunghee Cyber University 2; smileknock@gmail.com
[3] Hanyang University 3; rachelju@naver.com
*Correspondence: suho1793@hanmail.net Tel.: Mobile +82-10-3235-5643

Abstract: The purpose of this study is to analyze the effects of gender, age and educational experience on total scores for determining completion (over 60) and under commission (under 60). For this study, I collected log data (such as regularity of learning start interval, total number of learning and total learning time) and personal background data (number of courses of experience). The data were collected from 1,130 learners for 15 weeks and analyzed based on their learning analysis. The results were as follows: First, among the log data, the total learning time and number of learning had significant effects on the total score. The more the number of accesses to the LMS, the longer the learning time and the higher the total score. Second, among the personal background data, age had a significant effect on the total score. As they get older, they have a clearer sense of purpose for learning, which means that they are more likely to complete.
The data used in this study was used when the learner started signing up to LMS for learning and collected the consent in advance from the personal information agreement.

***Keywords**: Learning Analytics, Lifelong Distance Education, LMS (Learning Management System), Log data, Personal background data.*

1. Introduction

In a distance education environment, accurate information about a learner's learning status acts as an important factor in determining success or failure of learning. There is an increasing need to provide learning status information to learners [1]. At this time, if the data stored in the learning management system (LMS) is analyzed, the learner's learning type can be grasped, and by using this, it can be sufficient educational information data that can be used to predict the learner's learning status [1] [11]. In addition, since all records from the start of learning to the end are stored, a vast amount of data is created and numerous information data are accumulated. However, since most LMSs do not provide tools for instructors to track and analyze all learners' learning activities, learning analytics is a useful approach in predicting a learners' completion of data generated through LMSs. It can be such as [31]. Learning analytics can collect and analyze a learner's activity data generated during the learning process on the LMS to find the regularity in the data and predict future events based on this, as well as analyze learner-related data generated on the LMS. Nowadays, it is receiving attention in terms of being able to evaluate and manage learning outcomes [26].

Prior studies using LMS data in relation to the Department of Learning Analysis include regularity of learning time intervals, total learning time, number of learning connections, attendance, degree of interaction between instructors and learners, assignment of tasks, participation in discussions, detailed predictors of major areas, etc. Academic achievement is predicted by analyzing various factors [1] [2] [9] [19] [20]. Among these, a learner's studies is important information that can be observed and analyzed to understand a learner's online behavior so as to identify improvement points. It affects the learner's total score and helps their learning patterns to be regular. The number of times and total learning time invested for learning were selected as log data. In addition to log data, personal background data is stored in the LMS of the Distance Continuing Education Center. There are previous studies that studied the possibility of incompleteness based on age, gender, final education, occupation, grade, department, area, high school, admission selection, department satisfaction and major suitability corresponding to personal background data [3][14][29]. However, because research is mainly focused on the learner's personal environmental factors, attempts to analyze what the learner's personal background cause may be are insufficient, which may be a direct cause of the completion. It simply remains in a negative view of limiting the completion to a learner's personal problem or to a maladjustment problem [12]. Therefore, this study compares and analyzes the actual log data (regularity of learning start interval, total number of learnings and total learning time) and personal background data (gender, age, educational experience and the number of courses experienced) of learners of distance learning centers. Through this, we would like to verify how each factor affects the total score based on the learning analysis.

2. Theoretical Background

2.1. Concept of Learning Analytics

In the recent years, as the amount and types of learning data have increased and analytical techniques for them have been developed, learning analysis is getting more attention [17].

Learning analytics includes taking a step further from an educational data mining analysis method whose main goal is efficiency in data analysis and controlling learning outcomes by applying teaching and learning prescriptions using data mining results [4] [22]. Therefore, learning analytics is a research field that applies the predictive model found in the system of education to education beyond the simple technical meaning applied to education data. It provides a platform for learners to successfully complete learning or is useful as a new prediction tool. It can be utilized in [13] [15]. In the online learning environment, learning analytics prescribes learning by predicting learning outcomes and providing advice by analyzing data generated by various learning activities and digital footprints that learners left in the learning process via a computer. Thus, learning analysis becomes possible [15]. Therefore, by providing the results of the learning activities in a learner's online learning environment visually, it helps the learners in self-directed learning, setting goals for themselves and managing the course and time to reach the results [25] [30].

2.2. Learning Analytics Analysis Procedure

The analysis process of learning analytics starts from the data collection stage. In this step, personal learning activities and data related to learning elements within the LMS are collected. Next, after processing data, such as grouping of data, classification and analysis of related data, the result is provided to the learner through a dashboard or other visualized graphic means [21].

2.2.1. Data Collection and Processing

In the data collection and processing stage, it can be divided into static data that does not change well in a short time and dynamic data generated during the learning process. Static data has a refined form, i.e., it is systematically arranged according to the scale for each clearly distinguished variable, and it is easy to combine between different databases through a relational database design. Static data includes the characteristics of learners managed by the learning management system (LMS), demographic information (age, gender) and results of prior learning (number of courses experienced, educational experience). Dynamic data is data generated during the course of learning, and includes learner log data (regularity of learning start interval, total learning time and number of learning) and learning performance data (attendance, assignment, test and total score). Raw data of dynamic data is usually unstructured 'big data', which has no fixed form or cannot be operated on. Therefore, it cannot be analyzed in full without pre-processing and filtering. The data collection and processing stage is a process of filtering only the main data required through the characteristics of the data and converting it into an analytical form [15] [28].

2.2.2. Data Analysis and Prediction

For successful analysis and prediction, it must be possible to sufficiently reflect the characteristics of learners who fail or succeed in learning, along with selecting an effective algorithm. In the analysis and prediction phase, the processed data are analyzed in earnest [15] [24].

Prediction has the advantage of analyzing the learning patterns of learners with or without the possibility of unfinished risk, and suggesting appropriate measures at individual time points. For example, a warning message can be sent to a learner approaching an unfinished risk level and guided in learning activities to enter the normal or safety level.

3. Research Method

3.1. Research Problem

The total score that determines whether the LMS log data (regularity of learning start interval, total learning time and total number of learning) and personal background data (number of subjects, gender, age and educational experience) in the distance learning education are completed (60 points or more: completed, less than 60 points: not completed) How does it affect?

1.1 How does the regularity of learning start interval affect the total score?
1.2 How does the total learning time affect the total score?
1.3 How does the total number of lessons affect the total score?
1.4 How does the number of course experiences affect the total score?
1.5 How does gender affect your total score?
1.6 How does age affect your total score?
1.7 How does education experience affect the total score?

3.2. Research Subject

The study subjects were learners who had conducted remote classes at A Remote Lifelong Education Center for 15 weeks, with 771 male students and 359 female students, totaling to1,130 students. [Table 1] shows the total learner ratios.

Table 1. Subject Gender Ratio

Male		Female		Total	
Number of people	ratio(%)	Number of people	ratio(%)	Number of people	ratio(%)
771	68.2	359	31.8	1,130	100

3.3. Definition of Terms

3.3.1. Log Data

In this study, log data refers to 'regularity of the learning start interval' , 'total learning time', and 'total learning frequency' calculated based on all records made when the learner accesses the LMS.

1. Regularity of learning start interval

This refers to how regularly the learner accesses the learning window and the standard deviation of the values obtained in minutes by the difference between the start times of each learning.

2. Total learning time

For learning, the sum of all the time from the learner to access the learning window, start learning, complete learning, and leave the learning window was calculated as the total learning time.

3. Total number of learning

The total was calculated by adding all the records that the learner accesses the corresponding parking learning window for learning within a predetermined learning period. Even if it was accessed multiple times in a parking lot, all of the values were added and included in the total number of lessons.

3.3.2. Personal Background Data

This refers to the gender, age, educational experience entered by the learners when they sign up for membership, and the number of courses experienced by learners automatically managed by the LMS.

4. Course Experience

The number of course experiences refers to the number of courses one learner has taken at the institute, that is, the course history.

5. Education background

This is the information entered by learners when they first signed up for membership on the website, and the types were collected in three types: high school graduate, college graduate, and 4-year university graduate. The three variables used in the analysis were analyzed by dummy variables

6. Gender

Gender was used for the actual input value, and was used for analysis by dummy variable.

7. Age

It was analyzed using the actual age.

8. Total score

The total score represents the final score calculated based on the 100 points at the time when all the learning processes are completed according to the evaluation criteria. The final grade consists of midterm (30%), final exam (30%), attendance (15%), and assignment (25%). The total score was extracted by selecting one subject with the highest number of participants and using only 1,130 data excluding missing values.

3.4. Data Analysis Results

3.4.1. Descriptive Statistics

The average and standard deviation, minimum and maximum values, skewness and kurtosis of the total scores of the dependent variables were obtained through general trends for the analysis variables and are presented in [Table 2]. It can be seen that all the variables corresponding to the measured variables form a normal distribution with a skewness of ± 2 or less and a kurtosis of ± 7 or less.

Table 2. Descriptive Statistics

Variable	Average	Standard Deviation	Minimum value	Maximum value	Skewness	Kurtosis
Regularity of learning start interval	4.037	.782	.731	8.660	1.111	3.162
Total learning time	1,389.404	406.451	3.4	3,744.0	1.699	5.118
Total number of learning	47.009	14.247	7.000	129.000	.708	1.489
Course Experience	10.050	7.6497	1.0	44.0	1.611	2.437
Gender	.682	.466	0.0	1.0	-.784	-1.388
Age	27.804	5.986	20.0	53.0	1.275	1.485
Educational background	1.453	.641	1.0	3.0	1.105	.081
Total score	74.387	16.380	0.0	98.0	-1.928	4.684

3.4.2. Correlation Analysis

In this study, the correlation between each variable was analyzed. The regularity of the learning start interval showed a negative (-) correlation with the total learning time and number of learning, age, educational experience and total score, but did not show any correlation with the number of subjects and gender. The total score was significant for all the variables except for the number of courses. Negative (-) correlation with regularity at the beginning of the learning interval was found to be positive (+) between the remaining variables [Table 3].

Table 3. Correlation Analysis

Variable	1	2	3	4	5	6	7	8
1. Regularity of learning start interval	1							
2. Total learning time	-.228**	1						
3. Total number of learning	-.500**	.302**	1					
4. Course Experience	.016	-.003	-.072*	1				
5. gender	-.001	.037	.059*	087**	1			
6. age	-.075*	.159**	.095**	.095**	.342**	1		
7. Educational background	-.070*	.051	.087**	.045	.156**	.216**	1	
8. Total score	-.190**	.240**	.325**	.021	.084**	.168**	.108**	1

* $p < .05$, ** $p < .01$

3.4.3. Regression

Table 4 shows the results of the multiple regression analysis. The final model with all independent variables had an F value of 26.429 and significant probability of .000, which showed a significant result in the statistical regression model at a significance level of .05. It was concluded that 15.9% of the total score (adj. $R^2 = 15.3$) was accounted for by independent variables.

Table 4. Multiple Regression Analysis

Dependent variable	Independent variable		Denormalization factor		Standardization factor	T	p
			B	Standard error	β		
Total score		(a constant)	43.060	4.700		10.013	.000
	Log data	Regularity of learning start interval	-.332	.707	-.016	-.469	.639
		Total learning time	.004	.001	.101	3.513	.000
		Total number of learning	.221	.031	.257	7.186	.000
	Personal Back ground data	Course Experience	-.071	.057	-.037	-1.258	.290
		Gender dummy	-1.378	.927	-.043	-1.486	.138
		Age	.344	.075	.131	4.581	.000
		Educational background dummy 1	2.654	.961	.080	2.762	.006
		Educational background dummy 2	-.568	1.666	-.010	-.341	733
R^2(adj. R^2) = .159(.153) F = 26.429, p = .000							

*p < .05, n = 1,130 , High school =0

4. Discussion

Through this study, we tried to find out how log data and personal background data stored on the web affect each the total score. Through this, a study was conducted to find a way to predict incompletely in advance, to inform learners at risk of unfinished risks, as well as to inform risks of unfinished risks by using SMS, as well as various warning messages such as telephone and e-mail, and induce them to actively study.

5. Conclusion

This study is based on objective data because research was conducted based on actual LMS data of learners using distance learning centers. The results of the study were summarized as follows.

First, it was found that the total learning time and number of learning had a significant effect on the total score. In a previous study, Kim (2003) stated that the longer the time required for learning access, the higher the total score. Kang, Kim and Park (2009) showed that the attendance rate, number of discussions, classroom access time and bulletin board posting times significantly influence the prediction of the total score. In addition, Rau and Durand (2000) suggested that the learning time accumulated in LMS is the main variable for predicting the total score. Kwon (2009) showed that the higher the number of connections, responses, and discussions indicating participation in learning help in predicting the higher the total score. In other words, it suggested that it is difficult to expect good learning results without active participation of learners. Jeon and Han (2015) showed that the number of LMS accesses and time of learning in the university online learning process are variables that can significantly predict the total score.

Second, among the personal background data, age and college graduates were found to have a significant effect on the total score. Jeong (2004) stated that the higher the learner's age, the higher the objective consciousness for study and the higher the probability of completion. Lim (2007) said that the higher the learner's age, the less likely they are to finish. Kim (2012) admitted admission, age, gender and job. It was analyzed that only age among classification, residence and final education had a significant relationship with uncompleted. It can be estimated that the younger the age, the more economic burden of tuition provision leads to incompleteness. In particular, in the 20s, when work and study were performed in parallel, there was a clear tendency to give priority to work, and there was a high probability of not completing the study due to lack of learning time [8] [16]. Next, I would like to make the following suggestions for further research. In this study, research was conducted in advance to predict and classify learners who may be at risk of not completing fees. However, research on learning support measures, such as dashboards, in which feedback from professors can be effectively delivered to learners has not been conducted.

References

[1] J. H. Kim, "Impact of Learner's Time Management Strategieson Achievement: A Learning Analytics Approach", The Graduate School of Ewha Womans University, Seoul Korea, 2013.

[2] Y. M. Kim, "Impact of Regularity of Learning Interval, Total Learning Hours and Number of Access to E-learning in a Corporate E-learning Environment on Academic Achievement", The Graduate School of Ewha Womans University, Seoul Korea, 2011.

[3] S. Y. Kim, "An Analysis of College Student Dropouts' Mobility Paths and Structure", The Journal of Educational Studies, Vol. 43, No. 3, pp. 131-163, (2012).

[4] K. G. Kim, "Learner Activity Modeling Based on Teaching and Learning Activities Data", Software and Data Eng, Vol. 5, No. 9, pp. 411-418, (2016).

[5] S. T. Kim, "A correlation between Learning Behavior and Achievement Level of Learners in e-Learning," Korea University of Technology and Education, Chungcheongnam-do, Korea, 2003.

[6] M. S. Kang, J. I. Kim, and I. W. Park, "The Examination of the Variables related to the Students' e-learning Participation that Have an Effect on Learning Achievement in e-learning Environment of Cyber University", Korean Society For Internet Information, Vol. 10, No. 5, pp. 135-143, (2009).

[7] S. Y. Kwon, "The Analysis of differences of learners' participation, procrastination, learning time and achievement by adult learners' adherence of learning time schedule in e-Learning environments", Journal of Learner-Centered Curriculum and Instruction, Vol. 9, No. 3, pp. 1-86, (2009).

[8] Y. R. Kim, "Investigation of the relationship between university students characteristics and dropout factors", Graduate School of Education, Kyungpook National University, Kyungpook Korea, 2012.

[9] J. W. You, "Dropout Prediction Modeling and Investigating the Feasibility of Early Detection in e-Learning Courses", The Korean Association of Computer Education, Vol. 17, No.1, pp. 1-12, (2014).

[10] H. Y. Lee, "Development of prediction models based on the clustered online learners' behavioral patterns in university e-Learning environment", Graduate School of Ewha Womans University, Seoul Korea, 2016.

[11] Y. W. Lim, "A substantial study on the Relationship between students' variables and dropout in Cyber university", Journal & Artical management System, Vol. 11, No. 2, pp. 205-219, (2007).

[12] M. L. Ahn, Y. Y. Choi, Y. H. Bae, Y. M. Ko, and M. H. Kim, "A Literature Review on Learning Analytics: Exploratory study of empirical researches utilizing log data in Korea", Journal of Educational Technology. Vol. 32, No. 2, pp. 253-291, (2016).

[13] J. Y. Chung, M. S. Sun, and M. J. Jeong, "An Analysis of Institutional Factors Affecting on College Dropout Rates" Education Research Institut, Vol. 16, No. 4, pp. 57-76, (2015).

[14] I. H. Cho, Y. J. Park, and J. H. Kim, Understanding Learning Analytics, Seoul : Young sa Park, 2019.

[15] Y. J. Joo, W. J. Shim, and S. M. Kim, "A Study on the Factors Affecting the Drop-out in Corporate Cyber Learning", The Journal of Educational Information and Media, Vol. 14, No. 4, pp. 5-25, (2008).

[16] Y. H. Jeong, "Education Analytics", MEDIA & EDUCATION, Vol. 5, No. 1, pp. 44-49, (2015).

[17] B. Y. Jeong, "Analyses of Learning Achievement and Satisfaction on Demographic Characteristics in Cyber Universities : A Case Study", The Journal of Educational Information and Media, Vol. 10, No. 3, pp. 127-150, (2004).

[18] D. H. Han, and B. J. Chun, "Analysis of Learning Type Factors that Affect e-learning Performance : Centering on the Comparison Analysis of Whole Learners Log and Excellent Learners", Journal of the Korean Data Analysis Society, Vol. 17, No. 2, pp. 897-912, (2015).

[19] S. R. Hwang, "Impact of Learner's Learning Behavior on Achievement: The Moderating Effect of Learning Motivation", The Graduate School of Education Ewha Womans University, Seoul Korea, 2016.

[20] Dyckhoff, A. L., Zielke, D., Bultmann, M., Chatti, M. A., & Schroeder, U, "Design and implementation of a learning analytics toolkit for teachers", Educational Technology & Society, Vol. 15, No.3, pp.58-76, (2012).

[21] Elias, T, Learning Analytics: Definitions, Processes, and Potential. Creative Commons, 2011.

[22] Liao, S. H., Chu, P. H., & Hsiao, P. Y, "Data mining techniques and applications–A decade review from 2000 to 2011", Expert Systems with Applications, Vol. 39, No. 12, pp.11303-11311, (2012).

[23] Long, P, & Siemens, G, Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, Vol. 46, No.5, pp.30-32, (2011).

[24] Macfadyen, L. P., & Dawson, S, "Numbers are not enough. why e-learning analytics failed to inform an institutional strategic plan", Educational Technology & Society, Vol. 15, No. 3, pp. 149-163, (2012).

[25] Rau, W., & Durand, A, "The Academic Ethic and College Grades: Does Hard Work Help Students to "Make the Grade"?", Sociology of Education, Vol. 73, No.1, pp.19-38, (2000).

[26] Shum, S, B, LEARNING ANALYTICS, UNESCO Institute for Information Technologies in Education, Policy Brief. November 2012.

[27] Tinto, V, Leaving college: rethinking the causes and cures of student ttrition. Chicago: University of Chicago Press, 1987.

[28] Wagner, E., & Ice, P, "Data changes everything: Delivering on the promise of learning analytics in higher education", EDUCAUSE Review, Vol. 47, No. 4, pp. 32-36. (2012).

[29] Zorrilla, M. E., Menasalvas, E., Marin, D., Mora, E., & Segovia, J, Web usage mining project for improving web-based learning sites. In Web Mining Workshop Cataluna, pp.1-22, (2015).

