

Research on Online Learning Behavior Analysis and Its Influencing Factors Based on SPOC Data

Wei ZHENG*

School of Literature and Media, Jincheng College of Sichuan University, Chengdu, China

523740505@qq.com

*Corresponding author

Keywords: MOOCS, SPOC, Big data, Learning behavior.

Abstract. MOOCS is an emerging online course model based on *Internet & Education*, and SPOC is a sophisticated online and offline learning model following MOOCS. This teaching model is used by many top universities such as Harvard, MIT, Stanford, UC Berkeley, etc. and SPOC teaching mode provides a guarantee for MOOCS to take root in university campus. In this paper, in order to explore the students' learning behavior and influencing factors under the SPOC teaching mode and provide reference for educational workers to design teaching resources and teaching activities that are more in line with students' learning needs and learning habits, the author makes use of the data of the students' online learning behavior under the SPOC teaching mode, and analyzes the students' learning needs, learning patterns and learning preferences.

Introduction

Following the *Massive Open Online Course* (MOOC), the *Small Private Online Course* (SPOC) is a new type of teaching model which used by large number of top universities such as Harvard, MIT, Stanford, UC Berkeley, etc. The SPOC was first proposed and used by Professor Armand Fox of the UC Berkeley. The small scale and restrictiveness of the SPOC is relative to the large scale and open ended in the MOOC. Small scale means that there are usually dozens to hundreds of students, and the restrictiveness refers to the requirement that students need to meet the restrictions to enter SPOC courses. In the process of using the SPOC platform, there will be a large amount of detailed and reliable login log data and learning behavior data, which provide important support for the analysis of online learning behavior.

The research of SPOC-based online learning behavior and the influencing factors in this paper is aimed at the students in each learning section of the SPOC platform, tracking the acquisition of detailed and accurate learning behavior data, including the allocation and length of login time, the degree of interactive participation, and the accuracy and pass rate of class assignments and quizzes, etc., which are difficult to accurately collect in traditional classrooms. The analysis of learning behavior data based on the SPOC platform can help us understand the performance of students throughout the learning process from multiple perspectives, such as learning speed, learning ability, learning preferences, and collaborative learning. Analyzing the data of the SPOC platform allows teachers to better understand students' learning behavior characteristics, so as to design teaching resources and teaching activities that are more in line with students' learning needs and learning habits, thereby improving the teaching quality and teaching effectiveness.

Analysis of Student's Online Learning Behavior

Based on the record data of the teacher-student learning process on the SPOC platform, the analysis of students' online learning behavior is mainly about the statistical analysis and visual analysis of the behavior patterns (login, browse resources, watch videos, post, do homework, test) of the behavioral subject (teachers, students), the used condition of the behavior object (various resources, online course module), and the time of occurrence of the behavior. Combining the

demographic data of teachers and students to analyze and explore the factors that affect online learning behavior.

The development of foreign network technology is earlier than in China. At the same time, the study of online learning behavior technology in foreign countries takes a long time and has a deeper intensity, and there are a large number of rich empirical research results. George Siemens of Technology Enhanced Knowledge Research Institute at Athabasca University in Canada analyzed data on students' online learning activities to understand and discover students' learning situations and problems. He used the data analysis results to help teachers predict and adjust teaching, to meet the students' learning needs to the greatest degree, and to achieve the possibility of personalized learning. Prof [1]. Leah P. Macfadyen and Professor Shane Dawson collected data from the Biology class on the Blackboard platform for analysis and empirical research at the University of British Columbia in Canada through five undergraduate classes. The collected data were used for correlation analysis, multiple regression analysis, and logistic regression analysis. And the early warning monitoring software system is developed for teachers, who can use this system to identify students at risk and to intervene and provide timely assistance [2].

Compared with the diversity and practicality of foreign research, most of the studies on online learning behavior in China focus on the analysis of the influencing factors of online learning behavior and the development and application of related monitoring software.[3]Xie Weiqi (2004) used data exploration methods to explore web server logs and back-end databases.[4]Song Jiangchun and Chen Wenlin (2005) used web usage exploration to analyze students' visit behavior, frequency, content, and residence time.[5]Yuan Ming and Chen Weijie (2006) conducted research on the learning platform of the online education institute as an exploration target.[6]Xiao Aiping and Jiang Chengfeng (2009) pointed out that the influencing factors of online learning include the network environment, curriculum resources, online interaction, and the students' own reasons.[7]Wei Shunping (2012) used the login data, resource browsing data and formative test data generated by the course as the analysis objects, and used data exploration methods to obtain the characteristics and influencing factors of online learning activities such as login, resource browsing, homework, test, etc.

In addition, there are also studies on the analysis of online learning behavior through the establishment of a network learning monitoring system.[8]Yang Jinlai et al. (2008) realized learning behavior monitoring research based on an online learning platform.[9]Shao Yanli, Sun Lina (2008) realized the online learning behavior monitoring based on the electronic profile.[10]Mao Lijie, Cao Xiaoying (2009) realized the research and design of learning monitoring functions based on the online teaching platform.

Analysis of the Characteristics of Online Behavior and Its Influencing Factors

Based on the data of the learning analysis in the SPOC[11] platform, more learning behavior characteristics can be discovered from the perspectives of learning time, learning activities, learning resources, test scores, etc. The analysis of influencing factors can also focus more on internal factors that are the analysis of students' own factors.

The paper selects the *Web Design*, the professional compulsory course of the Department of Literature and Media of Jincheng College of Sichuan University, as the research object. *Web Design* is a professional compulsory course for undergraduate who majors in the Department of Literature and Media of Jincheng College of Sichuan University. And the course is equipped both with text materials and video teaching materials. According to the 2016-2017 fall semester (According to the calendar of the Jincheng College of Sichuan University, the second semester of 2016-2017 starts from February 20, 2017 to July 9, 2017 for 20 weeks. The course time is from February 20, 2017 to June 11, 2017 for 16 weeks.) as an example, there are 86 undergraduates of the Department of Literature and Media of Jincheng College of Sichuan University who have chosen this course. The 86 students were used as research samples to conduct online learning behavior descriptions and analysis of influencing factors by obtaining these students' online learning behavior records and basic information (such as gender, major).

Through the data table of the background login behavior, the basic login information of the student can be obtained, as shown in Table 1.

Table 1 Basic login information

Course Name	Total Number of Students	Average Number of Student Logins	Maximum Number of Student Logins	Average Student Login Days	Maximum Number of Student Login Days
Web Design	86	29	105	15	47

Then, statistical units of *week*, *day*, and *time period* are used as the classification fields to count students' login status at each period. First, two terms are introduced: weekly login rate and daily login rate. The weekly login rate is that the number of students who log on to the platform in a week is divided by the total number of visits, and the daily login rate is that the number of students who log on to the platform on a single day divided by the total number of students. In this example, the number of logins in a certain time unit is used for the statistics, but the number of logins is not used as the counting unit, because the total number of students is small. If the daily login rate method is used, the weekly login rate will exceed 100%, so the number of logins counts the distribution of logins for each week. The reason for the introduction of the daily login rate is that the number of logins in a short period of time may be affected by repeated logins by individual students, resulting in inconsistency between students' learning activities and facts. For this reason, the introduction of a daily login rate provides statistical analysis of the student's short-term login.

Statistical Analysis Based on Login Time as Classification Field

Count Visits by Week. The first day of the course starts on February 20, 2017. The starting time and ending time for the following weeks are shown in Table 2, and there are 16 weeks of statistical data.

Table 2 Weeks and start and end times of the learning cycle

Relative Order	The Start of the Week	The End of the Week	Relative Order	The Start of the Week	The End of the Week
1	2017-02-20	2017-02-26	9	2017-04-17	2017-04-23
2	2017-02-27	2017-03-05	10	2017-04-24	2017-04-30
3	2017-03-06	2017-03-12	11	2017-05-01	2017-05-07
4	2017-03-13	2017-03-19	12	2017-05-08	2017-05-14
5	2017-03-20	2017-03-26	13	2017-05-15	2017-05-21
6	2017-03-27	2017-04-02	14	2017-05-22	2017-05-28
7	2017-04-03	2017-04-09	15	2017-05-29	2017-06-04
8	2017-04-10	2017-04-16	16	2017-06-05	2017-06-11

The login rate (The number of weekly visits divided by the total number of visits.) was calculated by week. And a histogram of *Learning Time & Weekly Login Rate* was plotted, as shown in Fig. 1.

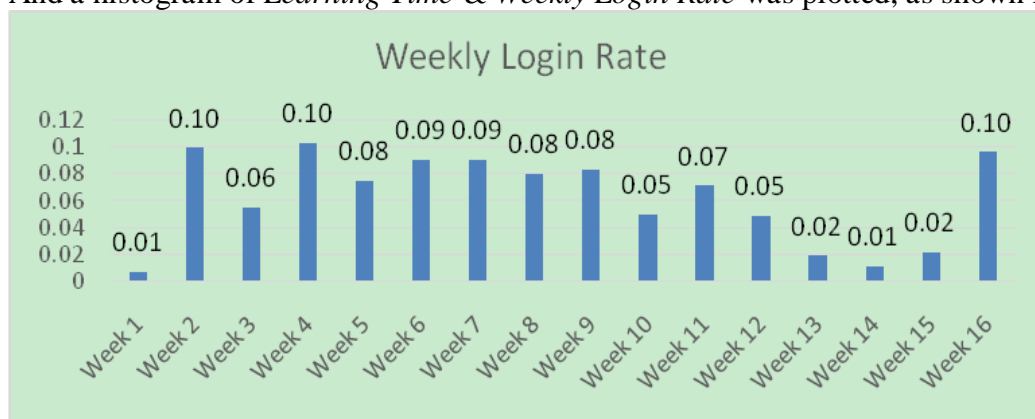


Fig. 1 Learning time & weekly login rate

Explore the Cycle of Group Learning by Week.

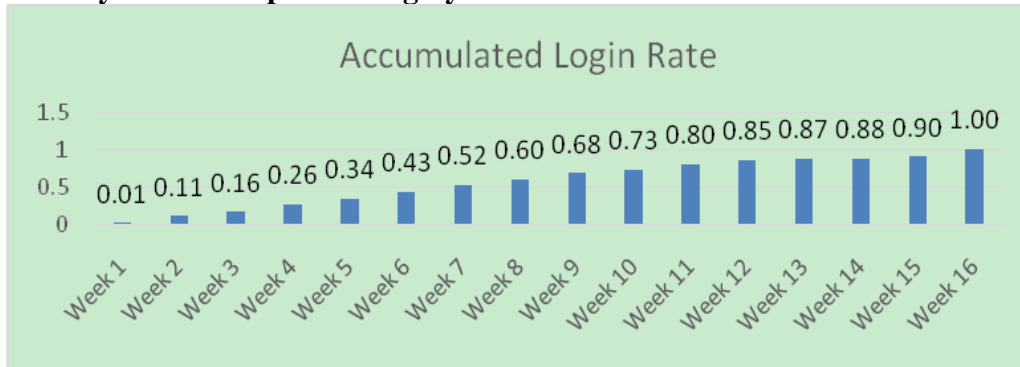


Fig. 2 Time distribution of cumulative student login rate

Count the Number of Logins by Day. As shown in the diagram, if the number of students who log in the course platform is not less than 1/3 of the total number of students on the course platform as the standard for all students to actively participate in the study, there will be 29 days of active learning for the students. The entire learning lasted for 114 days. And there are still 82 days of learning time for students except for the 32 day of rest period (16 weeks). As a result, students' overall coursework is active for slightly more than 1/3 of the entire learning period.

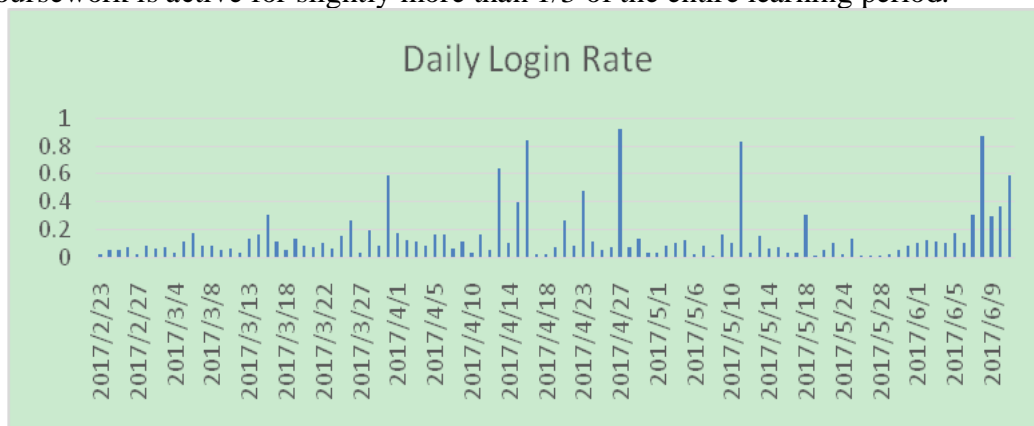


Fig. 3 Learning time & daily login rate

Count the Number of Logins by Time Period. The login rate is calculated for four periods (divided into 0:00-8:00 in the morning, 8:00-12:00 in the morning, 12:00-20:00 in the afternoon, and 20:00-24:00 in the evening). And the login rates for the four periods are as shown in Table 3.

Table 3 Release of learning period

Period	Login Times	Login Rate
Early Morning	37	0.015
Morning	296	0.121
Afternoon	1619	0.661
Evening	497	0.203

Statistical Analysis Based on Student ID. Each student's individual is used as a categorical field for statistics to discover some characteristics of individual learning. For each learning individual, a certain amount of study time is required to complete the necessary learning tasks. "Web Design" as a 2-credit course requires 32 hours of school hours. If the online assessment and offline assessment are to be distributed in a ratio of 2:8, the duration of online learning and offline learning will also be distributed in a ratio of 2:8. Then, the time for online learning should be 6.4 hours. Assuming that each student logs on to the platform one day and their average learning time is 0.5 hours, then one student should have at least 13 logins to the platform to learn. Therefore, the number of days that each student has a login record is first queried. Then, the number of students corresponding to each

class of days is calculated, and a function for accumulating the number of students and the number of login days is finally established, as shown in Eq.1.

$$R_s(t_i) = \sum_{i=1} \frac{S(t_i)}{S_0} \quad (1)$$

Among them: S_0 represents the total number of students, which is 86. t represents the number of days of login. The number of days of login for all students is sorted from highest to lowest and numbered from 1. t_i is the number of i -th login day. $S(t_i)$ indicates the number of students corresponding to the i -th login day. $R_s(t_i)$ indicates the cumulative number of students as a percentage of the total number of students as of the i -th login day.

According to this function and the actual data, draw the curve as shown in Fig. 4.

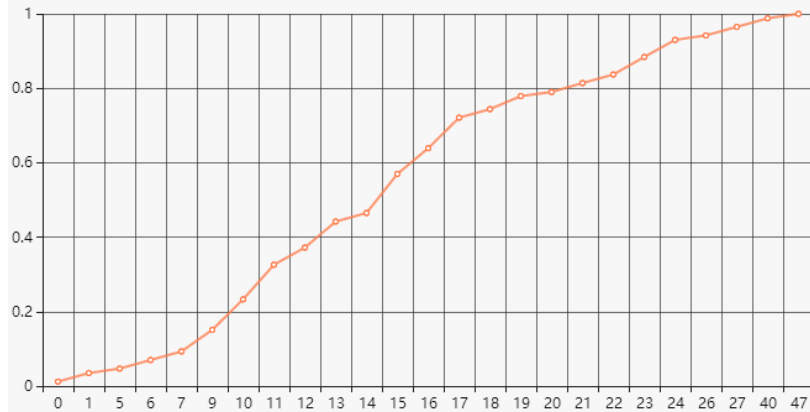


Fig. 4 Relationship between login days and login rates

Fig. 4 shows that the number of students with less than 7 login days is less than 10% of the total number of students, and about 56% of students have more than 13 leaning days, and only 4% have less than 5 leaning days. This shows that most of the students (57%) can devote the necessary time to conduct online learning.

Analysis of the Influencing Factors of Students' Login Days

Relationship between Days of Login and Gender. The students' login days were divided into different groups according to gender groups. Independent sample t test was used, and the results showed that there was a significant difference in the number of days of registration between students of different genders. As shown in Tables 4 and Tables 5.

Table 4. Statistics on days of login for students of different genders

	Number of Samples	Average Value	Standard Deviation	Standard Error
Female	54	18.52	7.748	1.054
Male	32	10.44	4.377	.774

Table 5 Test results of independent samples T

Independent Sample Test										
		Levene's Test for Equality of Variances		The Average Value of t-test for Equality						
		F	Significant	t	Degree of Freedom	Significant (two-tailed)	Significant (two-tailed)	Standard Deviation Difference	95% confidence Interval for the Difference	
									Lower Limit	Lower Limit
Login Times	Hypothesis of Homogeneity of Variance	6.082	.016	5.403	84	.000	8.081	1.496	5.107	11.055
	Non-hypothetical of Homogeneity of Variance			6.179	83.870	.000	8.081	1.308	5.480	10.682

Relationships between Login Days and Professional Levels. According to the professional level (*Chinese International Education*, *Advertising* and *retake course*) grouping to analyze the students' login days, the results show that there are significant differences in the number of days of login between students at different professional levels. As shown in Tables 6 and Tables 7.

Table 6. Statistics on days of login for students of different professional levels

	Number of Samples	Average Value	Standard Deviation	Standard Error	Average 95% Confidence Interval		Minimum Value	Maximum Value
					Lower Limit	Upper Limit		
Chinese International Education	29	16.31	6.714	1.247	13.76	18.86	5	40
Advertising	49	16.43	7.840	1.120	14.18	18.68	6	47
Retake course	8	7.00	5.806	2.053	2.15	11.85	0	15
Total	86	15.51	7.736	.834	13.85	17.17	0	47

Table 7 Analysis results of one-way variance

(I) Major	(J) Major	Average Difference (I-J)	Standard Error	Significant	95% Confidence Interval	
					Lower Limit	Upper Limit
Chinese International Education	Advertising	-.118	1.715	1.000	-4.31	4.07
	Retake Course	9.310*	2.924	.006	2.17	16.45
Advertising	Chinese International Education	.118	1.715	1.000	-4.07	4.31
	Retake Course	9.429*	2.792	.003	2.61	16.25
Retake Course	Chinese International Education	-9.310*	2.924	.006	-16.45	-2.17
	Advertising	-9.429*	2.792	.003	-16.25	-2.61

*. The significance level of average difference is at a level of 0.05.

Relationships between Login Days and Academic Performance. Scores of 85 or above are excellent grades, 70 to 85 are good grades, 60 to 70 are passing grades, and 60 or below are failing grades. It is found that there are significant differences in the number of visits with excellent grades and good grades, passing grades, and failing grades. As shown in Tables 8 and Tables 9.

Table 8 Statistics of student login days for different grades

	N	Average Value	Standard Deviation	Standard Error	Average 95% Confidence Interval		Minimum Value	Maximum Value
					Lower Limit	Upper Limit		
Excellent Grades	35	18.26	6.007	1.015	16.19	20.32	9	40
Good Grades	40	15.20	8.312	1.314	12.54	17.86	5	47
Passing Grades	5	8.20	4.658	2.083	2.42	13.98	0	11
Failing Grades	6	7.67	5.922	2.418	1.45	13.88	1	15
Total	86	15.51	7.736	.834	13.85	17.17	0	47

Table 9 One-way ANOVA results

(I) Level of Grade	(J) Level of Grade	Average Difference (I-J)	Standard Error	Significant	95% Confidence Interval	
					Lower Limit	Upper Limit
Excellent Grades	Good Grades	3.057	1.653	.408	-1.41	7.53
	Passing Grades	10.057*	3.415	.025	.82	19.29
	Failing Grades	10.590*	3.156	.007	2.06	19.12
Good Grades	Excellent Grades	-3.057	1.653	.408	-7.53	1.41
	Passing Grades	7.000	3.388	.252	-2.16	16.16
	Failing Grades	7.533	3.127	.109	-.92	15.99
Passing Grades	Excellent Grades	-10.057*	3.415	.025	-19.29	-.82
	Good Grades	-7.000	3.388	.252	-16.16	2.16
	Failing Grades	.533	4.325	1.000	-11.16	12.23
Failing Grades	Excellent Grades	-10.590*	3.156	.007	-19.12	-2.06
	Good Grades	-7.533	3.127	.109	-15.99	.92
	Passing Grades	-.533	4.325	1.000	-12.23	11.16

*. The significance level of average difference is at a level of 0.05.

Summary

Through the analysis of login data, resource browsing data and formative test data generated by students when they are learning *Web Design* courses, some basic characteristics and influencing factors of online learning behaviors based on the SPOC platform are discovered.

From a group perspective, the distribution of online learning activities in time is mainly concentrated in the afternoon and evening. From an individual perspective, the intrinsic factors affecting students' investment of time of online learning are gender and professional level. Among them, girls spend more time than boys, and there is no obvious difference in time between students of *Advertising* and Chinese international education, and the retake students spend the shortest time. The following research focuses include: how to analyze online learning behavior data from multiple perspectives and levels, how to deeply explore the behavioral characteristics of students in different categories and the impact of student-assisted learning behaviors, and how to improve the analysis of learning behavior to realize the guidance of assessment methods.

Acknowledgement

This research was financially supported by the *Education Department of Sichuan*(No. 18SB0350), China.

References

- [1] Siemens,G.(2010).What are learning analytics [EB/OL].
<http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/>.
- [2] Leah P. Macfadyen, Shane Dawson. Mining LMS data to develop an “early warning system” for educators a proof of concept, J. Computer & Education. 54 (2010) 588-599.
- [3] Xie Weiqi. Web data mining based on distance education platform (in Chinese), J. China Edu Info. 10 (2004) 62-56.
- [4] Song Jiangchun, &Chen Wenlin. Web data mining and its application in learning support service of distance education (in Chinese), J. China Distance Education. 7 (2005) 62-64.
- [5] Yuan Ming, &Chen Weijie (2006). A study on analyzing on-line learning behavior of learner from online education (in Chinese) [J]. Zhejiang Educational Technology, (4) :6-9.
- [6] Xiao Aiping, &Jiang Chengfeng. Investigation and research on web-based learners (in Chinese), J. Open Education Research. 15 (2009) 75-80.
- [7] Wei Shunping. Analysis and research on the characteristics of online learning behavior and its influencing factors (in Chinese), J. Open Education Research. 4 (2012) 81-90.
- [8] Yang Jinlai, Zhang Yixiang, &Ding Rongtao. A study on monitoring learners’ learning behavior based on online learning platform (in Chinese), J. Computer Education. 11 (2008) 65-68.
- [9] Shao Yanli, &Sun Lina. Monitoring of online learning behavior based on electronics (in Chinese), J. China Distance Education. 5 (2008) 38-40.
- [10]Mao Lijie, &Cao Xiaoying. Research and design of monitoring learning function based on online learning platform (in Chinese), J. Software Guide. 9 (2009) 79-81.
- [11]Kang Yeqin. An analysis on SPOC: Post-MOOC era of online education (in Chinese) , J. Tsinghua Journal of Education. 35 (2014) 85-93.