

# Learning Behavior Big Data Tells Us Which Students Can Get a MOOC Course Certificate in Chinese University MOOC

Yong LUO<sup>1,\*</sup>, Zhi-hao XIAO<sup>1</sup>, Jian-ping LI<sup>1</sup>, Zheng XIE<sup>1</sup> and Xiao XIAO<sup>2</sup>

<sup>1</sup>College of Science, National University of Defense Technology, Changsha, Hunan, China

<sup>2</sup>Chinese University MOOC, Beijing, China

\*Corresponding author

**Keywords:** MOOC video, Learning behavior analysis, Multiple regression analysis.

**Abstract.** MOOC (Massive open online courses) is developing rapidly, but the pass rate of the course is very low. It is significant to study the characteristics of learners who have obtained certificates. This article analyzes the video and completion behavior of various learners. The results show that the learner who obtained the certificate did not watch enough videos, but often completed a lot of exercises. It means they have acquired knowledge from other sources. MOOC is only a means of consolidating knowledge. The results of this study can help us find the right application model for MOOC. At the same time, MOOC operators are also reminded to improve the MOOC model for more learners to accept.

## Introduction

The rise of MOOC[1] (Massive Open Online Course) has enabled a large number of high-quality course resources to be accessed and used free of charge. Online open courses are also growing rapidly in China[1]. Future trends in the university triggered by MOOC include: university virtualization, onlineization of university courses, modularization of university courses, the scale of university teaching, and the personalization of university teaching.

The online open course affects the learning model, but it also has many flaws that are difficult to overcome. Among them, poor learning is the biggest problem. Often only 1% of learners are able to obtain a course certificate. What are the characteristics of these learners who have obtained the course certificate, can they help others? Studying their characteristics is very important for improving the pass rate of the course. Of course, these studies are based on the behavior of learners. We should look for their characteristics from the perspective of learning behavior.

Learning behaviors include the rules of online learning, and mining these rules can help learners improve their learning. By mining and analyzing a large amount of learning behavior data from more than 80,000 students in the course, Jiang [3] strives to embody many aspects of learning activities in the MOOC. At the same time, according to the characteristics of Chinese MOOC learning behavior, learners are divided into several groups, and then the relationship between their learning behavior and performance is deeply studied. Geza Kovacs [4] analyzes learners' participation in in-video testing, and studies the peak period of students' thinking. Kloft [5] proposed a method for clickstream data. Wang [6] studied the relationship between the number of discussions, the content of the discussion, and the learning outcomes. And extend content analysis to all communications that occur in the MOOC.

These studies provide many good methods for analyzing learning behavior data. We hope to find out through the research which learning behaviors can help learners obtain course certificates. Our study found no difference in the viewing video behavior between passers and losers based on data from Chinese University MOOC. However, there are significant differences in the behavior of completing the course exercises. This result is surprising, but it is an objective fact. This also shows that MOOC is still only an aid to classroom teaching.

### Data Description and Learner Classification

Since 2015, our team has been working with Chinese university MOOCs to study online learning behaviors, and they collect and provide behavioral data for us. These data were collected by the platform without any filtering or pre-processing. They can reflect the current situation of learners more objectively.

Table 1. Course information

Name	Videos	Active Learners	Exercises
Game Theory	38	14749	8
Advanced Mathematics	129	27664	19
C Programming	81	24684	2

Among these data, we have selected three courses (Table 1) as specific analysis objects. Table 1 also lists the number of videos for each course, the number of active learners, and the number of course exercises. Active learners should include at least one learning record, and those who only register for the course without learning behavior are filtered out.

Each learner corresponds to each row in the data table. Attributes include how long the learner watches the video, how many exercises are completed, and how many times the discussion takes place. The viewing time acquisition method detects the learner's state every 10 seconds or 20 seconds. Therefore, the watch video duration data is an integer multiple of 10(Advanced Mathematics) or 20(Game Theory and C Programming). The unit of data is seconds.

In our research, learners are divided into three categories: Passers, Losers and Abandoners. According to the final test scores, Abandoners refers to those who did not take the test. Losers took the exam but did not pass. Passers are the learners who get the certificate.

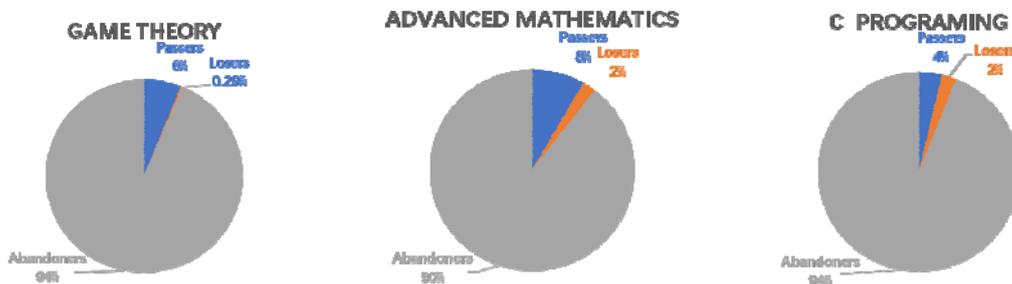


Figure 1. The proportion of three types of learners for 3 courses

Figure 1 shows the percentage of the three types of learners in the 3 courses. It was found that abandoners accounted for the largest proportion (about 90% of all learners). The pass rates for 3 courses are between 4% and 9%. Among the three types of learners, the percentage of students who took the exam but did not pass was the smallest. This shows that students who dare to take the test are likely to pass the test.

### Video Viewing Behavior Analysis

MOOC's main learning path is to watch videos, so video is the most important resource for online courses[7]. The learner who obtained the certificate should watch a large number of videos, and this guess seems reasonable. However, the results of the data analysis are very different from ours. First, we define the behavior of a full watch video, which refers to the behavior that the watch time exceeds 80% of the total video duration. Let's analyze the distribution of the number of full viewing videos. The proportion of learners who watched  $n$  videos is  $p_n$ .

The abscissa of Figure 2 is the number of possible possibilities for the learner to watch the video. When the course contains  $n$  videos, the possible value of the number of videos to watch is  $\{0,1,2,\dots,n\}$ , the corresponding ratio can be written as  $\vec{P} = \{p_0, p_1, \dots, p_n\}$ .  $\vec{P}$  can reflect the

distribution of the number of people watching the video by the learner, reflecting the characteristics of the learner. The ordinate of Figure 2 is the ratio, since all  $\{p_0, p_0, \dots, p_n\}$  values are less than 1, so the upper limit of the ordinate is 1.

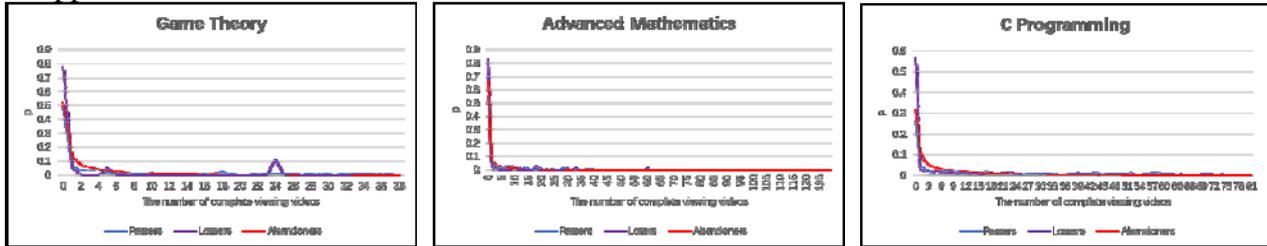


Figure 2. The probability distribution of the complete viewing videos number for 3 courses

Compare the pictures of the three courses and find them amazingly similar, the three curves of each course almost coincide. In other words, Passers (who have obtained course certification) and Abandoners (drop out) see the same amount of video. So, the question is, how does Passers learn knowledge? Without the ability to watch videos, it is impossible to get complete knowledge through online courses. Therefore, we suspect that most of the Passers have mastered the course knowledge, and they don't need to learn by watching the video at all. They are either synchronous learners in the university classroom, and the real knowledge comes from offline learning. The purpose of obtaining a certificate is to test the knowledge they have learned or to seek a challenge.

Therefore, we can't distinguish learner categories by watching video behavior because they have the same characteristics.

### Completing Exercise Behavior Analysis

Below, let's analyze the learner's behavior. There are differences in the difficulty of the coursework. In order to make the analysis and comparison more reasonable, we ignore the homework score. We analyze the number of homework completed by the learner, which is similar to the statistical method of watching the number of videos.

Assume that the course has  $s$  exercises. For three types of learner, count the number of learners who completed the  $k$  exercises and calculate the probability  $\{p_0, p_0, \dots, p_s\}$ .

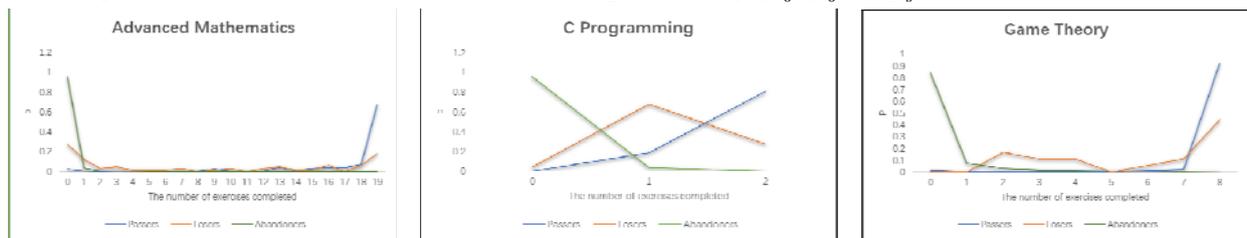


Figure 3. The probability distribution of completing  $k$  ( $k=0,1,2,\dots,s$ ) exercises for 3 courses

Figure 3 shows the probability distribution of the number of exercises completed in the three courses. The abscissa is the possible number of completed exercises. When there are  $s$  exercises in a course, the number of completed exercises is in the range of  $\{0,1,2,\dots,s\}$ . The abscissa of Figure 2 is the number of exercises that the learner may complete. The ordinate is the probability value. Curves of different colors represent different learners. It can be found that different categories of learners have significant differences in completing homework. Most of the Passers have done a lot of homework, and Abandoners have done very few homework.

Many courses take the completion of the assignment as a prerequisite for the issuance of the course certificate. In addition, synchronous learners are also willing to complete online exercises to strengthen their knowledge.

## Conclusion

We analyzed the characteristics of the certificate learner in watching the video and completing the assignment. Learners are divided into passers, losers, and abandoners based on their test scores. The problem we study is the characteristics of learning behavior of each category.

Analysis of the behavior of watching video shows that there is no significant difference between the Passers and Abandoners. This also shows that the current MOOC video is not attractive, and learners are not willing to use it as a way to learn knowledge.

After analyzing the learning behavior of the completed exercise, it was found that the learners who obtained the certificate almost completed a lot of exercises in the course. Completing the course exercises is clearly an important feature of the learners who get the certificate.

Thus, it can be concluded that MOOC is currently only an aid to classroom instruction. The course video is unattractive and has very low utilization. This video plus practice mode must be improved to make MOOCs easier to stick to.

Of course, the courses we studied were all from the Chinese university MOOC. Our research can only explain the existence of this phenomenon in it. We expect researchers with similar data on other platforms to give answers.

## Acknowledgement

This study was funded by The 13th Five-Year Plan for Education Science in Hunan Province (grant number XJK18CXX010).

This study was funded by the Hunan Province Teaching Reform Research Project (grant number [2018]436).

This study was funded by Open Fund of University Mathematics Teaching Research and Development Center (CMC20190305).

## References

- [1] Kay, J., Reimann, P., Diebold, E., & Kummerfeld, B. (2013). MOOCs: So many learners, so much potential... *IEEE Intelligent systems*, 28(3), 70-77.
- [2] Zhou, M. (2016). Chinese university students' acceptance of MOOCs: A self-determination perspective. *Computers & Education*, 92, 194-203.
- [3] Zhuoxuan, J., Yan, Z., & Xiaoming, L. (2015). Learning behavior analysis and prediction based on MOOC data. *Journal of computer research and development*, 52(3), 614-28.
- [4] Kovacs, G. (2016, April). Effects of in-video quizzes on MOOC lecture viewing. In *Proceedings of the third (2016) ACM conference on Learning@ Scale* (pp. 31-40). ACM.
- [5] Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs* (pp. 60-65).
- [6] Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating How Student's Cognitive Behavior in MOOC Discussion Forums Affect Learning Gains. *International Educational Data Mining Society*.
- [7] Guo, P. J., Kim, J., & Rubin, R. (2014, March). How video production affects student engagement: an empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41-50). ACM.