

# A Study on the Correlation Between Curriculum Achievement and Student Behavior on Network-aided Teaching Platform

Junying Yuan, Shaohui Lin, Sijie Liu and Haishan Chen\*

Nanfeng college of Sun Yat-sen University, Guangzhou, China

\*Corresponding author

**Abstract**—It's well-known that a high-quality network-aided teaching platform can promote the teaching efficiency through resource and information sharing. In this light, the analysis of students' online behaviors on the platform would help assessment curriculum achievements and learning rules, which in turn can assist teaching method adapting and improve the teaching and learning effects. In this paper, a qualitative correlation analysis is made between student achievement data and students' online behavior data in a widely used network-aided teaching platform. Specifically, the curriculum characteristics in the teaching platform and its influence on students' achievements are explored to provide some practical reference information for curriculum teaching reform and for assessment of curriculum performance.

**Keywords**—network-aided teaching platform; curriculum achievement assessment; correlation analysis

## I. INTRODUCTION

With the rapid development and wide application of modern information technologies, the informatization in higher education has entered a high-speed development stage in China since the beginning of the 21st century. As one of the organization methods of the teaching process, the network-aided teaching platform has shown its advantages in openness, interactivity and resource sharing [1]. By providing a variety of functionalities and services such as resource sharing, information sharing and online communication, it brings great convenience to curriculum teaching organization and students' learning. By far, various network-aided teaching platforms have been widely used in higher education teaching activities.

In developed countries such as Europe and the United States, where education informatization has developed much earlier, there have been a series of research results for online teaching or online teaching auxiliary platforms [2]-[5]. Although the modern informatization of higher education in China started late, with the wide application of network teaching platform, there have been many researches [1][6]-[8] on the application methods, methods and teaching effects of network teaching platform in the field of education. Relevant researches have promoted the further maturity and wide application of network-aided teaching platforms.

In recent years, with the development and wide application of big data technology, there have been quite a few researches on the correlation between learning behavior and learning effect with network-aided teaching platforms [6][9]-[11]. Some literatures [6, 9, 10] take MOOC network teaching platforms as the research object to study learning behavior and effect under the condition of online learning. Literature [11] predicts the learning effect of 70 students through their usual scores and video learning rate on the online teaching auxiliary platform, but there are some problems such as small amount of data, single course type and insufficient analysis of online learning behavior. However, there is little research on students' learning efficiency with network-aided teaching platforms.

In this paper, we first collect curriculum data of Nanfeng College of Sun Yat-sen University in Scholat [12], which is one of the most commonly used network-aided teaching platform in China, then observe students' various online behavior statistics in the teaching platform, and then explore the correlation between the behavior rules of students in the course and the evaluation of the course results so as to provide assessment information for curriculum teaching and learning.

## II. DATA SET

### A. Data Collection and Cleaning

By far, Scholat is the largest academic information exchange platform in the field of higher education in China. The platform brings together teachers and users from many universities, and provides them with network-aided teaching functions, supporting the sharing of teaching resources, attendance, online communication, assignment, assignment submission and evaluation, etc. In this platform, the metadata related to a curriculum includes: the number of login, the number of web clicks, the number of resource downloads, and the status of homework submission, etc.

To make an analysis on the correlation between the data of the network-aided teaching platform and the data of curriculum achievements (CAs), we further collected students' CA data of relevant curriculums, including ordinary time achievement, final test score and curriculum GPA (grade point average). Among them, GPA is the weighted average of ordinary time achievement and final test score.

Since some of the collected data are not complete and thus not suitable for direct data analysis, data pre-processing is performed before data analysis by filtering away the incomplete data, including:

- Curriculums with too few activity data, i.e. the student number is less than 20, and the curriculums with too few web clicks (less than 40).
- Curriculums with no corresponding achievement data.

### B. The Employed Data Set

After data pre-processing, the remaining data are employed the data set for analysis. The employed data set consists both the students' behavior data from Scholat and the data of curriculum achievement. The metadata of students' behavior for each student in each curriculum includes the number of login, the number of web clicks, the number of resource downloads, and the number of homework submission, while the metadata of student's CAs consist of ordinary time achievement, final test score and curriculum GPA.

The metadata in the data set are all generated between Feb. 2015 and Sept. 2018 (7 semesters), from 86 curriculums and 5596 students. In all, there are 323200 logins, 1078800 web clicks, 123054 homework submissions, 440231 resource downloads and 18865 CA records.

## III. STATISTICS OF ONLINE BEHAVIOR DATA

This section digs into the characteristics of students' online behavior. Specifically, the basic online behavior statistics are illustrated, and a normalized measure is designed to describe the curriculum vitality in the teaching platform is defined.

For simplicity, the following notations are used,  $1 \leq c \leq N_c$  denotes a curriculum,  $N_c = 86$  is the number of all curriculums,  $n_s(c)$ ,  $n_{lg}(c)$ ,  $n_c(c)$ ,  $n_{rd}(c)$  and  $n_{hw}(c)$  represent, for a curriculum  $c$ , the numbers of students, logins, web clicks, resource downloads and homework submissions, respectively.

### A. Behavior Statistics

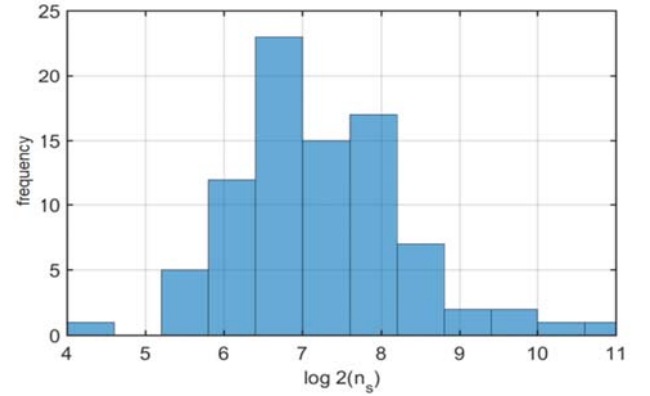
Figure 1 provides an illustration of the online behavior statistics. In Figure 1(a), the histogram of curriculum student numbers in exponential form, from which it's not difficult to see that the number of students in the exponential form of the course follows approximately Poisson distribution.

Figure 1(b) further present the relationship of  $n_l(c)$ ,  $n_c(c)$ ,  $n_{rd}(c)$  and  $n_{hw}(c)$  against  $n_s(c)$  in exponential form by arranging  $\log_{10}(n_s(c))$  in an ascending order. Generally speaking, as the number of students increases, the number of activities in the curriculum accordingly increases. However, since the data of some curriculums are quite different, the activity data fluctuates as the increment of student number in the curriculum. The reason is that the more the functions of the teaching platform are used, the more the activity data of students are. However, different teachers may have different ways in using the network-assisted teaching platform, i.e. some

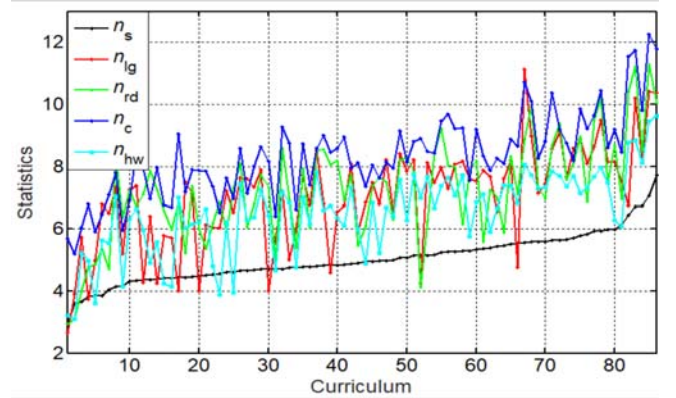
may only focus on online homework submission, some focus on resource sharing, and some courses use most functions.

### B. Designing a Curriculum Vitality Indicator

With the behavior statistics in Figure 1(b), we can conclude that different curriculums have different characteristics. Since different teachers have different emphases in using the teaching platform, the students' online activity is also different, and thus the curriculum vitality varies.



(a) Histogram of curriculum student numbers



(b) Behavior statistics in exponential form

FIGURE 1. BEHAVIOR STATISTICS.

In generally, whether a curriculum is active or not can be described by online behavior statistics. Among the five kinds of student's online behavior metadata, the number of students in a curriculum has no direct influence on the vitality of the curriculum, while, the remaining four types of data, including the number of logins, the number of web clicks, the number of resource downloads and the number of homework submission, reflect the vitality of the course to some extent. In this light, the concept of curriculum vitality is put forward according to the five kinds of statistical data of the course.

Let  $\psi$  denotes the curriculum vitality, the value of  $\psi$  should be normalized for convenient comparison, therefore  $\psi \in [0,1]$ . In the design of  $\psi$ , the influence of the number of students in a curriculum should be minimized. Therefore, curriculum vitality is defined with equation (1)

$$\psi = f(\bar{n}_{lg}, \bar{n}_c, \bar{n}_{rd}, \bar{n}_{hw}). \quad (1)$$

where  $\bar{n}_{lg} \in [0,1]$ ,  $\bar{n}_c \in [0,1]$ ,  $\bar{n}_{rd} \in [0,1]$  and  $\bar{n}_{hw} \in [0,1]$  are the normalized impact factors of  $n_{lg}$ ,  $n_c$ ,  $n_{rd}$  and  $n_{hw}$ , respectively.

Since  $\bar{n}_{lg}$ ,  $\bar{n}_c$ ,  $\bar{n}_{rd}$  and  $\bar{n}_{hw}$  are the impact factors of curriculum vitality  $\psi$ , there normalization must be done in consideration of eliminating the impacts from curriculum student number, the duration measured in number of weeks. Therefore, the normalization of  $n_{lg}$ ,  $n_c$ ,  $n_{rd}$  and  $n_{hw}$  are performed with equation (2), where  $N_{wk}=18$  is the number of weeks in a semester,  $N_r$  is the number of resources in a curriculum,  $T_{lg}=2$ ,  $T_c=5$ ,  $T_{rd}=2$ ,  $T_{hw}=1$  are normalization factors with heuristically given values for  $n_{lg}$ ,  $n_c$ ,  $n_{rd}$  and respectively. The values of normalized factors generally fall in the range of  $[0,1]$ , however, there may exist cases when the normalization factors are not big enough, therefore their maximum values are restricted to 1.

$$\begin{aligned} \bar{n}_{lg} &= n_{lg} / n_s / N_{wk} / T_{lg}, \\ \bar{n}_c &= n_c / n_s / N_{wk} / T_c, \\ \bar{n}_{rd} &= n_{rd} / n_s / N_r / T_{rd}, \\ \bar{n}_{hw} &= n_{hw} / n_s / N_{wk} / T_{hw}. \end{aligned} \quad (2)$$

Figure 2 shows the change of normalized impact factors along with the increase of student number. Note that there exists some correlation between normalized login times, click times, resource downloads and homework submissions, but the correlation is no obvious.

Based on the observations from figure 2, the curriculum vitality in equation (1) can be simplified by adopting the weighted sum method, as specified by the equation

$$\psi = \omega_{lg} \bar{n}_{lg} + \omega_c \bar{n}_c + \omega_{rd} \bar{n}_{rd} + \omega_{hw} \bar{n}_{hw}, \quad (3)$$

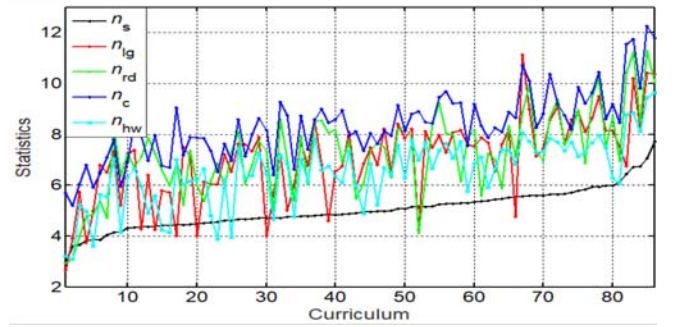
where  $\omega_{lg} + \omega_c + \omega_{rd} + \omega_{hw} = 1$ . Even though the value for each weight can be solved by optimization methods, they are simply set to  $\omega_{lg} = \omega_c = \omega_{rd} = \omega_{hw} = 0.25$  so as to improve the generality and simplicity of the definition of curriculum vitality.

Fig. 3(a) shows the distribution of vitality  $\psi$  along with the increase of student numbers in the curriculum. From the smoothed version of the curve, it can be seen that curriculum vitality generally increases with the increment of student numbers, but the growth is quite limited. Therefore, the curriculum vitality defined by equation (3) is less affected by the student number in a curriculum, which means it can better reflect the average vitality of students and can fairly measure the actual vitality of different curriculums. Fig. 3(b) low-level vitality curriculums. Therefore, the measure of curriculum

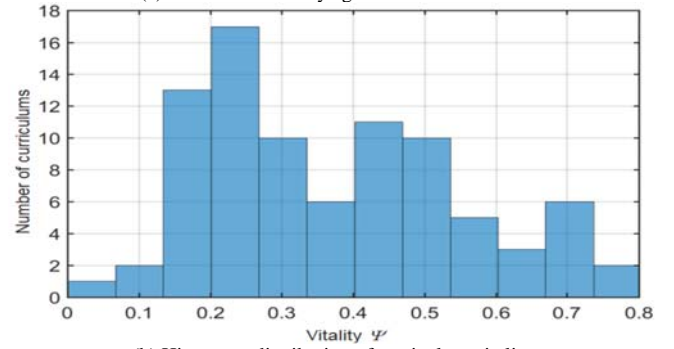
vitality proposed in this paper can be used to evaluate the use of the curriculum in the network-aided teaching platform.

#### IV. THE CORRELATION BETWEEN ONLINE BEHAVIORS AND CURRICULUM ACHIEVEMENTS

In this section, the measure of curriculum vitality is used to explore the correlation between online activities and the assessment of CAs. The CA in this paper are all evaluated in a 100-point system. In order to facilitate analysis, this paper first uses the average course score as the analysis object, and then normalizes the average course score (i.e. divided by 100 out of 100) to obtain the normalized average course score.



(a) Curriculum vitality against student number



(b) Histogram distribution of curriculum vitality

FIGURE II. DISTRIBUTIONS OF CURRICULUM VITALITY.

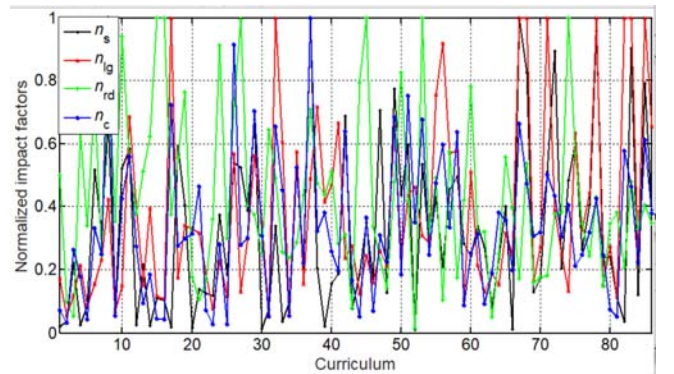


FIGURE III. DISTRIBUTIONS OF NORMALIZED IMPACT FACTORS ORDERED BY NS.

In order to simplify the description, we use  $g_n$ ,  $g_c$  and  $g$  to express the normalized ordinary time achievement, final test

score and curriculum GPA, respectively. Generally speaking, CA measures  $g_o$ ,  $g_e$  and  $g$  generally falls in the scope of  $[0.5, 1]$ , while  $\bar{n}_{lg}$ ,  $\bar{n}_c$ ,  $\bar{n}_{rd}$ ,  $\bar{n}_{hw}$  and  $\psi$  take values within  $[0, 1]$ . To better illustrate the correlations, in figures, between CA and  $\psi$ ,  $\bar{n}_{lg}$ ,  $\bar{n}_c$ ,  $\bar{n}_{rd}$ ,  $\bar{n}_{hw}$  and  $\psi$  are projected to the same of CA measures. The projected values of  $\bar{n}_{lg}$ ,  $\bar{n}_c$ ,  $\bar{n}_{rd}$ ,  $\bar{n}_{hw}$  and  $\psi$  are denoted by  $\bar{n}'_{lg}$ ,  $\bar{n}'_c$ ,  $\bar{n}'_{rd}$ ,  $\bar{n}'_{hw}$  and  $\psi'$  respectively.

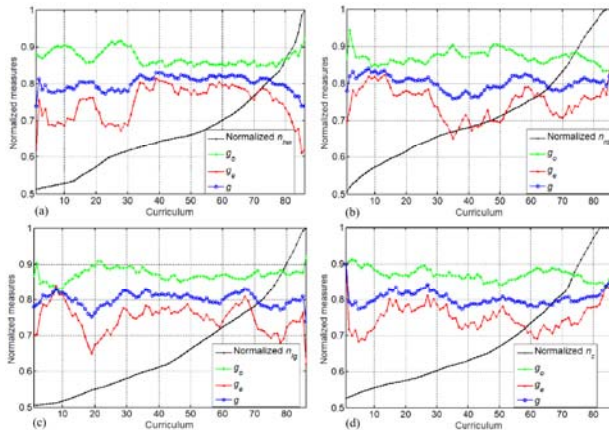


FIGURE IV. CORRELATION BETWEEN ONLINE BEHAVIORS AND CA MEASURES.

Figures 4(a)-(d) illustrate the trends of CA measures  $g_o$ ,  $g_e$  and  $g$  along the increment of  $\bar{n}'_{lg}$ ,  $\bar{n}'_c$ ,  $\bar{n}'_{rd}$ ,  $\bar{n}'_{hw}$ , respectively. Among them, the relationship between the number of homework submissions and the course scores is obvious (see figure 4(a)). When  $\bar{n}'_{hw} < 0.63$ , the ordinary time achievement  $g_o$  is usually higher but the examination score  $g_e$  is relatively lower. When  $0.63 < \bar{n}'_{hw} < 0.8$ , the gap between  $g_o$  and the examination score  $g_e$  is smaller. When  $\bar{n}'_{hw} > 0.8$ , the ordinary time achievement  $g_o$  is improved but the examination score  $g_e$  and the GPA  $g$  decay significantly.

There exists some kind of correlation between the number of resource downloads of and the CAs. As shown in figure 4(b), when the number of resource downloads  $\bar{n}'_{rd}$  overpasses a certain value (0.67), the examination score  $g_e$  presents an increasing trend with the increment of  $\bar{n}'_{rd}$ . However, the relationship is not obvious for the number of logins and web clicks against the CA results, as shown in figure 4(c)(d). In summary, as has been illustrated in figure 4, the correlation between the single online behaviors and the assessment of curriculum achievements is not obvious.

Figure 5 presents the correlation between the vitality  $\psi'$  and the CA measures  $g_o$ ,  $g_e$  and  $g$ . Compared with that in figure 4, these relationships are obviously clearer. When  $\psi' < 0.73$  along with the increment of  $\psi'$ , the score  $g_e$  and GPA  $g$  improve gradually, while the ordinary time achievement  $g_o$  decays, thus the gap between  $g_e$  and  $g_o$  narrows steadily.

However, when the curriculum vitality  $\psi' > 0.73$ , the CA results shows an opposite trend: the examination score  $g_e$  and the GPA  $g$  decrease gradually, while  $g_o$  shows an increasing trend, and thus the gap between  $g_e$  and  $g_o$  widens rapidly. The correlations presented by figure 5 indicates that the designed curriculum vitality  $\psi$  (see equation (3)) can reflect the characteristics of curriculum achievement assessment.

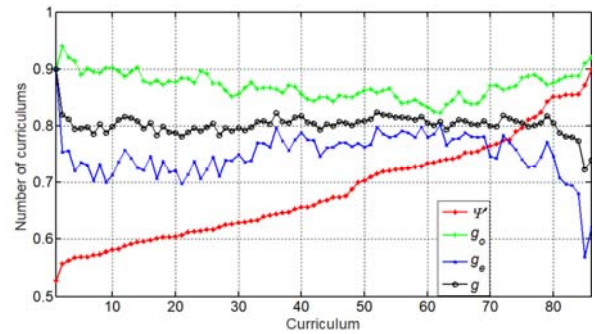


FIGURE V. CORRELATION BETWEEN VITALITY  $\psi$  AND CURRICULUM ACHIEVEMENTS.

#### ACKNOWLEDGMENT

This work is supported in part by the Guangzhou Science and Technology Plan Project of China under grant 201904010276 and 201804010292, and in part by the General Programs of Humanities and Social Sciences Research of Guangdong Provincial Education Department, China.

#### REFERENCES

- [1] G. Zhao, X. Wang, C. Che, L. Hu, and Q. Wang, "The auxiliary function of network platform in undergraduate ophthalmology teaching," *China Medical Education Technology*, vol. 28 no. 04, pp. 402-404, 2014. (In Chinese)
- [2] L. P. MACFAYDEN and S. DAWSON, "Mining LMS data to develop an EarlyWarning system for educators: a proof of concept," *Computers & Education*, vol. 54, no. 2, pp. 588-599, 2010.
- [3] L. Ivan, II Harre and L. Bower Beverly, "Student characteristics that predict persistence in community college online courses," *American Journal of Distance Education*, vol. 25, no. 3, pp. 178-191, 2011.
- [4] A. Ramesh, D. Goldwasser, B. Huang, H. Daumé, L. Getoo, "Modeling learner engagement in MOOCs using probabilistic soft logic," *NIPS Workshop on Data Driven Education*, 2013[2014-06-0]. <http://lytics.stanford.edu/datadriveneducation/papers/>
- [5] G. Balakrishnan, "Predicting student retention in massive open online courses using hidden markov models," *UCB/EECS-2013-109[R/OL]*. Berkeley: University of California, Berkeley, 2013.
- [6] Y. Zong, H. Sun, H. Zhang, Q. Zheng, and L. Chen, "A logical regression analysis of learning behavior and learning effect in MOOCs," *Distance Education In China*, vol. 05, pp. 14-22, 2016. (In Chinese)
- [7] Z. Zhu, "Research and practice of assistant teaching on college physics network teaching platform," *Contemporary Education Research and Teaching Practice*, vol. 02, pp. 30-31, 2018. (In Chinese)
- [8] W. Ma, "Research on the Application of Network Teaching Platform to Assist Classroom Teaching," *Digital Communication World*, vol. 08, pp. 246-259, 2018. (In Chinese)
- [9] L. Jiang, X. Han, and J. Cheng, "An analysis of MOOCs learners' characteristics and learning effects," *China Educational Technology*, vol. 11, pp. 54-59+65, 2013. (In Chinese)

- [10] J. Jia, J. Miu, and Q. Wang, "Big Data Analysis of MOOC Learning Behavior and Effect: A Case Study of 6 MOOC Classes in Peking University," *Industry and Information Technology Education*, vol. 09, pp. 23-29, 2014. (In Chinese)
- [11] Y. Bai, "Research on the forecast of curriculum achievement assisted by network teaching platform," *Jiangsu Salt Science & Technology*, vol. 45, no. 02, pp. 51-52, 2018. (In Chinese)
- [12] Y. Tang, Scholat, <http://www.scholat.com/>