Measuring Student's Utilization of Video Resources and its Effect on Academic Performance

Huan He[1, 2], Qinghua Zheng[1, 2], Bo Dong[3, 4], and Hongchao Yu[1, 2]

[1] Shaanxi Province Key Lab of STN Tech. R&D, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

[2] School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

[3] National Engineering Lab for Big Data Analytics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

[4] School of Continuing Education, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

{hehuan, qhzheng, dong.bo}@mail.xjtu.edu.cn, yuhongchao@stu.xjtu.edu.cn

***Abstract*—Massive video resources were produced to meet the needs of learning knowledge and skills anytime and anywhere through internet. Therefore, whether these video resources were fully utilized by students is an important issue for schools and teachers. This paper proposes three indicators based on student's log data and course's video information to measure the utilization of video resources. In addition, the proposed indicators are applied in a case study to analyze how different utilization patterns affect students' academic performance in a large-scale online distance education context.**

Keywords-video utilization; evaluation indicator; academic performance

I. INTRODUCTION

With the rapid development of internet technology, online distance education (ODE) has been widely used to provide high-quality educational resources to students throughout the country [1]. Since the expansion of enrollment and the increase of the number of specialties, massive video resources were produced to meet the needs, which requires a significant investment of capital and time of teachers, teaching assistants and various types of professionals [2]. Therefore, it's an important issue for ODE schools and teachers to understand how these video resources are utilized by students and how the difference in the utilization affects students' academic performance, which may provide useful information to further adjust curricula and improve instructional design.

In order to measure how student interacts with the learning resources, previous studies have proposed lots of indicators. Based on log data of learning behaviors recorded and accumulated by learning management systems (LMSs), these indicators were used to discover patterns of learning behaviors [3, 4, 5], to explore the relationship between learning behaviors and academic performance [6, 7], and to predict learning achievement [8, 9]. Since these indicators were mainly applied in the analysis of learning behaviors in single course or a small number of courses, they may not be adapted to the analysis and comparison of the large-scale utilization of learning resources in ODE context.

As an integral part of higher education, ODE sets up curriculum of each specialty with reference to disciplines of full-time college. Take a representative ODE school in China as an example, there are approximately 25 courses including 1,300 videos with a total duration of about 750 hours for each specialty. Students are required to complete the full course load within two to five years and pass the course exams to get the degree. Since the courseware video is the typical teaching tool and the courses are diversified in disciplines, content and length, it's necessary to use unified indicators to understand how these courseware videos of various courses and specialties are used and how the utilization is related to student's academic performance.

The major contributions of this paper are summarized below. First, we propose the attendance rate (AR), utilization rate (UR) and watch ratio (WR) as the general indicators to measure student's utilization of courseware videos in multi-specialty and multi-course context. Secondly, we apply the proposed indicators to a real dataset from a representative ODE school in China to investigate the utilization of video resources and its effects on students' academic performance.

The rest of this paper is organized as follow. Section II proposes the indicators. Section III presents the analysis of utilization patterns. Section IV concludes.

II. METHODS

A. Sample and data collection

The dataset used for the case study was collected from the ODE platform of our university, including curricula data from teaching management system (TMS) and learning process data from LMS. The curricula data includes total number and duration of courseware videos of each course, and the learning process data includes students' academic status, course list and log file which contains student' watching behaviors.

After preprocessing, the dataset consists of 8,276 distinct students enrolled in 2014 and their learning logs for the period from 2014 to 2017 which covers all activities from enrollment (Mar 2014) to graduation on schedule (July 2016) and postponed graduations (Jan 2017, July 2017).

B. Measurements

We propose the attendance rate (AR), utilization rate (UR) and watch ratio (WR) as the indicators to measure student's utilization of courseware videos with combination of the number and duration of video.

Student's course AR:

$$ar_{s,c} = \frac{|W_{s,c}|}{|V_c|}$$

Table 1 Descriptive statistics of AR, UR and WR

Indicator	Median	Mean	SD
AR	0.03	0.07	0.10
UR	0.02	0.08	0.19
WR	0.69	1.00	2.90

Here, $ar_{s,c}$ is the attendance rate of student s in course c; V_c is the collection of courseware videos of course c, and $W_{s,c}$ is the collection of distinct videos watched by student s in course c, which means multiple watching of the same video is regarded only once. For example, if a course has a total of 10 videos and a student watches one of the videos 5 times, then the student's course AR is 1/10 = 0.1, instead of 5/10 = 0.5.

Student's course UR:

$$ur_{s,c} = \frac{\sum_{wt \in WT_{s,c}} wt}{\sum_{vt \in VT_c} vt}$$

Here, $ur_{s,c}$ is the utilization rate of student s in course c; VT_c is the collection of courseware video time vt of course c, and $WT_{s,c}$ is the collection of the each duration of watching videos by student s in course c. Unlike AR, the duration of repeatedly watching the same video is cumulative. For example, if a course has a total duration of 10 hours and a student repeatedly watches one video 5 times for 0.5 hours each time, then the student's course UR is 0.5 x 5 / 10 = 0.25, instead of 0.5/10 = 0.05.

Based on the above $ar_{s,c}$ and $ur_{s,c}$ for evaluating student's utilization in one course, we can further measure the overall specialty AR ar_s and UR ur_s as follows:

$$ar_s = \frac{1}{|SC_s|} \sum_{c \in SC_s} ar_{s,c}$$

$$ur_s = \frac{1}{|SC_s|} \sum_{c \in SC_s} ur_{s,c}$$

The SC_s is the course list of student s, which is designated by his/her specialty's training program. With ar_s and ur_s, the student's overall specialty WR is defined as follow:

$$wr_s = \frac{ur_s}{ar_s}$$

The wr_s represents how student watches the video since he/she opened it. For instance, wr_s=1 means that the student s completely watches the video since he/ she opened it. Because ar_s may be zero, the wr_s only applies to students with ar_s>0. The follow-up discussions in this study are all regarding student's overall specialty AR, UR and WR.

III. Results and Discussion

A. Statistical Analysis on AR, UR and WR

The descriptive statistics of the proposed indicators are listed in Table 1. The result of WR indicates that students watched most of its content since they opened a video, and some of them may watch same video multiply times. However, the performance of AR and UR are lower than expected as the WR. This result indicates that the videos of each course were not fully utilized by students.

To further investigate the utilization of video resources, the cumulative distribution of AR and UR were plotted. As shown in Figure 1, compared with the expected distribution of

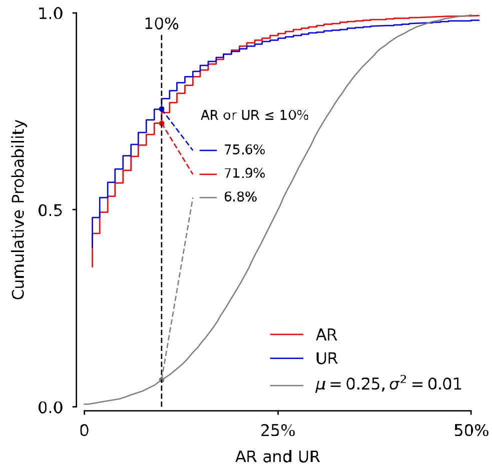

Figure 1 Cumulative distribution of AR and UR

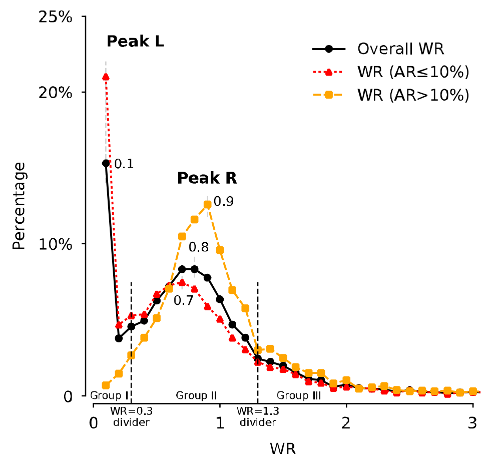

Figure 2 Distribution of WR

video utilization (the gray line, AR and UR ~ N (μ=0.25, σ=0.1)), the AR and UR of most students didn't exceed 50%, and more than 70% of students only watched less than 10% of all videos. The low video utilization is not only a waste of investment, but also may imply that there is a mismatch between current resources and the needs of students.

B. Patterns of utilization

As shown in Figure 2, there are two distinct peaks in the distribution of overall WR (black solid line with round dot), peak L (WR=0.1) and peak R (WR=0.8). Considering the AR difference shown in Figure 1, we further divide students into two series, series AR≤10% and series AR>10%. The WR distribution of series AR≤10% (red dotted line with triangle dot) is similar to the overall WR, except for higher proportion at peak L and the smaller value of peak R (WR=0.7). The WR distribution of series AR>10% (yellow dashed line with square dot) is obviously different from other series, which has only one peak R (WR=0.9).

This result indicates that there are different patterns in watching behaviors. The lower WR near peak L reflects that these students only watched for a short time after opening the

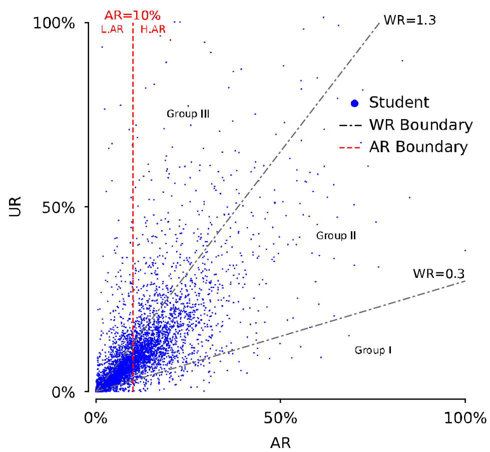

Figure 3 Distribution of students in AR and WR groups

courseware page, while the higher WR near peak R reflects those students watched most content of video. Meanwhile, these patterns are also related to the AR performance, since there is no peak L in series AR>10% and the peak R in series AR>10% is higher than that in series AR≤10%. These results may imply that different watching styles have mutual effect on utilization of video resources.

C. Effects of the utilization on academic performance

To further explore how the difference in AR and WR affect academic performance, we divided the students into three groups, WR≤0.3 (Group I), 0.3<WR≤1.3 (Group II) and WR>1.3 (Group III), respectively. The selected dividers were based on the distribution of WR shown in Figure 2. The lower point between peak L and peak R (WR=0.3) and the opposite point on the right side of peak R (WR=1.3) were selected. In addition, students were further divided by AR=10%. We used a scatter chat to visualize the distribution of students in each group. As shown in Figure 3, each blue dot represents a student with abscissa AR and ordinate UR and three lines divided students into different groups.

For ODE students, the primary criterion of academic performance is whether to graduate on schedule. Therefore, we used the academic status as the measurement which includes three situations, namely, Studying, Graduated (on schedule) and Postponed (graduation). As listed in Table 2, students with higher WR have relative higher rate of graduation on schedule (Group III 0.68 > Group II 0.65 > Group I 0.52). In addition, the result listed in Table 3 may imply that the AR have a greater impact on student's academic performance, since student with higher AR had a higher rate of graduation on schedule among groups or in same group.

IV. CONCLUSION

In this paper, we proposed using AR, UR and WR as indicators to analyze large-scale utilization of video resources and investigated the effects on students' academic performance. The results show that these indicators can reflect utilization patterns of watching video resources. In addition, the results also suggest that ODE school should pay more attention to the utilization of video resources, which may affect whether a student will graduate on schedule or not.

Table 2 Mutual distribution of academic status and WR groups

		Academic status		
WR Group	**Ratio**	***Studying***	***Graduated***	***Postponed***
Group I	P(A|G)	0.24	0.52	0.23
	P(G|A)	0.29	0.20	0.31
Group II	P(A|G)	0.18	0.65	0.17
	P(G|A)	0.56	0.63	0.56
Group III	P(A|G)	0.17	0.68	0.15
	P(G|A)	0.14	0.18	0.13

P(A|G) denotes the ratio of specified status in this group.
P(G|A) denotes the ratio of this group in specified status.

Table 3 Mutual distribution of academic status and AR+WR groups

		Academic status		
AR+WR Group	**Ratio**	***Studying***	***Graduated***	***Postponed***
L.AR + Group I	P(A|G)	0.25	0.51	0.24
	P(G|A)	0.29	0.18	0.29
L.AR + Group II	P(A|G)	0.22	0.59	0.19
	P(G|A)	0.43	0.37	0.41
L.AR + Group III	P(A|G)	0.19	0.64	0.16
	P(G|A)	0.11	0.11	0.10
H.AR + Group I	P(A|G)	0.08	0.73	0.19
	P(G|A)	0.01	0.02	0.01
H.AR + Group II	P(A|G)	0.12	0.75	0.13
	P(G|A)	0.13	0.26	0.16
H.AR + Group III	P(A|G)	0.13	0.76	0.11
	P(G|A)	0.03	0.07	0.03

L.AR represents AR⩽10%; H.AR f represents AR>10%

ACKNOWLEDGMENT

This research was partially supported by "The Fundamental Theory and Applications of Big Data with Knowledge Engineering" under the National Key Research and Development Program of China with Grant No. 2016YFB1000903, the MOE Innovation Research Team No. IRT17R86, the National Science Foundation of China under Grant Nos. 61721002, 61502379, 61532015, and Project of China Knowledge Centre for Engineering Science and Technology.

REFERENCES

[1] X. Ding, J. Niu, and Y. Han, "Research on distance education development in China," British Journal of Educational Technology, vol. 41, no. 4, pp. 582–592, 2010.

[2] S. Ros, R. Hernandez, T. Read, M. Rodriguez Artacho, R. Pastor, and G. Diaz Orueta, "UNED OER Experience: From OCW to Open UNED," IEEE Transactions on Education, vol. 57, no. 4, pp. 248–254, 2014.

[3] T. Kahan, T. Soffer, and R. Nachmias, "Types of Participant Behavior in a Massive Open Online Course," The International Review of Research in Open and Distributed Learning, vol. 18, no. 6, Sep. 2017.

[4] L.-Y. Li and C.-C. Tsai, "Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance," Computers & Education, vol. 114, pp. 286–297, Nov. 2017.

[5] F. Van der Sluis, J. Ginn, and T. Van der Zee, "Explaining Student Behavior at Scale: The Influence of Video Complexity on Student Dwelling Time," in Proceedings of the Third (2016) ACM Conference on Learning @ Scale, New York, NY, USA, 2016, pp. 51–60.

[6] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," Computers & Education, vol. 113, no. Supplement C, pp. 177–194, Oct. 2017.

[7] S. Joksimović, D. Gašević, T. M. Loughin, V. Kovanović, and M. Hatala, "Learning at distance: Effects of interaction traces on academic achievement," Computers & Education, vol. 87, pp. 204–217, Sep. 2015.

[8] N. A. Diep, C. Cocquyt, C. Zhu, and T. Vanwing, "Predicting adult learners' online participation: Effects of altruism, performance expectancy, and social capital," Computers & Education, vol. 101, pp. 84–101, 2016.

[9] J. W. You, "Identifying significant indicators using LMS data to predict course achievement in online learning," The Internet and Higher Education, vol. 29, pp. 23–30, Apr. 2016.