Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2019.DOI

How Learner Support Services Affect Student Engagement in Online Learning Environments

HUAN HE[1,2], QINGHUA ZHENG[1,2], DEHAI DI[3], and BO DONG[4,5]

[1]Shaanxi Province Key Laboratory of Satellite and Terrestrial Network 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]School of Management, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[4]School of Continuing Education, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[5]National Engineering Lab for Big Data Analytics Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

Corresponding author: Bo Dong (e-mail: dong.bo@mail.xjtu.edu.cn).

ABSTRACT Learning is challenging, especially in the self-paced online learning environment. Not all students start online learning with the skills to manage their learning plans, balance work-study, and find learning resources. Therefore, as an online educational institution, providing learner support services is essential to improving learning outcomes and student satisfaction. Existing studies have analyzed the impact of learner support services on student satisfaction, learning outcomes, and course retention. However, the relationship between these services and student engagement in the long-term learning process has not been fully examined. To address this issue, this paper investigates students' usage of two online learner support (OLS) services and its impact on student engagement. We examined four student groups categorized by their usage of two OLS services and analyzed student engagement with an online course using learning log data and visual analytics techniques. The findings indicate that the use of services and how many times the services are used are highly relevant to the pattern and level of student engagement, which suggests that the use of OLS services can be added to the indicators that reflect student learning status.

INDEX TERMS Learner support, online distance education, student engagement, visual analysis

I. INTRODUCTION

THE popularity of online education gives learners the opportunity to access global quality educational resources to improve their knowledge and skills, which also requires learners to face new challenges that are different from on-campus learning [6], [12], [40]. Due to the lack of face-to-face communication with course instructors, online learners are expected to plan their learning program, set the schedule for study, and balance their studies and daily work alone. Accordingly, learning at a distance requires a high level of motivation, multitasking coordination, and the ability to study independently [13]. Obviously, not all students have these skills when they start learning online. Therefore, online learning institutions have the responsibility to provide learner support services and corresponding resources to help students develop their skills and adapt to online learning and lay the foundation for learning success [4].

The importance of learner support services on students' satisfaction, motivation, engagement, retention, and success has been emphasized in specialized literature [4], [6], [12], [40]. As defined in [13], learner support services include the library, advising and counseling, academic skill assessment and development, community development, peer-to-peer support, and administrative services, which support the learning process but do not include direct subject teaching. Ideally, learner support services can be provided by instructors or teaching assistants to achieve better satisfaction and effectiveness. Existing studies have found that timely feedback and response from instructors were rated the highest support strategies, and novice online students may require more support and guidance from instructors [14], [15].

However, in large-scale online learning, the number of instructors and their available time for supporting learners is limited. Therefore, a variety of online learner support (OLS) services has been introduced to help students solve various problems from different aspects, such as a guidance section in online forums [21], instant messaging [35], chatbots [26], visual aids [19], and recommended learning strategies [9]. Through the learning guidance and course assistance provided by these support services, online learners may have a better learning experience and learning outcomes to achieve their learning goals. To the best of our knowledge, previous studies on learner support mainly focus on learner characteristics, satisfaction surveys, assessment of needs, and guidance for learner support practice, while the impact of the usage of learner support services on the student engagement of long-term learning process has not been examined. To fill this gap, we use visual analytics methods to explore the log data collected from the learning management system (LMS) of our online distance education (ODE) school to analyze the impact of two learning support services, namely, assistance messenger (AM) and weekly guidance (WG), on student engagement throughout a semester.

To better understand how these two services affect student engagement, we address the following research questions:

- **RQ1**: How frequently do students use OLS services in a semester?
- **RQ2**: Do students who use the OLS services show a higher level of course engagement?
- **RQ3**: Are there different patterns of course engagement among students regarding the usage of OLS services?

Section 2 presents a discussion of the state-of-the-art in OLS and student engagement, and a visual analysis of online learning. Section 3 further identifies the characteristics of the two OLS services provided in the LMS of the ODE school and the dataset used in this study. Section 4 describes the measurements for student engagement and data analysis procedure. Section 5 presents the findings to answer the research questions presented above. Section 6 discusses the results and findings. Section 7 summarizes our research and describes future work.

II. RELATED WORK

In this section, we first summarize recent works on OLS services. Then, we review the literature that is related to student engagement in the online learning environment. Finally, we discuss the use of visual analytics to mine student behavior patterns and discuss how our work extends existing research.

A. ONLINE LEARNER SUPPORT

To help online learners persist in their studies and meet their learning goals, the research related to learner support mainly focuses on the following three categories: theoretical insights, evaluation of learner support services, and guidance for learner support practice [13]. Since our study focuses on the impact of OLS services on student engagement, the following paragraphs mainly summarize the recent research on the evaluation of learner support services.

The research on the effectiveness of learner support services in an online learning environment confirms the role of various learner support services in promoting student performance and satisfaction. Many studies have analyzed the impact of support services on learning outcomes in terms of course content [17], [18], online forums [20], [21], chatbots [3], [25], [26], and learning strategies [9]–[11]. For example, Thistoll and Yates [23] suggested that helpful tutors and clear learning materials are essential factors for improving student engagement and course completion rates during the learning process in ODE. Gregori et al. [6] investigated learner support strategies that enable the success and completion of MOOCs and suggested that course designers should focus on their students during the second quartile of the course. Briton et al. [27] studied the discussion behaviors in online forums and found that the teaching staff's participation is positively correlated with a higher volume of discussion. Joksimović et al. [16] found that the time spent communicating with instructors has a significant, although negative, effect on the students' grades in core courses, which suggests a need for increased instructional support by those students who have difficulty with course materials. Martin et al. [14] suggested that instructor support and guidance are critical for students to gain the maximum benefits from student reflections.

In addition to learning outcomes, course retention and attrition have also been examined. Due to the self-paced, asynchronous properties of online learning, students must balance their study with other responsibilities to avoid dropout. In response, institutions should take an approach to support them. One common approach to improving retention is to use post hoc surveys to identify reasons for providing early anticipatory guidance to new students [8]. For example, Willging and Johnson [7] developed an online survey to investigate the reasons for leaving an online program. They identified a lack of interaction with the instructor and not enough support from technical staff as the program-related and technology-related reasons of dropping out of an online program .

Another approach to improving retention is to develop predictive models that consider learner demographics [7], course and institutional variables [16], and learning behaviors [6], [24] to forecast retention or attrition with the aid of statistical models and machine learning techniques. With the support of these models, instructors and institutions can better identify the factors that affect student dropouts and provide corresponding learner support in a timely manner. For instance, Gregori et al. [6] used an extreme learning machine based on a neural network to predict course completion to help instructors identify the determinants of course completion. Gardner and Brooks [28] proposed using the Friedman and Nemenyi two-stage procedure to evaluate the performance of features and models to choose an effective model for predicting dropout in the full population of learners in a course.

These studies provide valuable solutions and examples for building effective learner support services in the ODE school. However, due to differences in course content, students, and learning processes, existing support services on other platforms must be modified for adapting to our environment, and the effectiveness of the adjusted services still needs to be further verified.

B. STUDENT ENGAGEMENT IN AN ONLINE LEARNING ENVIRONMENT

Previous studies have typically identified student engagement as a multidimensional construct, such as behavioral engagement, emotional engagement, and cognitive engagement [37]. In this paper, we focus on behavioral engagement in the LMS. To measure student engagement in online learning environments, many indicators have been proposed from various perspectives.

The most commonly used indicators for measuring student engagement are based on student interactions with functions and resources in the LMS, including the number of accesses to resources [38], the time spent on tasks or resources [24], and the days spent on learning [39]. For instance, Guo et al. [17] used the time that a student spends on a video and whether a student attempts the follow-up problem after watching a video as the proxy for engagement. Van der Sluis et al. [30] used the time students spent viewing videos and how much of the video they actually viewed to measure students' interaction with course videos. Bote-Lorenzo and Gómez-Sánchez [31] defined 16 indicators to measure student engagement in each chapter of an online course, such as the percentage of lecture videos totally or partially watched, the percentage of finger exercises answered, and the percentage of assignments submitted. Li and Tsai [24] analyzed 14 indicators related to time spent on educational resources and found that the students' engagement with course videos shows great variety and can be clustered into three patterns. Singh et al. [29] proposed a content engagement score to measure the engagement experienced by the students towards a specific content, which consists of cognitive, emotional, and behavioral engagements using a comprehensive set of user activities.

These studies have provided a variety of indicators for measuring student engagement. In this study, we plan to extend the indicators related to viewing course videos to describe video utilization as part of student engagement.

C. VISUAL ANALYSIS IN ONLINE LEARNING

Previous research has successfully applied many statistical tools and machine learning methods to the analysis of OLS and student engagement, including logistic regression [8], [11], naive Bayes [27], support vector machines [31], neural network [6], k-means clustering [24], [38] and ensembles [39]. With these tools and methods, the patterns of learning activities can be revealed for the instructors and institutions to identify factors that influence student engagement.

However, these findings are usually presented in numbers and tables, which makes it difficult to show the results in a way such that both students and instructors can obtain an intuitive impression and conduct further interactive analysis, especially when the result involves many factors of multiple perspectives. To address this issue, a new trend in data science applies visualization technologies throughout the analysis process, from data exploration to presenting analysis results to users.

Using visual aids, such as charts [45] and visual badges [47], can better support students in their learning process. For example, Ilves et al. [45] used radar charts to support self-regulated learning and found that the lowest-performing students can benefit from this visualization. Ishizue et al. [46] presented a program visualization tool called PlayVisualizerC for novice C language programmers to learn the concept of memory management. Auvinen et al. [47] found that visual achievement badges can also have a positive impact on students' online study practice.

Additionally, instructors can use visual analytical tools to explore student engagement. For example, Coffrin et al. [41] proposed a state transition diagram to help instructors analyze student group movement between categories of engagement states. Chen et al. [42] developed a visualization tool called PeakVizor to investigate viewing patterns in clickstream data. Xia and Wilson [43] developed a comparative heatmap tool that enables instructors to explore and compare student video engagement. Fu et al. [44] developed a visual analytics tool, iForum, to examine the topic change patterns in online forums. These works apply a variety of visualization and interactive techniques in presenting and analyzing students' online learning activities. With the support of these tools, instructors can clearly and intuitively identify the patterns behind the data to provide better support for students.

In summary, although the existing research has studied the influence of various learner support services on student performance, retention, and course completion in many aspects, the influence of students' use of learner support services on their engagement with the LMS in the long-term learning process is still unclear. Inspired by these works, we plan to apply visualization techniques to explore the use of learner support services in large-scale online learning and its impact on the patterns of student engagement.

III. THE CONTEXT OF THIS STUDY

A. THE ODE LEARNING PROCESS

The entire ODE learning process, i.e., from entry to graduation, is similar to the online undergraduate programs provided by distance learning universities, such as the Open University in UK and the Athabasca University in Canada. Unlike popular online learning programs such as Coursera, edX, FutureLearn, Khan Academy and other MOOCs that provide training courses for specific skills or knowledge to any learners, ODE provides a long-term online academic program that includes a curriculum with references to the same majors of full-time colleges.

In our ODE school, according to the requirements of the curriculum of each major, students must take approximately 25 courses totaling 750 hours of courseware videos in at least 4 semesters over 2 years. Take computer science major as an example. There are 22 courses in this major, and each student studies 5-6 course per semester. In the first semester, students mainly study public basic courses, including 4 public foundation courses (distance learning 101, English, computing basics, and philosophy) and 2 major foundation courses

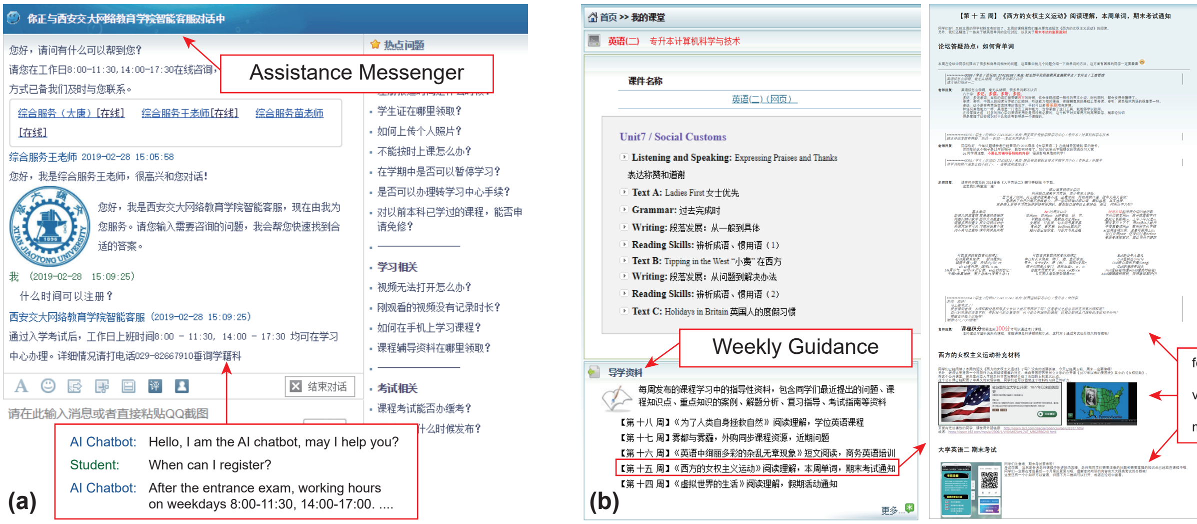

FIGURE 1. The screenshot of online learning support services in our LMS. (a) The AM service. (b) The WG service.

(network basics and programming language); In the second semester, there are 5 major foundation courses (e.g., data structure and discrete mathematics); In the third semester, there are 3 major foundation courses and 3 core courses (e.g., fundamentals of compiling and database theory); In the fourth semester, there are 5 major core courses (e.g., operating system and software engineering), and student may begin the graduation project in this semester.

To provide this long-term online learning, the ODE school has built the LMS that consists of many educational functions, such as video viewing, textbook reading, virtual experiments, online forums and exams. Moreover, various educational resources are also provided on the LMS platform, including courseware videos and reading materials, such as lecture notes and slides.

B. THE OLS SERVICES IN THE ODE SCHOOL

From the perspective of student perception, using a variety of functions in the LMS and learning such a large amount of materials offers the opportunity to master new knowledge and skills and new challenges, such as setting their study schedule and finding learning resources. Some experienced students can manage their learning on their own, but most students are not successful in the beginning.

To help all students cope with these challenges and achieve academic success, ODE schools have provided a variety of offline learner support services, including novice guidance videos, technical assistance hotline, and student manuals. These services, especially those with instructors involved, are highly rated by students and have contributed to their online learning process. However, these common learner support services can solve only general problems, such as LMS operations and technical failures. In addition, the instructors' available time is limited, and instructors cannot respond to requests all day.

Therefore, ODE school developed customized OLS services based on web technologies to further help students obtain help from instructors and faculty. Since these services are not compulsory in the normal learning process, they are designed as separate systems and embedded in the course page of LMS as a widget link. Students can decide whether to use according to their own progress. These OLS services consist of two parts: the AM service and the WG service.

The AM service is a web-based instant messaging tool that allows sending text messages, pictures, videos and audio files within the LMS. As shown in Figure 1(a), students can use this service to directly contact teaching assistants, technicians, or faculty during working hours to obtain support with studies, exams, systems or school events. Moreover, at times other than working hours, the AM service can also answer some common questions through a conversational chatbot, which is trained on the dataset of Q&A history.

The WG service is regularly released multimedia material for each course (usually 5 to 7 days depending on the course), which is edited by course instructors and teaching assistants. In each issue, they summarize the questions and answers recently raised by students, explain the difficulties in the exercises, and recommend various resources related to course learning, such as reading materials, thinking questions, audios, videos, and software. Figure 1(b) shows its location on the course page and a typical issue. Although some of the content of this service is similar to the discussions in the online forum, this service is still unique because of the following two characteristics: First, based on the feedback of the students and the current course progress in the teaching plan, instructors and teaching assistants dynamically adjust the content of each issue, which includes the course resources and the final exam focus, the schedule of offline Q&A, and career-related content. Second, due to the large number of students in each course, the flood of threads and topics often

TABLE 1. List of the datasets and related attributes.

Learner Log Data	
Student ID	Student identifier in the LMS
Timestamp	The time when student access a function
Function ID	Video player, forum, exam, AM and WG services, etc.
Course ID	The course identifier
Operation	Play video, post message, submit homework etc.
Courseware	Resource identifier
Viewing time[a]	Time spent on viewing video, in seconds
Course Data	
Course ID	Course identifier
Video Count	Number of videos in a course
Video List	List of the videos in a course
Courseware Video Data	
Video ID	Video identifier
Duration	Length of a video, in seconds

[a] The viewing time is measured by the JavaScript program on the browser side when viewing videos.

TABLE 2. The indicators for measuring student engagement.

The number of interactions	
SLCount	Total count of student-LMS interactions for a student, including login system, check calendar, manage profile, manage notification, etc.
SCCount	Total count of student-content interactions for a student, including view video, read material, listen audios, etc.
SSCount	Total count of student-student interactions for a student, including
The time of online study	
ODays	The number of days a student is online
VTime	Time spent on viewing courseware videos
The utilization of learning resources	
AR	The attendance rate measures the ratio of the number of viewed videos to the total number of videos by a student. Multiple watching of same video is regarded only once.
UR	The utilization rate measures the ratio of the total time spent by a student viewing videos to the total duration of all videos in a course. Unlike AR, the duration of repeatedly watching same video is cumulative.

overwhelm the information posted by instructors. Using this summative material and a separate link to this service page on the course cover can help students save time in finding and reading materials.

Although the technologies and forms that both the AM and WG services rely on have been used in other fields, such as social networks and online media (e.g., WhatsApp and Facebook), these features are not commonly used in ODE platforms, and the impact of these services on student engagement has not yet been analyzed. In addition, compared to common learner support services, the AM and WG services require more computing resources and more instructor involvement; thus, how these services are used is also a concern of ODE schools.

C. DATASET

The present study analyzed a dataset collected from our LMS, which consists of learner log data, course data and video data. Table 1 summarizes the types and main attributions of the dataset. The learner log data were recorded by the LMS platform and saved in the database for evaluating students' learning processes and research purposes. The raw learner log data contain many attributes, and some attributes are irrelevant to this study (e.g., browser type, and operation system). Thus, we developed a preprocessing program to remove irrelevant attributes and reserve those attributes listed in Table 1 for further analysis. To protect student privacy, no personal identification information was included in the log file. In addition, we collected the course data and detailed video information related to each course from the LMS. These two data were used in our indicators to measure student engagement. The details of these indicators are described in the next section.

The datasets used in this study were collected from the undergraduate program in the ODE school, which ran from 2015 to 2018, with a total of 4 or 5 semesters. In 2015, 14,939 students were enrolled in this program, and 10,529 of them were included in this study. The remaining 4,410 students were excluded due to different learning environments. For example, in some rural areas of the less-developed western regions, due to insufficient network bandwidth, ODE school provided course video DVDs and other materials for these students. We could neither collect their log data nor analyze their learning process.

In the first semester of this program, all students of each major must take approximately 6 courses and pass the course exams to earn credits, which include 4 public foundation courses (i.e., distance learning 101, English, computing basics, and philosophy) and 2 major foundation courses (e.g., network basics and programming language in the computer science major). Although there are various differences in each major, the public foundation courses are required for all majors. Therefore, we chose the English course and related log records for analysis. After preprocessing, the dataset included 14,025,342 rows of log records and 118 rows of video records.

IV. METHOD

In this section, we describe the measures used in the study and the analysis process performed on the collected datasets.

A. MEASUREMENTS

As discussed in Section 2.1, many indicators have been proposed in previous studies to measure student engagement. With reference to existing studies, we use the following three types of indicators as a proxy of student engagement: the number of interactions, the time of online study, and the utilization of video resources. The details are as follows.

The number of interactions. We counted three types of interactions to describe the students' online activities: student-LMS system (SLCount), student-content (SCCount), and student-student (SSCount) [6], [16]. These indicators were measured by counting the number of operations in the learner log data.

The time of online study. Research usually focuses on the time students spend on various activities, such as session time and task time. In this study, we measure the following two indicators to describe the time that students spend on online learning: online days (ODays) and viewing time (VTime). Since the learning time of a semester in an ODE program is much longer than common online courses (20+ weeks vs. 4-8 weeks), we use the number of days of online learning to measure the overall level of course engagement, which has been used and validated in many studies [24], [28], [39]. Moreover, watching courseware videos is the main activity of online learning in the ODE school, and the length of watched courseware videos is also measured to reflect the course engagement [16].

The utilization of video resources. Although there are various types of learning resources available on the LMS (e.g., courseware videos and textbooks), video is the most important resource since most knowledge is taught through video, and students spend most of their online time viewing courseware videos. As a result, we use the following two indicators of video utilization to measure student engagement: attendance rate (AR) and utilization rate (UR). For example, if a course has 10 videos and each one is 60 minutes long, and a student repeatedly watches the same video 5 times for 30 minutes each time, then, the student's course AR is $1/10 = 0.1$ and UR is $(5 \times 30)/(10 \times 60) = 0.25$. These two indicators combine the number and duration of course videos to eliminate the differences between courses; thus, they are also suitable for comparing the learning progress between different courses.

Pre-analysis data exploration included plots of probability density function and the cumulative distribution function of each indicator. These plots showed a long-tailed distribution in the count indicators (i.e., SLCount, SCount, and SSCount), which indicates that most students have a small amount of interactions, while a small number of students have a large number of interactions. In order to remove most of the nonlinearity, we applied a log transformation to each count indicators. These log-transformed indicators were used in the following models, while their original forms were used for visualization to facilitate interpretation.

In addition to the above indicators for measuring student engagement, we also count the number of AM and WG services used by each student during the semester for subsequent analysis.

B. DATA ANALYSIS

Using a program written in Python, we collected learner log data, course data and video data from the LMS and saved the preprocessed data into a MySQL database. To analyze the data, we first divided students into four groups according to their use of AM and WG services. Then, we used SQL and Python scripts to calculate the indicators listed in Table 2 to measure student engagement during the whole semester and the usage of AM and WG services. Next, correlation analysis was performed to explore the relationships between the use of AM and WG services and student engagement. Finally, a clustering analysis was conducted based on the t-distributed stochastic neighbor embedding (t-SNE) algorithm [32] and DBSCAN algorithm [49] using the Python scikit-learn library [36], which generated five clusters.

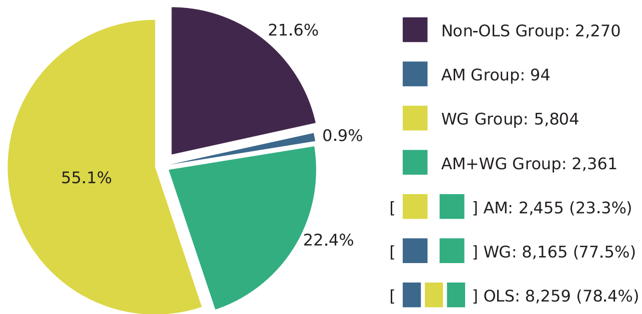

FIGURE 2. The proportion of students using AM and WG services.

TABLE 3. The number of using AM and WG services per student.

Service	Usage of OLS Services (Median (25%, 75%)) AM Group	WG Group	AM+WG Group
The AM Service	1 (1, 1)	N/A	1 (1, 2)
The CG Service	N/A	12 (6, 20)	15 (9, 24)

The t-SNE is a nonlinear dimensionality reduction algorithm in the field of machine learning that transforms the points in a high-dimensional space to a lower-dimensional space, typically the 2D plane. The t-SNE algorithm is sensitive to the local structure in the high-dimensional data and has advantages in revealing the structures, manifolds or clusters, which makes it suitable for exploring the patterns of high-dimensional data [32], [33]. In this study, we used t-SNE to embed the 7-dimensional engagement representation of each student (i.e., the 7 indicators listed in Table 2) in a 2D map for both visualization and analysis purposes. Due to the tunable parameters and slow initialization of t-SNE, we first ran the combinations of multiple sets of parameters in parallel to select the appropriate parameter values (e.g., perplexity, learning rate, and the number of iterations) using the Barnes-Hut approximation [34], [48]. Then, we used the DBSCAN algorithm as an exploratory approach to obtain the number of student engagement clusters based on the density of points on the 2D map.

The low-dimensional representation of student engagement generated by t-SNE reflects the probability distribution over pairs of high-dimensional objects; thus, the students with similar patterns of course engagement were placed closer, and the different clusters may indicate different patterns of engagement. Due to the randomness of t-SNE algorithm initialization, the low-dimensional representation is not the same every run, but the number and relative positions of the clusters are similar to the same parameters.

To compare the student engagement of the four groups and the five clusters, group comparison methods were conducted. Before analysis, the variables were checked for normal distribution and homogeneity of variance. As assessed by Shapiro-Wilk's test ($p < 0.01$), all variables violated the assumption

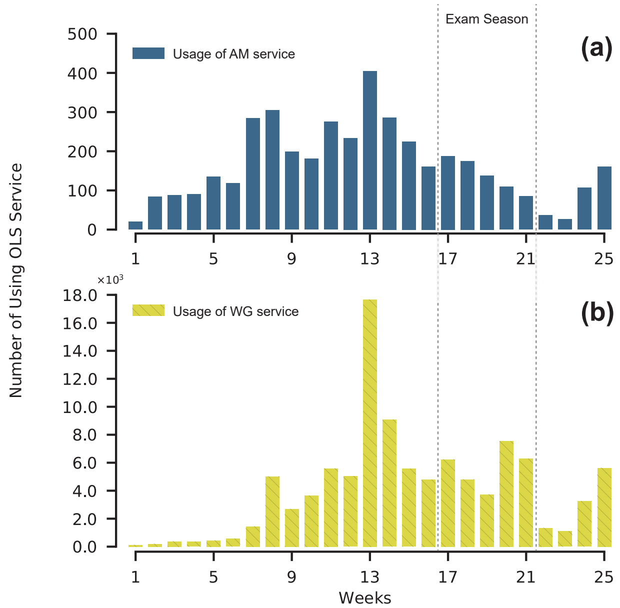

FIGURE 3. The weekly usage of (a) the AM service, and (b) the WG service.

of normality. To address these violations and for the sake of consistency, Kruskal-Wallis nonparametric tests were used. Since the Kruskal-Wallis tests were statistically significant, post hoc analyses were performed using Mann-Whitney U nonparametric tests.

V. ANALYSIS AND RESULTS

This section presents the results for answering the three research questions in the same order as the questions posed.

A. USAGE OF AM AND WG SERVICES

- **Non-OLS Group**. Students who did not use any OLS service (2,271 students, 21.5 percent).
- **AM Group**: Students who used only the AM service (94 students, 0.9 percent).
- **WG Group**: Students who used only the WG service (5,804 students, 55.1 percent).
- **AM+WG Group**: Students who used both AM and WG services (2,361 students, 22.4 percent).

The descriptive statistics (median, 25th and 75th percentile) on the usage of the OLS services during the semester are listed in Table 3. Overall, the students in the AM+WG group used both services more often than the other two groups. In addition, half of the students in both the AM group and AM+WG group used the AM service only once.

Figure 3 shows the statistics on the usage of the AM and WG services over the whole semester. In general, the use of the AM service (blue bars) was much less than the WG service (yellow bars). In the first few weeks of the semester, both AM and WG services were rarely used. Then, the use of both services increased significantly in the middle of the semester and reached the highest value in week 13. After the course exam was over near the end of the semester in week 22, the use of both services dropped.

TABLE 4. Descriptive statistics and correlation analysis (N=8,259).

	SL	SC	SS	OD	VT	AR	UR
SLCount	-						
SCCount	0.77**	-					
SSCount	0.40**	0.32**	-				
ODays	0.86**	0.78**	0.34**	-			
VTime	0.40**	0.57**	0.12**	0.40**	-		
AR	0.69**	0.77**	0.25**	0.76**	0.70**	-	
UR	0.65**	0.75**	0.23**	0.71**	0.91**	0.84**	-
N_{OLS}	0.41**	0.34**	0.33**	0.37**	0.09**	0.23**	0.22**
$N_{OLS\leq30}$	0.34**	0.27**	0.28**	0.31**	0.08**	0.20**	0.19**
$N_{OLS>30}$	0.13**	0.09**	0.03	0.04	−0.03	−0.00	0.00
Median	76	362	1	22	5.2	0.122	0.087
25%	44	193	0	13	1.8	0.033	0.017
75%	123	613	7	35	9.6	0.237	0.186

Note: * $p < 0.05$; ** $p < 0.01$

N_{OLS}: Number of OLS services used by a student

TABLE 5. The results of Kruskal-Wallis tests.

Indicator	4 Groups	5 Clusters	5 Clusters and 2 Subs
SLCount	4247.805**	4281.259**	5528.706**
SCCount	947.307**	6787.802**	9011.462**
SSCount	2013.067**	2572.283**	3090.377**
ODays	3228.552**	4101.983**	5186.717**
VTime	3526.648**	3158.681**	3885.050**
AR	4316.061**	3859.029**	5164.547**
UR	3824.955**	3570.623**	4703.843**

Note: * $p < 0.05$; ** $p < 0.01$

B. STUDENT ENGAGEMENT

Figure 4 illustrates the distribution of student engagement on each indicator. In general, students in the AM+WG group showed the highest level of course engagement among the four groups, followed by the WG group, AM group and non-OLS group. To examine whether the four groups were different in terms of their engagement, Kruskal-Wallis tests were conducted. The results revealed significant differences in the four groups on each indicator (Table 5). Pairwise Mann-Whitney U-tests revealed statistically significant differences in all comparisons (Table 6). These results indicate that the students who used OLS services were more engaged in online learning. Although the overall course engagement of the non-OLS group was lower than other groups, some non-OLS students showed a small gap in the SCCount indicator (i.e., they had more student-content interactions), which may indicate that different patterns exist in the non-OLS group.

The results of the correlation analysis of the study indicators are presented in Table 4. These results are in line with those obtained by Cerezo et al. [38]. As expected, due to the inherent relevance of online learning behaviors, the indicators related to student engagement in the LMS were positively related. For example, students must log in to LMS to access various resources and functions, so the number of online days (ODays) is clearly positively correlated with other indicators. Despite the overall relevance, due to the differences in the distribution of these indicators, student

TABLE 6. The results of pairwise Mann-Whitney U-tests between 4 groups.

Groups	SLCount	SCCount	SSCount	ODays	VTime	AR	UR
Non-OLS vs. AM	42159.5**	89285.5**	71080.0**	70889.5**	76110.5**	72624.5**	75650.0**
Non-OLS vs. WG	1163091.5**	4303189.0**	3407478.5**	1797974.0**	1322975.5**	798012.0**	1157484.0**
Non-OLS vs. AM+WG	204328.5**	1501391.5**	924497.0**	463018.5**	491294.5**	263433.0**	393836.0**
AM vs. WG	164740.0**	189013.5**	218969.5**	148119.5**	125974.5**	108698.5**	114153.5**
AM vs. AM+WG	40314.0**	64337.5**	65003.0**	43674.5**	47818.0**	37951.5**	39904.5**
WG vs. AM+WG	4755501.0**	5348489.5**	5373261.0**	5543819.5**	6511264.0**	5940166.0**	6061197.0**

Note: * $p < 0.05$; ** $p < 0.01$

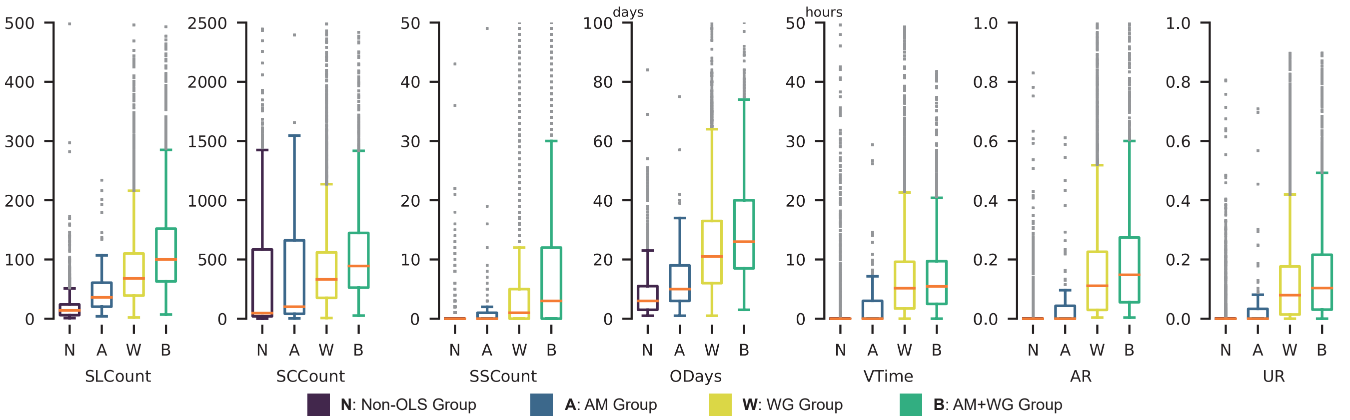

FIGURE 4. Student engagement among four groups.

engagement may shows different patterns, which will be described in next section.

In addition, we further examined the correlation between the number of OLS services used and each indicator in different ranges. The results showed that when the number of OLS services used was greater than 30, there was no significant correlation between all indicators except student-LMS interaction and student-content interaction, which had a significantly weak positive correlation.

C. THE PATTERNS OF STUDENT ENGAGEMENT

To address the third research question, we used t-SNE and DBSCAN algorithms to analyze the patterns of student engagement and identified five clusters. Figure 5(a1) shows a typical case of the 2D map of all students' course engagement produced by t-SNE, in which each dot represents a student, and its color is encoded by group. Figure 5(a2) and Figure 5(a3) use the same layout as that of Figure 5(a1) but use different color encoding. Figure 5(a2) shows the five color-encoded clusters, and each color represents a cluster. Figure 5(a3) shows the distribution of OLS service usage, where each student's OLS service usage is represented by a color (red indicates usage greater than median=13, while blue indicates less). The proportion of students in each cluster is shown in Figure 5(b), and the engagement among these clusters is shown in Figure 5(c). The characteristics of each cluster are summarized as follows.

- *Cluster 1*. In this cluster, most of the students belonged to the non-OLS group (88 percent), followed by the WG group (9 percent). Cluster 1 shows the lowest performance on all 7 indicators, which indicates that they rarely used the features and course resources on the LMS. Therefore, cluster 1 was labeled as "few engagement students".
- *Cluster 2*. This cluster is characterized by a moderate level of student engagement. Moreover, since cluster 2 is the largest of all clusters (N=6,771 (64 percent)) and the proportion of WG groups is also the highest (68.2 percent), cluster 2 was labeled as "common engagement students".
- *Cluster 3*. This cluster is similar to cluster 2 in the proportion of the WG group (63.9 percent) for all indicators except for the student-student interaction (SSCount). In cluster 3, the student-student interaction is significantly higher than that of other clusters, which indicates that the students in cluster 3 viewed and posted more messages than the students in the other clusters. Thus, cluster 3 was labeled as "extensive discussion students".
- *Cluster 4*. This cluster is similar to cluster 1, but the students in this cluster spent more days accessing course contents. Nevertheless, cluster 4 viewed few videos, while the student-content interaction was quite high, which indicates that the students in cluster 4 accessed more text materials than videos, so we labeled cluster 4 as "text preferred students".
- *Cluster 5*. In cluster 5, the students displayed much higher levels in all indicators than the students in other clusters, except for student-student interaction. Therefore, we labeled cluster 5 as "extensive engagement students". In addition, we found an interesting pattern

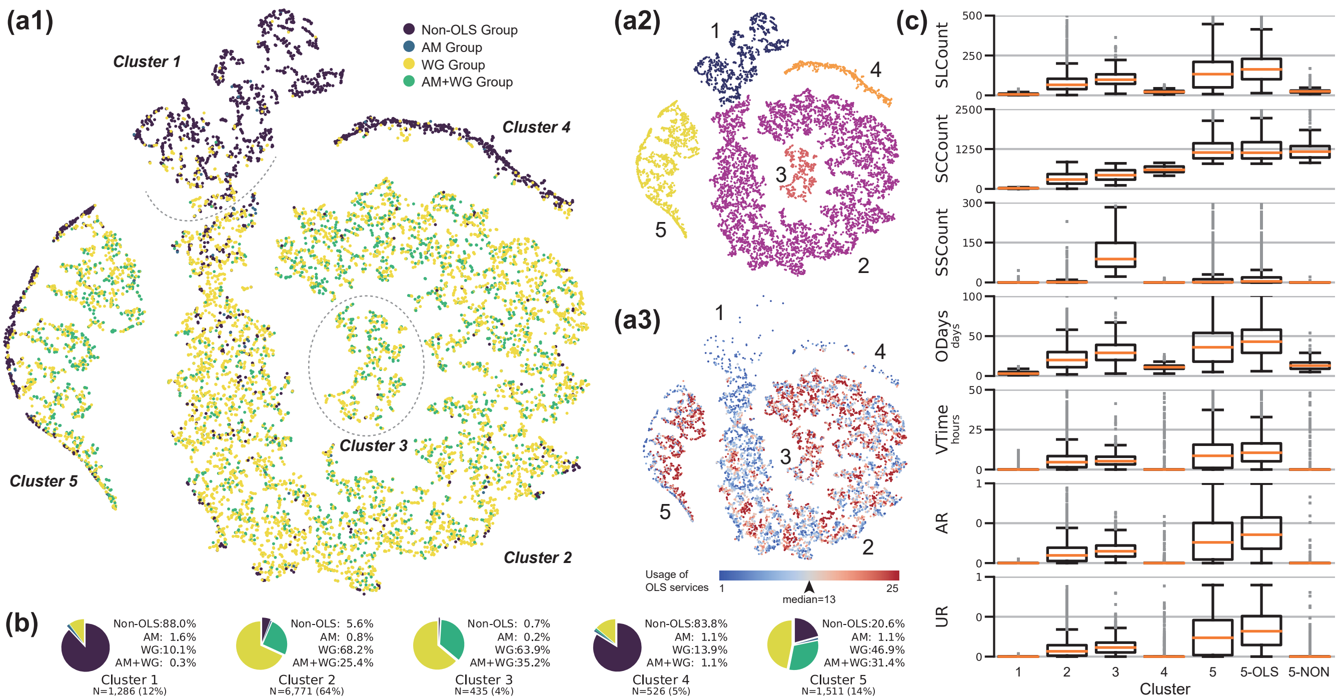

FIGURE 5. The five clusters. (a) The visualization generated by t-SNE. (b) Proportion of the four groups in each cluster. (c) The patterns of each cluster.

in which the students of the non-OLS group in this cluster were concentrated at the left edge and the proportion was high (20.6 percent). Therefore, we divided the cluster into two subclusters, i.e., students of other groups (subcluster 5-OLS) and students of the non-OLS group (subcluster 5-NON), to analyze the differences in engagement. As shown in Figure 5(c), in the 5-NON subcluster, the student-content interaction (SCCount) is quite high and similar to the 5-OLS group, while other indicators are low and similar to cluster 4.

In addition, Figure 5(a3) shows an uneven distribution of the usage of OLS services. The upper half of cluster 2, most of cluster 3, and the subcluster 5-OLS show higher usage of OLS services (i.e., above the median), while other parts and clusters show less. This result further indicates that students who used the OLS service may have a higher level of course engagement.

To clearly understand the student engagement patterns among the three clusters, we used Kruskal-Wallis tests to compare the five clusters in terms of all 7 indicators (Table 5). The results showed that the students in the five clusters significantly demonstrated different engagement behaviors. Pairwise Mann-Whitney U-tests revealed that statistical significance existed in all comparisons except in the student-student interaction (cluster 1 vs. cluster 4) (Table 7).

In addition, the results of comparisons among the two subclusters and other four clusters (Table 8) showed statistically significant differences in all comparisons except for the student-student interaction, viewing time, AR and UR (subcluster 5-NON vs. cluster 4), and the student-student interaction (subcluster 5-NON vs. cluster 1)

VI. DISCUSSION

The results of our study further contribute to the understanding of the importance of learner support services in online learning environments. Moreover, we also revealed some patterns of student engagement regarding the use of OLS services.

RQ1: How frequently do students use OLS services? For RQ1, the descriptive statistics revealed several interesting results. First, more than half of the students who used the AM service used it only once. The low utilization of the AM service may indicate that the AM service failed to meet the students' expectations. There could be several reasons behind this.

- Quality of services. The ODE school introduced an AI-based conversational chatbot as the default option to respond to students and answer common questions in a timely manner. When the students are not satisfied with the answer given by the chatbot, they can switch to the manual service provided by instructors or faculty. However, students may feel that the quality of the service provided by chatbot is not as expected; thus, they won't use it after the first try.
- Usability. Since the AM service provides learner support mainly through a conversational user interface, the design of the user interface and the interaction process may have an impact on the student's experience. For example, existing studies have suggested that learning support services should be able to be located within two

TABLE 7. The results of pairwise Mann-Whitney U-tests between 5 clusters.

Cluster	SLCount	SCCount	SSCount	ODays	VTime	AR	UR
C1 vs. C2	161157.0**	14653.0**	2275046.5**	218842.5**	448567.0**	223561.5**	354246.0**
C1 vs. C3	971.5**	0.0**	55.5**	1436.0**	12076.0**	2205.0**	5321.0**
C1 vs. C4	48778.0**	0.0**	**341053.0** ($p = 0.55$)	16656.0**	301809.0**	303459.5**	300841.0**
C1 vs. C5	27606.5**	0.0**	458754.5**	4523.0**	213067.5**	190217.5**	201248.5**
C2 vs. C3	936086.5**	991859.0**	11322.0**	951668.0**	1321147.5**	1122800.0**	1133550.0**
C2 vs. C4	3153402.5**	453330.5**	2633419.5**	2692265.5**	2963032.5**	3217349.5**	3018657.5**
C2 vs. C5	3250624.5**	555.0**	4389513.0**	2926244.0**	4021426.5**	3457528.0**	3304278.5**
C3 vs. C4	224259.5**	51781.0**	228810.0**	212729.5**	195564.0**	215885.5**	201828.5**
C3 vs. C5	274243.5**	5.5**	602466.0**	273306.0**	269694.5**	252992.0**	240167.0**
C4 vs. C5	80721.5**	35.5**	186933.0**	80756.5**	163579.0**	103659.5**	126577.0**

Note: * $p < 0.05$; ** $p < 0.01$

TABLE 8. The results of pairwise Mann-Whitney U-tests between two subclusters (5-OLS, 5-NON) and other four clusters.

Cluster	SLCount	SCCount	SSCount	ODays	VTime	AR	UR
5O vs. C1	1569.5**	0.0**	253381.0**	702.5**	33049.0**	8630.5**	21785.0**
5O vs. C2	1443291.5**	478.0**	2811933.0**	1468661.5**	2289208.5**	1521349.5**	1546455.0**
5O vs. C3	143702.5**	5.5**	54819.0**	153429.5**	155458.5**	123192.0**	123062.0**
5O vs. C4	8300.0**	35.5**	103641.0**	17901.0**	81448.0**	20551.5**	44524.0**
5N vs. C1	26037.0**	0.0**	**195858.5** ($p = 0.08$)	3820.5**	180018.5**	181587.0**	179463.5**
5N vs. C2	305219.0**	77.0**	534972.0**	654969.5**	380334.0**	176373.5**	354728.5**
5N vs. C3	5179.0**	0.0**	0.0**	15843.5**	21484.0**	5920.0**	18615.0**
5N vs. C4	72421.5**	0.0**	**80820.0** ($p = 0.20$)	62855.5**	**81981.0** ($p = 0.49$)	**81004.0** ($p = 0.33$)	**82053.0** ($p = 0.50$)

Note: * $p < 0.05$; ** $p < 0.01$

clicks, should have an immediate response, and should be 24/7 online [4]. As a result, if students feel that they cannot express doubts or it is not easy to operate through the interface of the current AM service, they may no longer use the service.

This result indicates that instructors need to pay attention to the number of OLS services used, which may be related to the quality of the service and the student experience. These possibilities will be investigated in future studies.

Second, although most of the students used the OLS service at least once (Figure 2), 21.6 percent of the students did not. There may be two reasons to explain this result. First, although both AM and WG services were introduced to students when they started online learning in the LMS, some students may not have been able to understand what kind of support is provided; thus, they may not have been aware that they could use the OLS services to solve their learning problems. Second, considering that they spent very few days online (Figure 4), they may not have had enough time to use all the features of the LMS, including the OLS services.

Third, from the middle to the end of the semester (i.e., from week 13 to week 21), the number of WG services used was higher than in other periods. Nevertheless, after the middle of the semester (i.e., week 13), the use of the AM service declined. These results indicate that the number of students using the WG service was related to the ODE school's exam schedule, and students may pay more attention to the WG service that includes course review materials before the exams (the exams were arranged between week 19 and week 22).

RQ2: Do students who use the OLS services show a higher level of course engagement? We divided students into four groups according to which OLS service they used, namely, the non-OLS group, AM group, CG group and AM+CG group. Since withholding OLS services intentionally is unacceptable by both students and ODE school, we cannot divide some students into a control group and conduct a controlled experiment to investigate the effect of using the OLS service. Nonetheless, the students in the non-OLS group can naturally be considered a control group, while the students in the other groups can be considered a treatment group.

As shown in Figure 4 and Table 4, the results indicate that the students who used the OLS services had a higher level of course engagement than those who did not. This finding is consistent with existing studies on the effects of learner support. For example, Simpson [5] reported that in the UK Open University, students who received an early proactive, supportive telephone call showed higher course completion rates than those who did not.

Moreover, our results further demonstrate the quantitative relationship between the number of OLS services used and student engagement. Overall, there is a positive correlation between student engagement and the number of OLS services used, but within a certain range (e.g., $N_{OLS} \leq 30$ in our case). This finding may indicate that the use of OLS services promotes student engagement, but the effect works within only a certain range of usage. Possible reasons could be various and are summarized as follows:

- *Learning difficulty*. Some studies have suggested that learners may leave an online program for reasons such as the assignments or the program are too difficult and not enough support [7]. When learners requested any of the OLS service, AM or WG during their online learning process, it may be fair to assume that they had an intention to seek help due to a problem that they could not solve. The extensive use of OLS services may indicate that they had many unsolved problems, and thus, they may have felt frustrated and decreased online learning, which would result in nongrowth of course engagement.

- *Learning by reading* As mentioned in Section 2.2, the WG service provides some course-related reading materials, including a focus on the final exam and Q&A session; thus, students can reduce the time spent watching courseware videos by reading these materials, especially when near the final exam at the end of the semester. As a result, in this case, students used OLS service more than usual, but their engagement with courseware videos did not increase accordingly.

If these reasons can be identified from available datasets of students' activities in a future study, then it might be helpful for instructors and institutions to improve learner support services. For example, institutions may organize face-to-face offline discussions with teaching assistant to help those students who have difficulty in assignments to solve their problems. Additionally, instructors may adjust the content in the WG service to guide students to watch the courseware video instead of giving them the key content. In addition, these findings also suggest that the usage of learner support service could be used to represent students' learning status.

RQ3: Are there different patterns of course engagement among students regarding the usage of OLS services? We identified five different patterns of student engagement represented in the five clusters. The details of the engagement pattern in each cluster are as follows.

Cluster 1 - few engagement students. Considering that most of the students in this cluster did not use the OLS service at all, it indicates that the students in this cluster may lack time to study online for various reasons, such as work and family responsibilities. It requires considerable time to learn online, especially those undergraduate courses, which usually require watching many videos online for dozens of hours and spending more hours completing the assignment offline. In our case, if students do not have enough time to watch online videos, then they also may not have the time to use learner support services. Moreover, existing studies reported that "couldn't keep up with deadlines" and "course requires too much time" are the two main reasons for students to drop out of online learning [8]. Therefore, when it is found that some students have low course engagement and do not use OLS services, the instructors and institutions may pay special attention to them.

Cluster 2 - common engagement students. The students of this cluster account for the majority of the whole population; thus, this cluster represents the characteristics of the online learning process of most students in the ODE school. Although some students have high utilization of video resources, most students have both AR and UR of less than 50 percent, which means that more than half of the course videos are not viewed at all. In this sense, the OLS service has not fully guided students to learn course resources and answer their questions. Therefore, future research should focus on how to provide effective support services for these students to improve their course engagement.

Cluster 3 - intensive discussion students. The extensive use of online forums may indicate that these students are more active in interacting with others, both students and teachers. Although there are fewer students in this cluster than others, they may become a complement to learner support services. Existing studies have reported that the interaction between students is positively correlated with the course final grade [16] and may improve their course completion rate [1]. Additionally, the proportion of the non-OLS group in this cluster is the lowest, which may imply that students who often interact with others are more likely to use OLS services.

Cluster 4 - text preferred students. The students in this cluster mainly spent their days and efforts on accessing the course resources. However, considering that their video utilization was low, we could argue that rather than viewing course videos and discussing in the online forums as those in clusters 2 and 3 did, they prefer to read text materials such as lecture notes and slides. Notably, most of the students in this group did not use OLS services. Thus, it is also possible that the students had mastered the course content and needed to only browse the course materials to pass the exam (e.g., some students may use English frequently in their daily work, so it is no longer necessary for them to relearn the content of this English course).

Cluster 5 - intensive engagement students. This pattern may represent the students who meet the expectations of the ODE school, i.e., those who spent many days accessing course materials and fully using course resources and features in the LMS. However, the subcluster 5-NON (i.e., the students of the non-OLS group in cluster 5) showed different learning behaviors of extensively accessing reading materials while rarely viewing videos. This result indicates that although the student-content interaction counts may be similar, their actual use patterns of course resources are varied. It is interesting that the subcluster 5-NON is very similar to cluster 4, except for the student-content interaction. The subcluster 5-NON spent more effort on text materials than cluster 4, which may suggest that the students in the subcluster 5-NON have not mastered the course content, and they have to read a large amount of text material to pass the exam due to limited study time.

These results are interesting in two different senses. Whether or not to use OLS services and the number of OLS services used seems to be highly related to the patterns of student engagement; thus, instructors can use the usage of OLS services as indicators combined with existing indicators of statistics on interactions and resource utilization to better understand students' learning process and learning status. Additionally, according to the patterns reflected in the five clusters, it may help to improve the quality of existing OLS services to help students with different characteristics.

VII. CONCLUSION AND FUTURE WORK

In this study, we applied visualization techniques to explore the impact of learner support services on student engagement. First, we categorized students into four groups and found that the students' use of OLS services varied and that the usage of the AM service was lower than expected. Second, there was a significant positive correlation between OLS services usage and student engagement. However, when the number of OLS services used exceeded a certain range, there was no correlation in terms of video utilization. Third, using t-SNE and DBSCAN algorithms, we identified five clusters representing different patterns of student engagement and usage of OLS services.

These findings may imply that instructors and institutions can use OLS service as indicators to further understand students' learning processes, and the utilization of OLS services can also be used as a reference for improving service quality.

Since a potential limitation of this study is the reduced set of indicators employed to measure student engagement and learning performance, future research will include more detailed indicators such as interactions on various types of resources and final grades to further analyze the impact of OLS on learning processes and outcomes. In addition, we only analyzed the learning data of one course in the semester in this study, and the effect of using OLS services on student engagement in the subsequent semesters will also be analyzed in the future research.

ACKNOWLEDGMENT

This research was partially supported by "The Fundamental Theory and Applications of Big Data with Knowledge Engineering" under the National Key Research an d Development Program of China with Grant No. 2018YFB1004500, the MOE Innovation Research Team No. IRT17R86, the National Science Foundation of China under Grant Nos. 61721002 and 61532015, the Project of China Knowledge Center for Engineering Science and Technology, and the Project of Shaanxi Education Informatization Development and Research No. 18JX005.

REFERENCES

[1] A. S. Sunar, S. White, N. A. Abdullah, and H. C. Davis, "How Learners' Interactions Sustain Engagement: A MOOC Case Study," IEEE Transactions on Learning Technologies, vol. 10, no. 4, pp. 475–487, Oct. 2017.

[2] C. R. Henrie, R. Bodily, R. Larsen, and C. R. Graham, "Exploring the potential of LMS log data as a proxy measure of student engagement," J Comput High Educ, vol. 30, no. 2, pp. 344–362, Aug. 2018.

[3] J. Knobloch, J. Kaltenbach, and B. Bruegge, "Increasing Student Engagement in Higher Education Using a Context-aware Q&A Teaching Framework," in Proceedings of the 40th International Conference on Software Engineering: Software Engineering Education and Training, Gothenburg, Sweden, 2018, pp. 136–145.

[4] D. H. Cox and C. C. Strange, "Student Services at a Distance," in Achieving Student Success: Effective Student Services in Canadian Higher Education. Montreal, Quebec, Canada, McGill-Queen's University Press, 2010, pp. 221-230

[5] O. Simpson, "The impact on retention of interventions to support distance learning students," Open Learning: The Journal of Open, Distance and e-Learning, vol. 19, no. 1, pp. 79–95, Feb. 2004.

[6] E. B. Gregori, J. Zhang, C. Galván-Fernández, and F. de A. Fernández-Navarro, "Learner support in MOOCs: Identifying variables linked to completion," Computers & Education, vol. 122, pp. 153–168, Jul. 2018.

[7] P. A. Willging and S. D. Johnson, "Factors that Influence Students' Decision to Dropout of Online Courses," Journal of Asynchronous Learning Networks, vol. 13, no. 3, pp. 115–127, Oct. 2009.

[8] R. F. Kizilcec and S. Halawa, "Attrition and Achievement Gaps in Online Learning," in Proceedings of the Second (2015) ACM Conference on Learning @ Scale, Vancouver, BC, Canada, 2015, pp. 57–66.

[9] [R. F. Kizilcec, M. Pérez-Sanagustín, and J. J. Maldonado, "Recommending Self-Regulated Learning Strategies Does Not Improve Performance in a MOOC," in Proceedings of the Third (2016) ACM Conference on Learning @ Scale, Edinburgh, UK, 2016, pp. 101–104.

[10] R. F. Kizilcec and G. L. Cohen, "Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures," PNAS, vol. 114, no. 17, pp. 4348–4353, Apr. 2017.

[11] R. F. Kizilcec, M. Pérez-Sanagustín, and J. J. Maldonado, "Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses," Computers & Education, vol. 104, pp. 18–33, Jan. 2017.

[12] P. McAndrew and E. Scanlon, "Open Learning at a Distance: Lessons for Struggling MOOCs," Science, vol. 342, no. 6165, pp. 1450–1451, Dec. 2013.

[13] O. Zawacki-Richter and T. Anderson, "Learner Support in Online Distance Education: Essential and Evolving" in Online Distance Education: Towards a Research Agenda, Edmonton, Canada, Athabasca University Press, 2014, pp. 287-310

[14] F. Martin, C. Wang, and A. Sadaf, "Student perception of helpfulness of facilitation strategies that enhance instructor presence, connectedness, engagement and learning in online courses," The Internet and Higher Education, vol. 37, pp. 52–65, Apr. 2018.

[15] L. Fedynich, K. S. Bradley, and J. Bradley, "Graduate Students' Perceptions of Online Learning," Research in Higher Education Journal, vol. 27, pp. 311-323, Oct. 2015.

[16] 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.

[17] P. J. Guo, J. Kim, and R. Rubin, "How video production affects student engagement: an empirical study of MOOC videos," in Proceedings of the First (2014) ACM Conference on Learning @ Scale, Atlanta, GA, USA, 2014, pp. 41–50.

[18] G. Kovacs, "Effects of In-Video Quizzes on MOOC Lecture Viewing," in Proceedings of the Third (2016) ACM Conference on Learning @ Scale, Edinburgh, UK, 2016, pp. 31–40.

[19] B. Yousuf and O. Conlan, "Supporting Student Engagement Through Explorable Visual Narratives," IEEE Transactions on Learning Technologies, vol. 11, no. 3, pp. 307–320, Jul. 2018.

[20] Q. Zhao, S. Varma, and J. A. Konstan, "In-Video Reuse of Discussion Threads in MOOCs," in Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale, Cambridge, MA, USA, 2017, pp. 153–156.

[21] C. Kent, E. Laslo, and S. Rafaeli, "Interactivity in online discussions and learning outcomes," Computers & Education, vol. 97, pp. 116–128, 2016.

[22] J. Wang, W. J. Doll, X. Deng, K. Park, and M. G. (Mark) Yang, "The impact of faculty perceived reconfigurability of learning management systems on effective teaching practices," Computers & Education, vol. 61, pp. 146–157, 2013.

[23] T. Thistoll and A. Yates, "Improving course completions in distance education: an institutional case study," Distance Education, vol. 37, no. 2, pp. 180–195, May 2016.

[24] 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.

[25] L. K. Fryer, M. Ainley, A. Thompson, A. Gibson, and Z. Sherlock, "Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners," Computers in Human Behavior, vol. 75, pp. 461–468, Oct. 2017.

[26] L. K. Fryer, K. Nakao, and A. Thompson, "Chatbot learning partners: Connecting learning experiences, interest and competence," Computers in Human Behavior, vol. 93, pp. 279–289, Apr. 2019.

[27] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. F. Wong, "Learning about Social Learning in MOOCs: From Statistical Analysis to Generative Model," IEEE Transactions on Learning Technologies, vol. 7, no. 4, pp. 346–359, 2014.

[28] J. Gardner and C. Brooks, "Dropout Model Evaluation in MOOCs," in Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2018, pp. 7906-7912
[29] V. Singh et al., "A Content Engagement Score for Online Learning Platforms," in Proceedings of the Fifth Annual ACM Conference on Learning at Scale, London, UK, 2018, pp. 25:1–25:4.
[30] 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, Edinburgh, UK, 2016, pp. 51–60.
[31] M. L. Bote-Lorenzo and E. Gómez-Sánchez, "Predicting the Decrease of Engagement Indicators in a MOOC," in Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, British Columbia, Canada, 2017, pp. 143–147.
[32] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," Journal of Machine Learning Research, vol. 9, pp. 2579–2605, Nov, 2008.
[33] L. van der Maaten, "Accelerating t-SNE using Tree-Based Algorithms," Journal of Machine Learning Research, vol. 15, pp. 3221–3245, Oct, 2014.
[34] N. Pezzotti, B. P. F. Lelieveldt, L. v d Maaten, T. Höllt, E. Eisemann, and A. Vilanova, "Approximated and User Steerable tSNE for Progressive Visual Analytics," IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 7, pp. 1739–1752, Jul. 2017.
[35] F. Simui et al., "'WhatsApp' as a Learner Support tool for distance education: Implications for Policy and Practice at University of Zambia," Zambia ICT Journal, vol. 2, no. 2, pp. 36–44, Nov. 2018.
[36] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, Oct, 2011.
[37] K. F. Hew, "Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS," Br J Educ Technol, vol.47, no. 2, pp. 320–341, Mar. 2015.
[38] R. Cerezo, M. Sánchez-Santillán, M. P. Paule-Ruiz, and J. C. Núñez, "Students' LMS interaction patterns and their relationship with achievement: A case study in higher education," Computers & Education, vol. 96, pp. 42–54, May. 2016.
[39] W. Xing, X. Chen, J. Stein, and M. Marcinkowski, "Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization," Computers in Human Behavior, vol. 58, pp. 119–129, May 2016.
[40] K. S. Hone and G. R. El Said, "Exploring the factors affecting MOOC retention: A survey study," Computers & Education, vol. 98, pp. 157–168, 2016.
[41] C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy, "Visualizing Patterns of Student Engagement and Performance in MOOCs," in Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, Indianapolis, IN, USA, 2014, pp. 83–92.
[42] Q. Chen, Y. Chen, D. Liu, C. Shi, Y. Wu, and H. Qu, "PeakVizor: Visual Analytics of Peaks in Video Clickstreams from Massive Open Online Courses," IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 10, pp. 2315–2330, Oct. 2016.
[43] J. Xia and D. C. Wilson, "Instructor Perspectives on Comparative Heatmap Visualizations of Student Engagement with Lecture Video," in Proceedings of the 49th ACM Technical Symposium on Computer Science Education, Baltimore, MD, USA, 2018, pp. 251–256.
[44] S. Fu, J. Zhao, W. Cui, and H. Qu, "Visual Analysis of MOOC Forums with iForum," IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 1, pp. 201–210, Jan. 2017.
[45] K. Ilves, J. Leinonen, and A. Hellas, "Supporting Self-Regulated Learning with Visualizations in Online Learning Environments," in Proceedings of the 49th ACM Technical Symposium on Computer Science Education, New York, NY, USA, 2018, pp. 257–262.
[46] R. Ishizue, K. Sakamoto, H. Washizaki, and Y. Fukazawa, "PVC: Visualizing C Programs on Web Browsers for Novices," in Proceedings of the 49th ACM Technical Symposium on Computer Science Education, Baltimore, MD, USA, 2018, pp. 245–250.
[47] T. Auvinen, L. Hakulinen, and L. Malmi, "Increasing Students' Awareness of Their Behavior in Online Learning Environments with Visualizations and Achievement Badges," IEEE Transactions on Learning Technologies, vol. 8, no. 3, pp. 261–273, Jul. 2015.
[48] M. Wattenberg, F. Viégas, and I. Johnson, "How to Use t-SNE Effectively," Distill, vol. 1, no. 10, pp. e2, Oct. 2016.
[49] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, USA, 1996, pp. 226–231.

HUAN HE received the B.S. degree in software engineering, the M.E. degree in software engineering from Xi'an Jiaotong University, in 2007 and 2012 respectively. He is currently a Ph.D student at the school of electronic and information engineering, Xi'an Jiaotong University, supervised by Prof. Qinghua Zheng. His research interests include web visualization application, machine learning, and educational data mining.

QINGHUA ZHENG received the B.S. degree in computer software in 1990, the M.S. degree in computer organization and architecture in 1993, and the Ph.D. degree in system engineering in 1997 from Xi'an Jiaotong University. He did postdoctoral research at Harvard University from February 2002 to October 2002 and visiting professor research at HongKong University from November 2004 to January 2005. Since 1995, he has been with the Department of Computer Science and Technology at Xi'an Jiaotong University. He is currently a professor and serves as the vice president of Xi'an Jiaotong University. His research interests include intelligent e-Learning, network security and trusted software.

DEHAI DI received the B.S. degree in computer science and technology, the M.E. degree in management engineering, and the Ph.D. degree in management science and engineering from Xi'an Jiaotong University. He is currently an associate professor in school of management, Xi'an Jiaotong University, and serves as the dean of school of continuing education, Xi'an Jiaotong University. His research interests include e-Business, information system and Informatization engineering.

BO DONG received his Ph.D. degree in computer science and technology from Xi'an Jiaotong University in 2014. He did postdoctoral research in the MOE Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University from 2014 to 2017. He currently serves as the research director in the School of Continuing Education, Xi'an Jiaotong University. His research interests focus on intelligent e-Learning and data mining.