Internet and Higher Education 29 (2016) 23–30

Contents lists available at ScienceDirect

Internet and Higher Education

Identifying significant indicators using LMS data to predict course achievement in online learning

Ji Won You *

Department of Early Childhood Education, Gachon University, 1342 Sungnamdaero, Sujeong-gu, Sungnam-si, Gyeonggi-do 406-799, Republic of Korea

ARTICLE INFO

Article history:
Received 11 June 2015
Received in revised form 4 November 2015
Accepted 26 November 2015
Available online 26 November 2015

Keywords:
LMS data
Self-regulated learning
Course achievement
Learning analytics
Online learning

ABSTRACT

This study sought to identify significant behavioral indicators of learning using learning management system (LMS) data regarding online course achievement. Because self-regulated learning is critical to success in online learning, measures reflecting self-regulated learning were included to examine the relationship between LMS data measures and course achievement. Data were collected from 530 college students who took an online course. The results demonstrated that students' regular study, late submissions of assignments, number of sessions (the frequency of course logins), and proof of reading the course information packets significantly predicted their course achievement. These findings verify the importance of self-regulated learning and reveal the advantages of using measures related to meaningful learning behaviors rather than simple frequency measures. Furthermore, the measures collected in the middle of the course significantly predicted course achievement, and the findings support the potential for early prediction using learning performance data. Several implications of these findings are discussed.

1. Introduction

Online learning has become a conventional mode of learning in higher education. Not only has the number of online educational institutions increased, but an increasing number of traditional universities are offering online courses to meet students' needs. Furthermore, massive open online courses (MOOCs) are now being offered to the public. Thus, online learning has attracted many students and provides additional learning opportunities.

Several researchers have studied the factors that are important to improving online learning and have found self-regulation to be a crucial factor in this regard (Rakes & Dunn, 2010; Sun, Tsai, Finger, Chen, & Yeh, 2008; You & Kang, 2014; Yukselturk & Bulut, 2007). Online learners are responsible for initiating, planning, and conducting their studies, but many online learners have expressed how difficult it is to maintain their motivation and persistence throughout a course (Elvers, Polzella, & Graetz, 2003; Levy & Ramin, 2012; Michinov, Brunot, Le Bohec, Juhel, & Delaval, 2011). Previous research has shown that failure to study regularly leads to poor academic achievement, and procrastination and withdrawals have proven to be persistent problems in online learning. Therefore, the ways in which strategic support and the self-regulation of online learners influence learning should be investigated to keep students motivated, regulated, and participating in their courses.

In many studies of self-regulated learning, a self-report questionnaire is generally used to measure the level of self-regulation (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007), which raises concerns regarding whether self-reported data properly represent actual self-regulated learning behaviors in authentic learning contexts. However, self-regulated learning in an online learning environment can be traced because students' learning behaviors are automatically recorded by learning management systems (LMSs). At present, LMS use has become common in most institutions, and LMSs provide new opportunities to monitor students' learning participation and progress (You, 2015). Furthermore, analyzing LMS data allows instructors to discover meaningful patterns (Gašević, Dawson, & Siemens, 2015), to identify at-risk students, to provide proactive feedback, and to adjust instructional strategies (Dietz-Uhler & Hurn, 2013). This approach is called learning analytics, and it enables data-driven decision making while improving institutional productivity. Several researchers have predicted that educational data mining will be extensively employed to optimize institutional decision making, to resolve academic problems, and to enhance students' performances in higher education within a few years (Johnson, Adams Becker, Estrada, & Freeman, 2014; Reyes, 2015).

Although the field of learning analytics is still in its infancy, prior research regarding online learning has attempted to use log or LMS data to examine online learning success. According to studies that have utilized students' log data, frequency measures, such as the number of content views, the frequency of logins, and the time spent reading pages, are the most typical measures used to explain individual differences in online learning (Morris, Finnegan, & Wu, 2005; Qu & Johnson, 2005). Numerous studies (Campbell, Finnegan, & Collins, 2006; Johnson, 2005; Morris et al., 2005; Wang & Newlin, 2002) have reported

* Corresponding author.
E-mail address: uimagine@gachon.ac.kr.

http://dx.doi.org/10.1016/j.iheduc.2015.11.003
1096-7516/

a significant relationship between active participation in online courses and academic performance.

However, several studies (Hadwin et al., 2007; Misanchuk & Schwier, 1992) have claimed that frequency counts of events are minimally relevant to engaged learning and that such measures are limited to suggesting instructional interventions and providing practical learning guidance. From this perspective, researchers need to use LMS data to identify more meaningful measures that are congruent with learning and instructional theories. Hadwin et al. (2007) suggested that the use of elaborated time-based indicators from students' log data, rather than the simple time spent on a specific issue, enables descriptions of students' self-regulated learning. However, notably few attempts have been made to identify appropriate measures of self-regulated learning and to examine the effects on course success.

In this context, the present study aims to identify significant LMS data indicators, including self-regulated learning indicators, to predict course achievement. Additionally, this study examines whether the data collected in the middle of the course can successfully predict final course achievement, which would contribute to the possibility of early prediction based on the learning analytics approach.

2. Theoretical background

2.1. Self-regulated online learning

Self-regulation is defined as setting one's goals and managing one's own learning and performance (Driscoll, 2000), and self-regulated students are described as "metacognitively, motivationally, and behaviorally active participants in their own learning process" (Zimmerman & Martinez-Pons, 1988, p. 284). Many self-regulated learning studies in traditional learning contexts have generally indicated that learners who frequently use self-regulated learning strategies exhibit better academic achievement (Mega, Ronconi, & De Beni, 2014; Zimmerman & Martinez-Pons, 1988), more intrinsic motivation (Pintrich & Zusho, 2002), and greater persistence (Pintrich & De Groot, 1990) than those who use fewer self-regulated learning strategies.

In online learning, students need to be more responsible for their studies due to the autonomous nature of the learning environment (Dabbagh & Kitsantas, 2004; Joo, Joung, & Kim, 2014; You & Kang, 2014). Students who achieve success in online courses can be described as those who have an understanding of the responsibilities and discipline needed to complete the work. Successful students actively participate in their learning in terms of regularly accessing course notices, carefully studying and reviewing the course content, completing the assignments in a timely manner, self-evaluating their learning, asking questions when they need help, and attentively communicating with others. By contrast, unsuccessful learners are characterized by their failures in estimating the amount of time and effort required to complete tasks and their lack of time-management and life-coping skills (You & Kang, 2014; Yukselturk & Bulut, 2007). Furthermore, self-regulation failures in online learning contexts have been suggested to lead to greater detrimental effects (Dabbagh & Kitsantas, 2004; Jonassen, Davidson, Collins, Campbell, & Haag, 1995; King, Harner, & Brown, 2000; Warnock, Bingham, Driscoll, Fromal, & Rouse, 2012) compared with those obtained from failures in traditional learning environments. Self-regulation failures easily elicit academic procrastination, and procrastinating in online learning has a greater negative impact on achievement (Klingsieck, Fries, Horz, & Hofer, 2012; Tuckman, 2005; Wolters, 2003; You, 2015) and often results in dropout. Overall, online learning entails high degrees of initiation, organization, and regulation of studying by the students, and this self-regulation is the focus of online learning research (Artino, 2008).

Among various self-regulated learning strategies, multiple studies have empirically shown the importance of time-management skills in online learning success (Lee, 2002; Puzziferro, 2008; Song, Singleton, Hill, & Koh, 2004). Lee (2002) claimed that online learning environments are different from traditional learning environments and that learning strategies should differ according to the learning context. This researcher also suggested and delineated 11 online learning strategies, including self-directed learning, clear and active communication, the management of concurrent discussions, sociality in online learning, the management of information overload, information processing strategies, time management, information interpretation skills, the management of asynchronous tasks, self-efficacy in completing online course, and a positive attitude toward online courses. Additionally, this researcher investigated the relationship between online learning strategies and academic achievement and identified time-management skills as the most dominant predictor of achievement, followed by self-efficacy in completing online courses and a positive attitude toward online courses.

Furthermore, studies on online learning have frequently demonstrated that students in online courses predominantly access the learning materials immediately before an exam and that they do not often complete assignments on time and have an urge to drop out over time (Elvers et al., 2003; Levy & Ramin, 2012; Michinov et al., 2011). In general, students participating in online learning exhibit a lack of time management regarding self-regulated learning, for example, by cramming and procrastinating. Therefore, indicators that reflect regular study and time-management-related behaviors should be given more attention and be further investigated.

2.2. Examining trace data in LMSs and learning analytics

Several researchers have claimed that, although theories of self-regulated learning have been finely developed, few instruments adequately capture students' self-regulation (Hadwin et al., 2007; Pintrich, Wolters, & Baxter, 2000). Most studies on self-regulated learning have used self-report instruments, which not only are intrusive but also are limited to capturing actual self-regulated behaviors in learning contexts. However, as mentioned earlier, this issue can be resolved via LMS use, and such technologically mediated learning environments enable the collection of a comprehensive set of student learning behaviors that occur in learning environments (Pardo, 2014).

The typical data collected from online learning environments are text communication records, server log files, and LMS log files, but LMS log files are recognized as the most practical data source in terms of the level of information and the time and labor intensity of coding. LMS log files still require some work regarding data handling, but they are relatively easy to manage and contain large amounts of information regarding the frequency, patterns, and sessions of actual learning activities (Black, Dawson, & Priem, 2008). Furthermore, well-known LMSs, such as Blackboard and Moodle, provide analytic functions or summarized reports to instructors, and tracing usage data from LMSs is a tangible method of capturing students' self-regulated behaviors.

As considerable student data have become available in the education field, the attention to utilizing student data to improve academic success has increased. The use of analytic techniques in learning is called learning analytics, and learning analytics is defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs," according to the 1st Conference on Learning Analytics and Knowledge (Siemens, 2010). The key objectives of employing learning analytics involve identifying at-risk students by predicting student learning success, providing adequate interventions, and improving learning outcomes (Campbell, DeBlois, & Oblinger, 2007; Dawson, Gašević, Siemens, & Joksimovic, 2014).

Developed at Purdue University, Course Signals is a well-known application for student data analysis and modeling. The primary function of Course Signals is to evaluate students' performances during a course and to categorize students into the following three tiers in terms of their risk of failure: high risk, moderate risk, and no risk. The initial model of

Course Signals was based on four components: demographic characteristics, previous academic history, LMS usage data during a course, and performance in the course to date (Campbell, 2007; Clow, 2013). Depending on the risk level, red, orange, or green lights are presented in the form of traffic signals, and feedback is sent to the students accordingly. Research has empirically shown that implementing Course Signals helps reduce dropout among high-risk students and improves their academic achievement (Purdue University, 2013). Similarly, other studies have reported the benefits of utilizing learning analytics in terms of retention and the prevention of academic failure (Campbell et al., 2007; Dietz-Uhler & Hurn, 2013; Jayaprakash, Moody, Lauría, Regan, & Baron, 2014).

Although learning analytics appears very promising, some challenges must be overcome to enhance learning and teaching practices in the field. Because learning analytics involves sophisticated data mining, it heavily depends on statistical methods and numbers rather than learning theory and processes. Several researchers have argued that data should be interpreted from learners' perspectives (Ferguson, 2012), and the results should provide actionable recommendations (Gašević et al., 2015). Similarly, Gašević et al. (2015) have suggested that feedback after identifying at-risk students should be given to the students, and this feedback should include instructive messages and process feedback rather than low-level system-generated feedback. Therefore, the practice and interpretation of learning analytics should be aligned with a theoretical learning framework. Moreover, the learning analytics tools and functions should be easy for educators to understand and use. Specifically, instructors are the ones who are most curious about their students' progress and level of understanding, and simple and intuitive presentations of students' performances are required to convince instructors to accept and use learning analytics. Therefore, to build effective prediction models and suggest practical implications, the leading factors of students' learning behaviors in relation to performance should be identified and validated in various contexts.

2.3. Relationships between LMS data measures and achievement

2.3.1. Overall participation measures and achievement

Utilizing LMS data allows the identification of, for example, how students' behaviors change over a course, how students interact with peers, and how they perform on their quizzes and exams. Several studies that utilized LMS data have shown that participation indicators and patterns are strongly correlated with academic achievement (Asarta & Schmidt, 2013; Goldstein & Katz, 2005; Michinov et al., 2011; Rafaeli & Ravid, 1997) and engagement (Beck, 2004; Qu & Johnson, 2005). For example, Morris et al. (2005) examined log data regarding student participation to determine whether the level of student participation could differentiate between withdrawers and completers or between successful completers and non-successful completers of an online course. These authors found that among the multiple types of participation log data, the number of discussion posts viewed, the number of content pages viewed, and the time spent viewing discussion pages significantly predicted the students' final grades. These three variables explained 31% of the variability in course achievement. Furthermore, these authors concluded that the level of participation differed significantly between withdrawers and completers and between successful completers and non-successful completers. These results imply that active participation is essential to successful online learning.

Similarly, Campbell et al. (2006) examined the predictive value of LMS usage data on academic performance. These authors found that the addition of the ‘frequency of LMS logins’ variable to the regression model with students' SAT scores increased the power of prediction nearly three-fold compared with the regression model with SAT scores only. Moreover, these authors found that the students with low to moderate SAT scores could achieve good course grades if they exerted an above-average level of effort, as measured by the frequency of LMS logins. These findings convey an important message: students' active participation in learning is a more critical determinant than their ability. Other research has also reported a positive correlation between the total volume of usage data and performance (Johnson, 2005; Wang & Newlin, 2002).

2.3.2. Elaborated timing and access pattern measures and achievement

Despite the prominence of frequency measures in online learning, a few studies have focused on the quality rather than the quantity of online participation (Asarta & Schmidt, 2013; Baugher, Varanelli, & Weisbord, 2003; Chong, 1998). Asarta and Schmidt (2013) were particularly interested in access patterns in terms of 36 online lesson materials; they examined the effects of the timing, volume, intensity, and consistency of access on achievement. These authors defined and used unique time-based measures, such as pacing, anti-cramming, reviewing, completeness and consistency, in addition to conventional frequency measures. Pacing indicated whether a student kept up with the predetermined learning schedule as the course proceeded. Anti-cramming concerned whether a student avoided delaying his/her initial access of the materials, such that he/she would not cram shortly before an exam. Reviewing indicated whether a student encountered a lesson unit without delay and whether the student accessed the content again for review in advance of an exam. Completeness measured the total count of lesson units that a student accessed. Consistency measured whether a student regularly accessed lesson materials regardless of the lesson schedule, and consistency was determined in terms of the intervals of access. This study concluded that the overall frequency of access exhibited the weakest correlation with course achievement and that pacing, anti-cramming, and consistency were significant predictors of course achievement after controlling for grade point average and math skills. Notably, anti-cramming was the most significant factor. These findings clarified that keeping pace with the class schedule, studying the materials in advance of an exam without cramming, and accessing course materials regularly are vital factors for achievement. These findings support the notion that the quality of learning behaviors rather than simply the frequency of access should be taken into account.

Some other studies also support the importance of regular and timely study behaviors in online courses (Jo & Kim, 2013; You, 2015). You (2015) investigated the detrimental effects of academic procrastination on academic achievement using two observed procrastinating behaviors in LMS data. In this study, the students' degrees of failure in completing weekly assigned learning materials as scheduled and their delays in submitting assignments were deemed academic procrastination. The results revealed that academic procrastination negatively predicted course achievement, and a regression model with these two academic procrastination indicators explained 59.7% of the variability in course achievement in this study. These findings support the notion that consistent and regular study behaviors, which are related to time-management strategies, should be considered.

2.4. Research questions

The purpose of this study is to examine the significant behavioral indicators of learning using LMS data that may predict course achievement. Because LMS data generally present information about the frequency, patterns, and duration of actual learning behaviors, this study includes the frequency measures, the duration of study time, and other elaborated time measures that are indicative of self-regulated learning. Additionally, this study examines the potential use of early prediction based on student data collected in the middle of the course.

1. Which indicators from LMS data significantly predict online course achievement?
2. Do significant indicators (i.e., those from question 1) and performance data collected in the middle of the course significantly predict online course achievement?

3. Methods

3.1. Participants and context

The data used in the current study were collected at a mid-sized, four-year university near Seoul, South Korea. Although the university is a campus-based university, it provides several e-learning courses every semester. Among the e-learning courses available, one elective course that was open to a large number of students was chosen: "Introduction to Color". The course did not require any prerequisite, and 575 undergraduate students were registered. Twenty-five students who officially withdrew from the course within four weeks and another 20 students who did not complete the course were removed from the study sample. Particularly the LMS records of the 20 students who were removed from the course showed that they rarely accessed the course and did not take any exams or complete assignments; thus, these students were nearly equivalent to the withdrawers. The data from the remaining 530 students were used in this study.

The course comprised 13 online instruction units and two offline examinations. Instructional videos were the primary course materials, and the instructor uploaded other supplementary resources to the bulletin board. At the beginning of each week, a new instructional video was uploaded for the students to study, and the students were required to watch the instructional video during the assigned week. Four assignments with due dates were given at the beginning of the course. The students were able to ask questions via a Q & A board and email and to communicate with their instructor and peers via a discussion board. All of the learning activities and communication occurred online, with the exception of two examinations that were administered at an offline campus.

Among the participants (see Table 1), 251 were male (47.4%), and 279 were female (52.6%). Most participants were seniors ($n = 241$, 45.5%), and the remaining students included 126 juniors (23.8%), 109 sophomores (20.6%), and 54 freshmen (10.2%). Their majors varied and included computer science, business, physical education, accounting, administration, architecture, and others. The average age was 22.5 ($SD = 2.56$).

3.2. Data collection and measures

The data used in this study were collected from the LMS, with the exception of the two exam scores and the final course score. The midterm exam was administered at week 8, and the final exam was administered at week 15 on campus. The measures from the LMS data used in this study were regular study, total viewing time, sessions, the number of late submissions, proof of reading the course information packets (CIP), and the messages created. The measures from the LMS data were calculated at two different time points. The first included the data accumulated from weeks 1 to 8, and the second included the data accumulated from weeks 1 to 15. The LMS data collected at week 8 along with the midterm scores were used to test the potential of early prediction. Each measure is explained as follows.

First, two scores were used for course achievement. The course exam score was calculated by adding the midterm and final exam scores (which ranged from 0 to 60), and the final course score was also used (ranged from 0 to 100). Course grading was based on virtual attendance, the scores on the two examinations, and the four assignments. Because the final course score reflected attendance, the study used an accumulated exam score to avoid the confounding relationship between attendance (see the regular study section) and the final course score.

Second, regular study was measured by the virtual attendance score. The participating school set an attendance rule for all online courses. Because the instructional videos were the main course materials, the time point when the student first encountered the material and the length of time that the student watched the video were traced. There were 13 weeks of course material, and 0, .5, or 1 point was given per week as a weekly attendance score (ranging from 0 to 13, excluding the two offline exam weeks). No point was given if the student did not access the instructional video within the scheduled time period, and a half point was given when the students accessed the content but did not finish watching the video. One point was given a student watched the weekly assigned videos from beginning to end within the scheduled week. For example, if a student accessed the assigned instructional video after the scheduled time period, no attendance credit was given for that week. If a student accessed the materials within the scheduled time but the timespan of watching the video did not exceed the video length, he/she earned only half of a point. Therefore, the attendance score was used as an elaborated time-based measure to determine whether the students consistently accessed the course material without delay and completed the assigned learning content. In other words, the attendance data used in this study indicated the degree of regular study by integrating pacing, consistency, and completeness. The cumulative attendance score for weeks 1 to 8 was used as regular study $_{time\,=\,1}$, and the cumulative attendance score for weeks 1 to 15 was used as regular study $_{time\,=\,2}$.

Total viewing time was measured by the total time spent on watching 13 weekly instructional videos. The participants were unable to access upcoming instructional videos, but they were able to review past lessons as many times as they wanted. Although attendance was considered an absence if a student accessed a video after the specific time, the total viewing time refers to the cumulative time spent on the videos, regardless of the access time. The sum of the lengths of 13 weekly instructional videos was 1220 min. The total viewing time from weeks 1 to 8 was calculated and termed total viewing time $_{time\,=\,1}$, and the total viewing time from weeks 1 to 15 was calculated and termed total viewing time $_{time\,=\,2}$.

The 'sessions' variable refers to the frequency of a student's access of the online course. That is, this variable simply counts the number of course logins. The session data were collected twice from weeks 1 to 8 (sessions $_{time\,=\,1}$) and from weeks 1 to 15 (sessions $_{time\,=\,2}$).

In this study, late submission indicates the student's failure to submit assignments on time. All of the assignments and the submission schedule were fixed and announced at the beginning of the course. Four assignments were collected at weeks 4, 8, 12, and 15. When the due date for each assignment was approaching, the instructor posted a message on the bulletin board and sent text messages to the students' mobile phones to remind them about the assignment. When a student uploaded an assignment file to the assignment folder, a time stamp was generated in the LMS. One point was assigned for every late submission. Late submission scores were collected twice—at the end of week 8 (late submissions $_{time\,=\,1}$, ranging from 0 to 2) and at the end of week 15 (late submissions $_{time\,=\,2}$, ranging from 0 to 4).

Proof of reading the course information packets was measured by whether a student downloaded and read the course information.

Table 1
Demographic information ($n = 530$).

Variables	Frequency	%
Gender		
Male	251	47.4
Female	279	52.6
Year of study		
Freshman	54	10.2
Sophomore	109	20.6
Junior	126	23.8
Senior	241	45.5
Age		
18–20	117	22.1
21–23	239	45.1
24–26	152	28.7
Over 27	22	4.2

Although the instructor, who had previously taught the course, posted important course regulations and information, including the attendance rules, the assignment schedule, the weekly class schedule, and the evaluation criteria, some students either did not read the posts at all or did not read them thoroughly. Therefore, the instructor created a document that included all important course information, uploaded this document to the bulletin board, and advised the students to reply to the posting to prove that they had thoroughly read and understood the course information. This task was intended to encourage the students to give the important course information extra attention, but it was neither mandatory nor part of an assignment. The students who replied to the post within four weeks were coded as 1, and the others were coded as 0.

Messages created refer to the total number of postings created in the Q & A section, the discussion board, and any emails sent by an individual student within the course. The score for messages created was calculated twice, namely, for week 1 to week 8 (messages created $_{time\ =\ 1}$) and for week 1 to week 15 (messages created $_{time\ =\ 2}$).

3.3. Data analysis

To investigate the research questions, descriptive statistics and correlation analyses were performed, and hierarchical regression analyses were chosen as the statistical analysis methods to identify significant LMS measures with controlling the year of study. The variance inflation factor (VIF) and tolerance among the predictors were checked to determine whether any multicollinearity issues existed. A significance level of .05 was used for the hypothesis testing.

4. Results

4.1. Descriptive statistics and correlation analysis

The descriptive statistics for the study variables are presented in Table 2. The high mean of the regular study variable indicated that the participants visited and studied on a regular basis, but their total viewing times varied widely. However, the small number of messages created indicated that only a few students posted or emailed their questions. The class was very large, and it was designed more for individual learning than for collaborative learning. The results of the correlation analyses of the study variables are presented in Table 3.

4.2. Hierarchical regression analyses of course achievement

First, the student's year of study was entered as a control variable in Step 1, and regular study, total viewing time, sessions, late submissions, proof of reading the CIP, and messages created were entered in Step 2 to predict the final course score. A stepwise method was used, and four variables were revealed to be significant. As presented in Table 4, after controlling the year of study, regular study $_{time\ =\ 2}$ ($B = 3.47$, $p < .001$), late submissions $_{time\ =\ 2}$ ($B = -5.11$, $p < .001$), sessions $_{time\ =\ 2}$ ($B = .13$, $p < .001$), and proof of reading the CIP ($B = 3.97$, $p < .01$) were significant in predicting the final course score. The regression model explained 58.1% of the variance in the final course score ($R^2 = .581$, $F\,(5, 524) = 145.54$, $p < .001$).

Another hierarchical regression analysis was conducted on the course exam score. As presented in Table 5, after controlling the year of study, regular study $_{time\ =\ 2}$ ($B = 2.41$, $p < .001$), sessions $_{time\ =\ 2}$ ($B = .12$, $p < .001$), late submissions $_{time\ =\ 2}$ ($B = -2.03$, $p < .001$), and proof of reading the CIP ($B = 3.53$, $p < .01$) were significant in predicting the course exam score. The regression model accounted for 37.1% of the variance in the course exam score ($R^2 = .371$, $F\,(5, 524) = 61.73$, $p < .001$).

4.3. Hierarchical regression analysis of the early prediction of the final course score

To address the second research question, the four significant variables in the data collected at time 1 and the midterm exam score were selected as predictors, and a regression analysis was conducted to examine the predictability of the final course score. These data were collected in the middle of the course to determine whether this approach was suitable for early prediction. After controlling the year of study, regular study $_{time\ =\ 1}$, sessions $_{time\ =\ 1}$, late submissions $_{time\ =\ 1}$, and proof of reading the CIP were entered in step 2, along with the midterm exam score. The results (see Table 6) revealed that the regression model predicted 69.3% of the variance in the final course score ($R^2 = .693$, $F\,(6, 523) = 196.46$, $p < .001$), and all of the variables, with the exception of proof of reading the CIP ($B = 1.67$, $p = .10$), were significant.

5. Discussion and conclusions

The present study was conducted to investigate the significant behavioral indicators of learning in LMS data and their effects on course achievement. Because self-regulated learning is essential to online learning, measures that reflect the degree of self-regulation were specifically used. Regular study, total viewing time, sessions, late submissions, proof of reading the course information packets, and messages created were chosen as predictors. The results revealed that regular study was the strongest predictor of course achievement, followed by late submissions, sessions, and proof of reading the course information packets. After controlling the year of study, four indicators from the LMS data significantly predicted both the course exam score and the final course score. The regression model with four variables explained 58.1% of the variance in the final course score, and the predictability of the regression model appeared to be relatively high compared with results from previous research (e.g., 33% in Macfadyen & Dawson, 2010, and 31% in Morris et al., 2005). However, the total viewing time and messages created were not significant. In summary, these results clearly indicated the importance of self-regulated learning, particularly regular study in accordance with the course schedule, the timely completion of assigned tasks, frequent accessing of course materials, and the reading of important course information. These findings support those of previous research (Asarta & Schmidt, 2013; Jo & Kim, 2013; You, 2015), which has emphasized the quality of learning behaviors rather than the

Table 2
Descriptive statistics of the study variables ($n = 530$).

Variables	Weeks 1 to 8 (*time* 1)			Weeks 1 to 15 (*time* 2)		
	Mean	SD	Min.–max.	Mean	SD	Min.–max.
1. Course exam score	–	–	–	48.42	8.23	10.80–60.00
2. Final course score	–	–	–	79.63	10.31	24.00–93.00
3. Regular study	6.75	.53	3–7	12.25	1.18	3–13
4. Total viewing time	1257.77	708.54	395–5380	2410.08	1236.19	395–7909
5. Sessions	19.64	5.91	4–46	41.27	11.82	10–93
6. Late submissions	.13	.38	0–2	.36	.73	0–4
7. Proof of reading the CIP	.92	.26	0–1	–	–	–
8. Messages created	.13	.39	0–3	.27	.72	0 ~ 8

Table 3
Correlation analyses (n = 530).

Variables	1	2	3	4	5	6	7	8
1. Exam score	1							
2. Final course score	.97**	1						
3. Regular study$_{time\,=\,2}$	.54**	.65**	1					
4. Total viewing time$_{time\,=\,2}$	.00	.00	-.04	1				
5. Sessions$_{time\,=\,2}$	.41**	.45**	.41**	.25**	1			
6. Late submissions$_{time\,=\,2}$	−.41**	−.60**	−.45**	.04	−.30**	1		
7. Proof of reading the CIP[a]	.27**	.29**	.28**	−.02	.21**	−.11**	1	
8. Messages created$_{time\,=\,2}$	.08	.08	.01	.11*	.25**	−.04	.07	
9. Year of study[b]	.16**	.18**	.13**	−.06	.10*	−.19**	.06	−.04

* $p < .05$.
** $p < .01$.
a Proof of reading the CIP was analyzed by using point biserial correlation.
b Control variable.

quantity of learning. Furthermore, it is worth noting that the year of study, the control variable, was significant in the earlier model, but it was not significant after regular study and late submissions were included in the model. This result also supports the idea that self-regulation is critical for course success.

Additionally, the present study focused on examining the potential for early prediction during the course. The data regarding regular study, sessions, late submissions, proof of reading the course information packets, and the midterm exam score that were collected in the middle of the course were used for the early prediction of students' final course achievement. The regression analysis results revealed that the four indicators from the LMS data and the midterm exam score significantly predicted final course achievement. These findings imply that use students' performance data during the course will help instructors predict final course achievement and provide proactive feedback and adequate interventions to students. The results confirm the benefits of tracing and analyzing LMS data during courses.

The findings of this study have several implications. First, this study contributes to the identification of unconventional but more relevant self-regulated learning measures from LMS log data and their effectiveness. Although LMS log data provide a lengthy list of variables, including total visits, the number of discussion messages posted, the number of files viewed, the number of messages sent, the total time spent on online learning, and the number of assignments submitted, Huang and Fang (2013) claimed that merely adding more variables does not improve the predictability of mathematical models. Therefore, researchers and educational practitioners need to continue to identify or develop significant indicators that effectively capture learners' engagement and self-regulated learning.

In this study, the ‘regular study’ variable was considered an elaborated time-based measure and indicated more than the access frequency or time spent viewing. Such elaborated time-based measures can simultaneously indicate a student's access time and study patterns; in addition, they can serve as leading indicators. However, additional indicators that reflect students' levels of regularity, engagement, and motivation in online learning should be explored and tested in this field.

Furthermore, this study intentionally selected indicators from LMS data. Although there were many postings on the bulletin board, reading and understanding the course information was considered fundamental and essential for successful learning. The effect of reading the course information packets in the beginning of the course weakly but significantly affected course achievement. Because each online course has specific grading criteria, assignments, and course organization, students need to carefully review the information and be aware of important events in advance, such that they can regulate their studies and participate in course activities. This result implies that not all clicks are equal. Therefore, discerning the important learning behaviors and weighing the quality of actions will better predict student performance.

Second, these findings empirically prove the significant effects of self-regulated learning in online learning contexts. Although traditional classes have fixed time schedules that involve students attending class regularly, online courses usually do not have a specific schedule for accessing course material, and online students need to manage their time and efforts to keep their studies on track. The present study clearly demonstrates the benefits of regularly accessing course material and

Table 4
Hierarchical regression analysis results on final course score (n = 530).

Model	Predictors	Final course score			
		B	SE	ß	R^2
M0	Year of study[a]	1.78	.43	.18***	.032
M1	Year of study[a]	.97	.33	.10*	.437
	Regular study $_{time\,=\,2}$	5.62	.29	.64***	
M2	Year of study[a]	.46	.30	.05	.550
	Regular study $_{time\,=\,2}$	4.18	.29	.48***	
	Late submissions $_{time\,=\,2}$	−5.39	.47	−.38***	
M3	Year of study[a]	.42	.29	.04	.572
	Regular study $_{time\,=\,2}$	3.69	.30	.42***	
	Late submissions $_{time\,=\,2}$	−5.06	.46	−.36***	
	Sessions $_{time\,=\,2}$	.14	.03	.16***	
M4	Year of study[a]	.39	.29	.04	.581
	Regular study $_{time\,=\,2}$	3.47	.30	.40***	
	Late submissions $_{time\,=\,2}$	−5.11	.46	−.36***	
	Sessions $_{time\,=\,2}$	.13	.03	.15***	
	Proof of reading the CIP	3.97	1.16	.10**	

* $p < .05$.
** $p < .01$.
*** $p < .001$.
a Year of study: control variable.

Table 5
Hierarchical multiple regression analysis results on course exam score (n = 530).*

Model	Predictors	Course exam score			
		B	SE	ß	R^2
M0	Year of study[a]	1.25	.34	.16***	.025
M1	Year of study[a]	.71	.29	.09*	.297
	Regular study $_{time\,=\,2}$	3.67	.26	.53***	
M2	Year of study[a]	.63	.29	.08	.335
	Regular study $_{time\,=\,2}$	3.06	.27	.44***	
	Sessions $_{time\,=\,2}$	.15	.03	.22***	
M3	Year of study[a]	.46	.28	.06	.359
	Regular study $_{time\,=\,2}$	2.60	.29	.37***	
	Sessions $_{time\,=\,2}$	.13	.03	.19***	
	Late submissions $_{time\,=\,2}$	−1.99	.45	−.18***	
M4	Year of study[a]	.43	.28	.05	.371
	Regular study $_{time\,=\,2}$	2.41	.29	.35***	
	Sessions $_{time\,=\,2}$	.12	.03	.18***	
	Late submissions $_{time\,=\,2}$	−2.03	.45	−.18***	
	Proof of reading the CIP	3.53	1.13	.11**	

* $p < .05$.
** $p < .01$.
*** $p < .001$.
a Year of study: control variable.

Table 6
Hierarchical regression analysis on final course score ($n = 530$).

Predictors	Final course score			
	B	*SE*	ß	R^2
Step 1				
Year of study[a]	1.78	.43	.18***	.032
Step 2				
Year of study[a]	.795	.24	.08**	.693
Regular study $_{time\ =\ 1}$	3.47	.51	.18***	
Sessions $_{time\ =\ 1}$	.28	.05	.16***	
Late submissions $_{time\ =\ 1}$	−3.56	.66	−.13***	
Proof of reading the CIP	1.67	1.01	.04	
Midterm exam score	1.50	.06	.63***	

* $p < .05$.
** $p < .01$.
*** $p < .001$.
[a] Year of study: control variable.

keeping pace with the learning schedule. Furthermore, late submissions, which directly captured students' time-management skills, were revealed to be a significant indicator. The frequency of assessing the course materials was still identified as an important indicator in the present study. Because the instructors continued to post notices and upload extra materials throughout the course, frequent visit to the online course enabled students to keep up and engage in their learning.

Lastly, these findings support the potential of early prediction via the analysis of LMS data. A number of LMS developers have endeavored to develop more intuitive and informative dashboards for instructors, such that the instructors are able to monitor students' progress easily and to acquire insights from this information. This function is especially useful when the course size is large. For example, in 2011, a MOOC course on artificial intelligence at Stanford University registered 160,000 students. Such large-scale classes obviously present issues in relation to monitoring students' learning progress, identifying their levels of understanding, and providing them feedback (Macfadyen & Dawson, 2010). Therefore, the extraction and aggregation of meaningful indicators from the students' behavioral data and the development of reliable prediction models are valuable for instructors who seek to understand their students' learning statuses and to provide actionable feedback during the course.

6. Limitations and future research

Although the present study demonstrates the benefits of identifying significant measures from LMS data to facilitate successful online learning, several limitations should be noted. First, the online course in this study was a very large class that was designed for individual learning. Although the messages created did not significantly affect course achievement in the present study, these results could be different if the discussions were a central part of the instruction. Future studies should identify not only significant but also robust LMS measures that reflect engaged learning in various contexts. Second, although the year of study was used to control the results, other control variables, such as online learning experience, might be considered. Because the research context of the present study was a traditional university and online learning has become more conventional, students are assumed to have similar levels of online learning experience. However, the use of significant background information and student records is suggested to increase the predictability of students' performances. Finally, the relationships between LMS measures and course achievement that were obtained in this study were based on correlations, not necessarily causation. As such, these results should be cautiously interpreted.

References

Artino, A.R. (2008). Promoting academic motivation and self-regulation: Practical guidelines for online instructors. *TechTrends*, *52*(3), 37–45.

Asarta, C.J., & Schmidt, J.R. (2013). Access patterns of online materials in a blended course. *Decision Sciences Journal of Innovative Education*, *11*(1), 107–123.

Baugher, D., Varanelli, A., & Weisbord, E. (2003). Student hits in an internet-supported course: How can instructors use them and what do they mean? *Decision Sciences Journal of Innovative Education*, *1*, 159–179.

Beck, J.E. (2004). Using response times to model student disengagement. *Proceedings of the workshop on social and emotional intelligence in learning environment, 7th international conference on intelligent tutoring systems, Maceio, Brazil* (pp. 13–20).

Black, E.W., Dawson, K., & Priem, J. (2008). Data for free: using LMS activity logs to measure community in online courses. *The Internet and Higher Education*, *11*, 65–70.

Campbell, J. P. (2007). Utilizing student data within the course management systems to determine undergraduate student academic success: An exploratory study. Unpublished doctoral dissertation, Purdue University.

Campbell, J.P., Finnegan, C., & Collins, B. (2006). Academic analytics: Using the CMS as an early warning system. *Paper presented at the WebCT impact conference 2006, Chicago, IL*.

Campbell, J.P., DeBlois, P.B., & Oblinger, D.G. (2007). Academic analytics: A new tool for a new era. *Educause Review*, *42*(4), 40–42.

Chong, S. (1998). Models of asynchronous computer conferencing for collaborative learning in large college classes. In C.J. Bonk, & K.S. King (Eds.), *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship and discourse* (pp. 157–182). Mahwah, NJ: Lawrence Erlbaum Associates.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, *18*(6), 683–695.

Dabbagh, N., & Kitsantas, A. (2004). Supporting self-regulation in student-centered web-based learning environments. *International Journal on E-Learning*, *3*, 40–47.

Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014). Current state and future trends: a citation network analysis of the learning analytics field. *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 231–240). New Work, NY, USA: ACM.

Dietz-Uhler, B., & Hurn, J.E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning*, *12*(1), 17–26.

Driscoll, M.P. (2000). *Psychology of learning for instruction* (2nd ed.). Needham Heights, MA: Allyn & Bacon.

Elvers, G.C., Polzella, D.J., & Graetz, K. (2003). Procrastination in online courses: Performance and attitudinal differences. *Teaching of Psychology*, *30*, 159–162.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *The International Journal of Technology Enhanced Learning*, *4*(5/6), 304–317.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71.

Goldstein, P.J., & Katz, R.N. (2005). *Academic analytics: The uses of management information and technology in higher education*. Washington, DC: EDUCAUSE Center for Applied Research.

Hadwin, A.F., Nesbit, J.C., Jamieson-Noel, D., Code, J., & Winne, P.H. (2007). Examining trace data to explore self-regulated learning. *Metacognition Learning*, *2*, 107–124.

Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education*, *61*, 133–145.

Jayaprakash, S.M., Moody, E.W., Lauría, E.J.M., Regan, J.R., & Baron, J.D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Jo, I., & Kim, Y. (2013). Impact of learner's time management strategies on achievement in an e-learning environment: A learning analytics approach. *Journal of Educational Information and Media*, *19*(1), 83–107.

Johnson, G. (2005). Student alienation, academic achievement, and WebCT use. *Educational Technology & Society*, *8*(2), 179–189.

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2014). *NMC horizon report: 2014 higher* (education edition). Austin, Texas: The New Media Consortium.

Jonassen, D., Davidson, M., Collins, M., Campbell, J., & Haag, B.B. (1995). Constructivism and computer-mediated communication in distance education. *The American Journal of Distance Education*, *9*(2), 7–26.

Joo, Y.J., Joung, S., & Kim, J. (2014). Structural relationships among self-regulated learning, learning flow, satisfaction, and learning persistence in cyber universities. *Interactive Learning Environments*, *22*(6), 752–770.

King, F.B., Harner, M., & Brown, S.W. (2000). Self-regulatory behavior influences in distance learning. *International Journal of Instructional Media*, *27*(2), 147–156.

Klingsieck, K.B., Fries, S., Horz, C., & Hofer, M. (2012). Procrastination in a distance university setting. *Distance Education*, *33*(3), 295–310.

Lee, I. (2002). Relationships between e-learning strategies and learning achievement. *Journal of Educational Technology*, *18*(2), 51–67.

Levy, Y., & Ramin, M.M. (2012). A study of online exams procrastination using data analytics techniques. *Interdisciplinary Journal of E-Learning and Learning Objects*, *8*, 97–113.

Macfadyen, L.P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*, 588–599.

Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of Educational Psychology*, *106*(1), 121–131.

Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers & Education*, *56*, 243–252.

Misanchuk, E.R., & Schwier, R.A. (1992). Representing interactive multimedia and hypermedia audit trials. *Journal of Educational Multimedia and Hypermedia*, *1*, 355–372.

Morris, L.V., Finnegan, C.L., & Wu, S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, *8*, 221–231.

Pardo, A. (2014). Designing learning analytics experiences. In J.A. Larusson, & B. White (Eds.), *Learning analytics: From research to practice* (pp. 15–38). New York, NY: Springer Science.

Pintrich, P.R., & De Groot, E.V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40.

Pintrich, P.R., & Zusho, A. (2002). The development of academic self-regulation: The role of cognitive and motivational factors. In A. Wigfield, & J.S. Eccles (Eds.), *Development of achievement motivation* (pp. 249–284). San Diego, CA: Academic.

Pintrich, P.R., Wolters, C.A., & Baxter, G.P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw, & J.C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Buros Institute of Mental Measurement: Lincoln, NE.

Purdue University (2013). Course signals. (Retrieved from) http://www.itap.purdue.edu/learning/tools/signals/

Puzziferro, M. (2008). Online technologies self-efficacy and self-regulated learning as predictors of final grade and satisfaction in college-level online courses. *American Journal of Distance Education*, *22*(2), 72–89.

Qu, I., & Johnson, W.I. (2005). Detecting the learner's motivational states in an interactive learning environment. In C. -K. Looi, G.I. McCalla, B. Bredeweg, & J. Breuker (Eds.), *Proceedings of the 12th international conference on artificial intelligence in education: Supporting learning through intelligent and socially informed technology* (pp. 547–554) (Amsterdam, The Netherlands).

Rafaeli, S., & Ravid, G. (1997). Online, web-based learning environment for an information systems course: Access logs, linearity and performance. *Proceedings of information systems educators conferences, USA. 97*. (pp. 92–99).

Rakes, G.C., & Dunn, K.E. (2010). The impact of online graduate students' motivation and self-regulation on academic procrastination. *Journal of Interactive Online Learning*, *9*(1), 78–93.

Reyes, J.A. (2015). The skinny on big data in education: Learning analytics simplified. *TechTrends*, *59*(2), 75–80.

Siemens, G. (2010, July 22). About: 1st international conference on learning analytics and knowledge. (Retrieved from) https://tekri.athabascau.ca/analytics/about

Song, L., Singleton, E.S., Hill, J.R., & Koh, M.H. (2004). Improving online learning: Student perceptions of useful and challenging characteristics. *The Internet and Higher Education*, *7*(1), 59–70.

Sun, P.C., Tsai, R.J., Finger, G., Chen, Y.Y., & Yeh, D. (2008). What drives a successful e-learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education*, *50*(4), 1183–1202.

Tuckman, B.W. (2005). Relations of academic procrastination, rationalizations, and performance in a web course with deadlines. *Psychological Reports*, *96*, 1015–1021.

Wang, A.Y., & Newlin, M.H. (2002). Predictors of web-student performance: The role of self-efficacy and reasons for taking an on-line class. *Computers in Human Behavior*, *18*, 151–163.

Warnock, S., Bingham, K., Driscoll, D., Fromal, J., & Rouse, N. (2012). Early participation in asynchronous writing environments and course success. *Journal of Asynchronous Learning Networks*, *16*(1), 35–47.

Wolters, C.A. (2003). Understanding procrastination from self-regulated learning perspective. *Journal of Educational Psychology*, *95*(1), 179–187.

You, J.W. (2015). Examining the effect of academic procrastination on achievement using LMS data in e-learning. *Educational Technology & Society*, *18*(3), 124–134.

You, J.W., & Kang, M. (2014). The role of academic emotions in the relationship between perceived academic control and self-regulated learning in online learning. *Computers & Education*, *77*, 125–133.

Yukselturk, E., & Bulut, S. (2007). Predictors for student success in an online course. *Educational Technology & Society*, *10*(2), 71–83.

Zimmerman, B., & Martinez-Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology*, *80*, 284–290.