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