Internet and Higher Education 27 (2015) 44–53

Contents lists available at ScienceDirect

Internet and Higher Education

A multivariate approach to predicting student outcomes in web-enabled blended learning courses

Nick Z. Zacharis

Technological Educational Institute of Piraeus, Department of Computer Systems Engineering, Building B - Office DA5, Petrou Ralli & Thivon 250, 12244, Greece

ARTICLE INFO

Article history:
Accepted 12 May 2015
Available online 20 May 2015

Keywords:
Learning analytics
Blended learning

ABSTRACT

This study aimed to develop a practical model for predicting students at risk of performing poorly in blended learning courses. Previous research suggests that analyzing usage data stored in the log files of modern Learning Management Systems (LMSs) would allow teachers to develop timely, evidence-based interventions to support at risk or struggling students. The analysis of students' tracking data from a Moodle LMS-supported blended learning course was the focus of this research in an effort to identify significant correlations between different online activities and course grade. Out of 29 LMS usage variables, 14 were found to be significant and were input in a stepwise multivariate regression which revealed that only four variables – *Reading and posting messages*, *Content creation contribution*, *Quiz efforts* and *Number of files viewed* – predicted 52% of the variance in the final student grade.

1. Introduction

Internet-based information and communication technologies have opened up new potentials in higher education. Most universities today offer some coursework online and some have converted programs of study in order to make them entirely available online (Ward, Peters, & Shelley, 2010). Distance students can now use an assortment of online resources and learn at their own pace, collaborating with other learners rather than working in isolation. By their very nature, web-based learning modules offer the flexibility of self-directed learning and the opportunity to move away from teacher-directed approaches to teaching and learning. There are many instructional design models available to guide course design processes, but alignment among learning goals, assessment strategies and instructional activities is essential to a well-designed, student-centered learning course (Rubin, 2013).

Facilitating such a learning process in the online classroom is particularly challenging for the instructor due to the fact that most communication is asynchronous and lacks many of the emotional cues of the face-to-face environment (Sheridan & Kelly, 2010). Students with inadequate knowledge of the technologies being employed or with poor time management skills may experience delays in getting prompt feedback, feel unmotivated and procrastinate. Reflecting and in response to the students' specific experiences in the online classroom, instructors need to create an environment that encourages student feedback and engage students in intensive and fruitful interactions with the instructor, the material and the other learners (Mahle, 2011). Activities promoting critical thinking, collaborative learning and self-directedness contribute to students' engagement and learning (Ishtaiwa & Abulibdeh, 2012).

Inspired by social constructivist theories of learning and development, online learning environments have evolved to specialize in the provision of networked tools that allow learners to determine and direct their learning activities according to their personal needs and goals. A variety of nearly synchronous or synchronous communication and collaboration tools, combined with media-rich resources and classroom materials, are used to engage students in meaningful learning online. However, although online student enrolment rates are drastically moving upward in all schools which offer online education (Allen & Seaman, 2009), research indicates that students are substantially less likely to complete fully online courses and their dropout rates are higher compared to students who took the same course face-to-face (Jaggars, 2011; Jenkins, 2011; Xu & Jaggars, 2011).

In contrast, online learning platforms enhanced by conventional teaching methodologies that include instructor led meetings and seminars, have been found to be at least as effective, in terms of learning outcomes, as the face-to-face courses (Bowen & Lack, 2012). Although the lack of self-regulatory learning skills remains a serious impediment to the success of learning in the blended learning context, the higher student–instructor interaction results in higher levels of students' motivation, engagement and achievement (Xu & Jaggars, 2011; Zacharis, 2011). Because a blended teaching model shifts instructors' focus away from more traditional curricular and administrative tasks in the direction of working with data and providing more individualized support to students, the analysis and interpretation of tracking data of students' activities online should be a seamless part of a blended learning classroom workflow.

E-mail address: nzach@teipir.gr.

http://dx.doi.org/10.1016/j.iheduc.2015.05.002

Many empirical studies have been carried out so far to examine the relationship between LMS usage and student academic achievement. By analyzing specific subsets extracted from the large datasets stored in the LMS (e.g. discussion messages posted, files viewed, time spent online, total hits), researchers try to identify online activities/variables that provide a predictive view of student performance in order to inform instruction or to determine strategies to help students at risk of course failure (Ashford-Rowe & Malfroy, 2009; Milne, Jeffrey, Suddaby, & Higgins, 2012). However, choosing the right set of variables to build a predictive model that works as an early warning system for instructors is a difficult task. Usually, the researchers have determined beforehand which variables are the most important ones to be plugged into the model and the only thing remaining is to test its predictive power.

In this study, all the pertinent LMS data concerning online activities during the blended learning process were extracted and each activity was treated as the independent variable in a bivariate correlation analysis with student course grade. From a total of 29 potential explanatory variables, 14 variables with significant univariate association with course grade were chosen for inclusion in a multivariate regression analysis. View of resources, quiz engagement, reading and posting of messages (in forum, email and chat) and content creation (in wiki and blog) were the variables that best predicted final grade, explaining over 50% in variability in the data set. These findings support the viewpoint that few LMS-based online learning activities are able to accurately predict educational outcomes in blended learning courses. Therefore, the provision of online tools and resources that promote engagement with content, collaboration with team mates and connectedness with both peers and teachers, should be of high priority during the design and practice of blended learning.

2. Blended learning through LMS

With the prevalence of web applications, online learning has gained increasing popularity over the years and has evolved as a viable and flexible alternative to traditional brick-and-mortar academic approaches. Nowadays, colleges and universities worldwide are using various forms of electronic distance media to transmit educational courses to students without the limitations of location or time. Both synchronous and asynchronous communication and collaboration online tools provide learning opportunities that are flexible and responsive to learners' needs, learning styles and backgrounds. Although research on the comparative effectiveness of online instruction and traditional, face-to-face learning in higher education reveals no significant difference in learning outcomes (U.S. Department of Education, 2010), it is common sense that the higher dropout rates for full-time online students are mostly due to their poor self-regulating skills and a lack of interaction with tutors and other learners.

2.1. Blended learning course design strategies

As higher education institutions adopt learning platforms as part of their educational delivery portfolio, educators seeking to effectively serve individual students' needs and learning styles are turning to blended learning instruction as a way to provide opportunities for self-paced and self-directed learning. Blended learning systems combine face-to-face instruction with computer mediated instruction (Graham, 2006). Blended or hybrid courses integrate online learning with traditional face-to-face class activities in a planned, pedagogically valuable manner that not just supplement, but transform and improve the learning process. Surpassing the conventional role of delivering knowledge, the teachers' new roles include facilitation, student mentoring and differentiating instruction for individual learners. By moving learning resources online, educators can easily track students' progress, freeing up time in the face-to-face meetings to engage students in collaborative work and troubleshoot difficult concepts.

Blended learning courses employ active learning strategies through using a variety of pedagogical approaches (Glazer, 2012), ranging from fully online curriculums with face-to-face interaction to courses in which traditional, face-to-face, classroom instruction is integrated with online components that extend learning beyond the classroom (Eduviews, 2009). How much of the face-to-face instruction must be replaced by online coursework will vary greatly by class, discipline, and learning objectives. A simple, but inefficient to cover every blended learning configuration, guideline defines blended learning as a course where 30%–70% of the instruction is delivered online (Allen & Seaman, 2009).

No matter what the blend of technologies or learning approaches, properly designing and implementing blended learning starts with a re-examination of the intended learning outcomes of the course (Garrison & Vaughan, 2008). Intended learning outcomes should be clearly stated, meaningful and measurable. Content should be divided into suitable learning chunks and presented in various formats to students, taking into account their different backgrounds and diverse ways of learning. Authentic learning activities and assignments should engage students in active learning, promote the achievement of the intended learning outcomes and be aligned with each other (Reaburn, Muldoon, & Bookallil, 2009). The sequence of learning events should involve the right mix of student–content, student–instructor, and student–student interaction (Gradel & Edson, 2010). Regular feedback about student performance should be provided in a timely manner throughout the course (Salmon, 2013).

Searching for the most appropriate model for a particular blended learning setting, Moskal, Dziuban, and Hartman (2013) conclude that "there is no singular best model, and most institutions can achieve success with nearly any of them" (Moskal et al., 2013, p. 16). They argue that mode of delivery in general, and blended learning in particular, has a very weak statistical correlation with student success. Rather, a set of institutional variables (institutional goals and objectives, administrators and faculty members' goals alignment, organizational capacity, faculty development and course development support, support for online students and faculty, robust and reliable infrastructure, longitudinal data collection and assessment, proactive policy development and an effective funding model) has come to be accepted as critical factors for blended learning success (Moskal et al., 2013).

2.2. Using LMS tools to facilitate blended learning

Campuses have adopted LMSs, like Blackboard and Moodle, to facilitate online, onsite and hybrid courses through their functionalities for content creation, communication, assessment and administration (Piña, 2010). Besides centralizing and automating administration tasks, like creating and managing user accounts, creating syllabi and assignments, grading, etc., LMSs assemble and deliver learning content rapidly, personalize content and enable knowledge re-use (Ellis, 2009). In an LMS environment, teachers can create and maintain a learning structure or sequence (Hirumi, 2012), load and replace resource files, control access to resources, organize and support group activities, track activity of learners, customize learning sequences, mark and provide feedback. Based on advanced relational database software such as Oracle, Microsoft SQL, or MySQL, which emphasize data independence, interconnectedness and security, LMSs incorporate a variety of login roles (instructor, student, guest) permitting the instructor to interact privately with one student or create discussion groups and teams with different profiles (Kats, 2010).

Through LMS platforms, students can access learning materials, like documents, spreadsheets PowerPoint presentations, hyperlinks, audio or video of lectures, submit assignments, track their progress and interact with professors and peers. All the social and collaborative aspects of blended learning can be facilitated by both the asynchronous and synchronous LMS tools. Asynchronous (non real-time) tools include email, threaded discussion boards, wikis, blogs, calendars, course announcements and file sharing. Synchronous (real-time) tools found in a typical LMS include text chat, whiteboard, and web-conferencing tools (audio and video communication, online presentations, breakout rooms, application sharing) and might be available only after installing additional plug-ins or integration packs.

By combining LMS tools with different learning approaches, such as self-directed learning on individual computers, small group instruction with intervention teacher or collaborative online content creation, a wide variety of activities can be undertaken individually, in small groups or in one large group: discussion and debate, reflective diary, collaborative writing, peer review, enquiry-based learning, resource-based learning, online note taking and peer comment, knowledge repository on course concepts, pre-module questionnaires, online quizzes, and electronic mind maps (Macdonald, 2008; Nissen & Tea, 2012). Given the ease, convenience and accessibility of online resources available through an LMS, it is not surprising that students' satisfaction with blended learning course delivery is very high (So & Brush, 2008). The integration of human interaction to online learning, the balanced combination of self-paced and team activities and a mix of spoken, written and interactive media, have been proven to be effective in supporting learning for all personality types – visual, auditory and kinesthetic – in diverse blended learning environments (Naveh, Tubin, & Pliskin, 2010; Picciano, 2009; Zacharis, 2010).

Recently, the capture and analysis of large data sets related to students' learning activities, has drawn the attention of academic researchers as a possible solution to challenges, such as retention and learner support (Siemens, 2013). This kind of work has been called learning analytics and its focus is on how students access the information, how they navigate through the materials, how long it takes them to complete activities, and how they interact with the materials to transform the information into measurable learning (Mattingly, Rice, & Berge, 2012). Several, mainly commercial, applications, such as SPSS, Stata, and NVivo and Web analytics tools, such as Google Analytics and Adobe's Digital Marketing Suite, can be used as learning analysis tools to create predictive models of potential at risk students or to develop personalized learning recourses or model the collective learning community behavior (Dawson, Heathcote, & Poole, 2010; Siemens, 2013). Since most blended learning courses today combine in-class activities with the support of an LMS, teachers can easily track student activity by processing the digital trails that every online interaction leaves in the system's log file.

3. Previous work on learning analytics

Web-based learning platforms have the ability to capture large streams of data through which educators and administrators can attempt to improve the learning experience. The collection, analysis and reporting of data about learners' activities and interests, in order to assess academic progress, predict future performance, and spot potential issues, is the core proposition that leads the development of the newly emerging field of learning analytics (Buckingham Shum & Ferguson, 2012; The Horizon Report, 2013). Information derived from learning analytics aims to gain insights about student behavior and learning, inform instructional practice in real time, and help in the design of course management systems that personalize education (The Horizon Report, 2013). Although large stores of data already exist at most colleges and universities that could be used to make data driven decisions able to support optimal use of both economic and pedagogical resources, to date there has been limited application of this data within higher education (Dawson et al., 2010).

3.1. Engagement and frequency of LMS use

In an LMS, all students' logins, time spent on each activity, number of discussion messages posted, downloads, etc., are recorded in the website log file, which can then provide teachers with activity reports that can be further filtered by student name, date, task and action. Once there are large amounts of those records, one of their possible usages is the automatic prediction of students' performance. A number of papers have explored the relationship between the frequency of student LMS usage and academic success. Vengroff and Bourbeau (2006) provide evidence that supplemental online resources benefit undergraduates and conclude that students who consistently used Blackboard did on average better on examinations than students who used it less frequently. Searching for a link between LMS activity and student grades, Dawson, McWilliam, and Tan (2008) found significant differences between low and high performing students in the quantity of online session times, total time online and the amount of active participation in discussion forums. Damianov et al. (2009) used the time spent online as a measure of effort in a multinomial logistic model and found a positive and significant relationship between time spent online and grades, especially for students who obtained grades between D and B.

Besides measuring student engagement and participation by total time spent online, many researchers measure student effort by the frequency of accessing course materials within the LMS. Baugher, Varanelli, and Weisbord (2003) collected data from a web-augmented course and found that hit consistency – a measure of how routinely a student accessed the course LMS – is a better predictor of success than total hits. Chanchary and Haque (2007) examined an LMS activity log of 112 undergraduate students to find relationships between students' LMS access behavior, study habits and overall performances, and concluded that students having low LMS access rate obtained poor grades. Biktimirov and Klassen (2008) also found that hit consistency is a strong predictor of student success. They counted accesses to various content areas within an LMS and concluded that homework solutions accessed was the only activity that significantly correlated with course grade. Beer, Clark, and Jones (2010) analyzed large numbers of records captured from both Blackboard and Moodle LMSs and found a general correlation between the number of student clicks within each LMS and their final grade. Crampton, Ragusa, and Cavanagh (2012) analyzed the web access data from the Sakai LMS and demonstrated that students who accessed the most resources, in terms of diversity and percentage of available resources, achieved higher grades.

3.2. Instructional design and participation

Effective online teaching strategies are heavily dependent upon instructional designs in which each mode of online interaction – student/content, student/teacher, and student/student – has its own positive impact on student achievement. Coldwell, Craig, Paterson, and Mustard (2008) studied the impact of student participation on student performance, using as components of participation the total time spent online, the number of discussion messages read, the number of discussions posted and the number of content files viewed. Data from a fully online information technology course were analyzed, showing a positive relationship between students' participation in the course and their performance as measured by final grades in the course. Dawson et al. (2008) examined the types of LMS tools utilized by students and staff and found that the discussion forum was the dominant tool representing over 80% of all interactions in the online environment. Greenland (2011) compared three instructional designs of varying degrees of synchronous and asynchronous learning activities, and found that asynchronous learning activities correspond with greater levels of interaction. Nandi, Hamilton, Harland, and Warburton (2011) found that the number of posts increases during weeks when assignments and examinations are due and that the assessment marks of high achieving students were correlated with their online participation throughout the semester.

All of the articles listed above have used LMS log files to extract unbiased information from activity and performance data and indentify relationships between different independent variables and final grade. Most of these studies have been based on univariate analyses, focusing on a single variable or a single set of variables that were expected to have an impact on course outcomes. However, academic performance is a difficult area of education to fully understand or measure. Students' learning habits and success rates are complex interrelated phenomena that cannot be attributed to a single variable. The majority of the researchers mentioned above have noted the need for more sophisticated models that can take into account multiple variables that may contribute to student engagement and achievement. Although little information is given about how and why displayed variables were selected for tracking, it is obvious that the researchers included only the variables (tools and actions) that they believed to be most closely correlated with student grades.

3.3. Social learning analytics

Factors affecting academic performance have been the focus of research for many years and it remains an active research topic (Buckingham Shum & Deakin Crick, 2012; Buckingham Shum & Ferguson, 2012), indicating the inherent difficulty in both measurement of learning and modeling the learning process, particularly in tertiary education (Gray et al., 2014). Positive learning dispositions have been shown to contribute towards improved learner engagement by different pathways. Dispositional language describes learning as a combination of learning inclinations, self-regulated skills, patterns of behavior and interaction, motivation (e.g., self-efficacy, goal setting) and cognitive ability (Tempelaar, Rienties, & Giesbers, 2014). Buckingham Shum and Ferguson (2012), proposed a combination of learner data generated by LMSs with self-reported data gathered through surveys in an effort to focus more on learning process details, individual performance and team interactions. These social learning analytics depend on users self-disclosing 'metadata' about themselves and can create meaningful toolkits that support particular kinds of learning, especially in high diversity courses (Buckingham Shum & Ferguson, 2012). This study however, based on the homogeneity (same nationality, language, university entry grades, technical background, age) of the cohort selected as sample, restricted its attention in searching only for factors that predict success objectively and avoided using data from self-reported questionnaires which could be biased or inaccurate.

3.4. Predicting success with multivariate analysis

Although there seem to be widespread agreement on what the purpose of learning analytics is, there are a variety of opinions on what data is important and should be gathered and processed to improve teaching and learning. Agudo-Peregrina, Iglesias-Pradas, Conde-González, and Hernández-García (2014) argue that it is very difficult to identify what is the net contribution of each type of interaction – interactions between students and teachers, interactions among students as well as student–content interactions – to the learning process. Their research showed that student–student interaction had a lower influence than student–teacher interaction on academic performance, a result that appears to contradict previous research (Dawson et al., 2008; Macfadyen & Dawson, 2010; Nandi et al., 2011). Differences in instructional design (fully online or blended learning courses, mandatory or non-mandatory e-learning components, granularity of learning, learning objectives, assessment and feedback loops), kind of data extracted, statistical techniques and predictive modeling used, understanding and interpretation of the patterns observed, etc., can probably explain inconsistencies between the results of different studies.

Due to the lack of standardization in systems, measures and procedures, researchers usually capture, analyze and interpret educational data via slow and tedious manual processes (Agudo-Peregrina et al., 2014). Student interaction via LMSs is the main focus of research since log data have the advantage of being non intrusive, readily available and free (Greenland, 2011). In order to assess the usefulness of LMSs in tracking data and predicting student success Macfadyen and Dawson (2010), conducted a pilot study of a fully online undergraduate biology course, offered at The University of British Columbia during 2008. Data collected on each student included counts for frequency of usage of course materials and tools supporting content delivery, engagement and discussion, assessment and administration/management. In addition, tracking data indicating total time spent on certain tool-based activities (assessments, assignments, total time online) offered a total measure of individual student time on task. Two statistical models, a multiple regression model and a binary logistic regression analysis, were applied to predict grade as a function of total number of discussion messages posted, number of completed assignments and number of messages sent, and test the reliability of the model in predicting whether or not an individual student is considered at risk of failure. Their regression model explained more than 30% of the variation in student final grade while logistic modeling demonstrated that the predictive model developed using these tracking variables correctly identified students at risk of failure with 70.3% accuracy.

This study uses the same approach as Macfadyen and Dawson (2010), but in a different learning setting — a blended learning course instead of a fully online course. Taking into account the increased opportunities for interaction between instructor and students in the face-to-face sessions, and the students' need for out-of-class communication and collaboration on projects and assignments, the underpinning research question was to determine if the data collected from the LMS log file was enough to predict grades in a hybrid learning environment. In this effort, all LMS activities related to blended learning were treated equally in a search for significant correlations with student grade. From these activities/variables, 14 were found having significant relationship with final course grade and were included in a multiple regression analysis, in order to develop a predictive model of outcomes in blended learning settings.

4. Research methodology

Data from an introductory Java programming course, taught in a blended mode, was gathered using the Moodle LMS activity tracking feature. The research methodology had three phases. In the first phase, a bivariate correlation between 29 online activities with student grade resulted in 14 variables with strong impact on student final achievement, which were then used as the input in a regression analysis. Given the fact that exploratory univariate regression analyses for student age, gender, previous grades, working status and ethnicity revealed that none of these variables had a significant effect on course grade, they were dropped from further analysis. In the second phase, these 14 strong independent variables were entered into a stepwise regression analysis to determine the significance as well as the amount of variance associated with each significant variable. This regression analysis resulted in a model with four independent variables, explaining 52% of variance in student grade. In order to validate the model, the WEKA 3.7 data mining software was used for creating a 10-fold cross-validation, which confirmed the robustness of the model. In the third phase, a binary logistic regression analysis was performed in order to evaluate the power of the model to discriminate between those students who are at risk of dropping from the course and those who are not. Students with actual grade lower than 5.5 were considered to be at risk of failure and coded as "0" while the others were coded as "1". The overall accuracy of classification was 81.3%.

4.1. Research questions

This study employed an ethnographic research design, which is well-suited when the focus is a single group without experimental control over conditions of assignment and acting of participants, to explore patterns of LMS usage and significant correlations between student actions and academic performance. Students' behavior was studied in their everyday life context and data was gathered from observation without having a detailed plan set up at the beginning of what actions should be recorded or how the categories of observations should be interpreted. The analysis of the data was guided by the following two questions:

1) Which online activities correlate significantly with student grades?
2) Which course LMS tracking variables form the best model predicting student success?

4.2. Participants and study context

This study investigated the relationships between student achievement and online activities of 134 university students who took a blended course, which combined traditional lectures and lab sessions with online interactions and online access to course materials. All students participating in this study were freshmen majoring in Computer Science and Computer Engineering at a large technological university during the Spring 2013 semester. The course under study, Introduction to Computer Science using Java, is designed for students who have no prior programming experience and aims to teach them the basic Java syntax and control flow, compilation and debugging, class and object usage, graphical interfaces, program–user interaction, and the Java API. During the last decade, the course has undergone extensive redevelopment, resulting in substantial changes in both content presentation and delivery format. From being a traditional campus based full-time course, it has evolved into a blended learning course in which online activities replaced 16 h of classroom lectures with online self-study modules. The traditional lecture and linked laboratory format is combined with a Moodle LMS-supported learning environment in which students could access online content, collaborate in projects, communicate with the instructors, submit homeworks, take quizzes and control their learning.

All lecture notes, assignments, course announcements and other course-related activities were available through Moodle LMS. Moodle (Modular Object Oriented Developmental Learning Environment) is a user-friendly, highly interactive, free, open-source LMS that enables the creation of flexible and engaging learning courses. Moodle's integrated tools allow teachers to organize online classroom activities and create new content, assignments, quizzes, etc., and students to access digital resources, engage in discussions, collaborate via forums, chats, wikis and blogs, view their progress, etc. Moodle analytics is also of great importance as it enables teachers to monitor and keep track of the learners' activities, accumulating information about students' online behavior and progress. The activities of the students are recorded in the system's log file, providing information about the name and IP address of each student, the time of the visit, and the action performed.

4.3. Data gathering

Once the course had been completed, the Moodle reporting feature was used to extract student clicks from the system's log file to MS Excel. A single spreadsheet, containing the complete set of data logs for the entire semester, was created and filtered with several Excel macros to create entries that illustrate, for each student, the total time spent online, the number of online sessions, the total number of messages sent, etc. For the purposes of this study, an online session was defined as the sequence of a user's interactions with course web page, from entering after login until logging out or staying 40 min inactive. Time spent on each activity was computed by summing the time from the first click engaging the activity until the time of the next click onto another course object. Next, the log file was joined with final course grade for each participant, resulting in a consolidated data file that combined student LMS activities and final course score (Fig. 1).

Because the completion of assigned projects required extensive interaction between students, reading and posting messages via either discussion forum, chat or emails, were summed and produced a new variable, *RePo Messages*, which was examined in parallel with the original ones. This combined variable was meant to capture the flexibility of team work and describe the effect of interaction, via the system tools, to the course grade. Students collaborated on projects and discussed a number of topics using Moodle's communication facilities and, of course, all other means available outside the course web-page, like Skype, Facebook, blogs, etc. Tracking only the system's log file to discover indicators of success, seems to be an imperfect attempt to deal with such a complicated phenomenon as the human collaboration process. Furthermore, since not all text message posts are of the same quality and given that there was no way to know if a user read one or all of the posts in a forum page or if her email is just a media gossip or a thorough analysis of a program code, the *RePo Messages* variable was created to test the whole system-logged interactions against student final grade. All the variables composing the *RePo Messages* variable, were also examined individually to find useful associations with course grade.

Content creation in class wiki and site blog was treated as a separate variable, *CCC* or *Content creation contribution*, measuring student contribution and collaboration with other students. This variable was also examined in parallel with wiki and blog related variables. Quiz attempt related actions, such as *quiz attempt*, *quiz continue attempt* and *quiz close attempt*, were combined into a single variable, *Quiz efforts*, due to very few cases in the sample, and *quiz view* and *quiz review* actions were joined to echo another variable, *Quiz view*, since the effect of both on student performance is not of primary interest from a pedagogical perspective. Because some resources (pdf, ppt, word, excel files) when uploaded on the server are recorded as *resource view* in the log file, while others (html pages within Moodle, external urls) are recorded as *page view* or *url view*, a combined variable, *Files viewed*, was used

	A	B	C	D	E	F	
1	User	grade	total hits	time online	# online sessions	# mgs posted	# file
2	CS02-st001	6,3	2549	35	279	54	
3	CS02-st002	8,4	5787	42	204	62	
4	CS02-st003	6,9	5899	37	187	48	
5	CS02-st004	7,5	7403	25	294	61	
6	CS02-st005	7,9	6885	29	256	37	
7	CS02-st006	5,8	6409	38	192	50	
8	CS02-st007	7,3	7703	47	307	68	
9	CS02-st008	5,4	6641	18	282	56	
10	CS02-st009	8,3	8660	25	273	41	

Fig. 1. Final data file.

Table 1
Student end-of-course grades.

Students with final grade	#	Mean	Mode	Min	Max	Std. dev.
Up to 5	23	4.10	4.10	3.10	5.00	0.57
Between 5 and 5.5	20	5.30	5.30	5.20	5.50	0.09
Between 5.5 and 7	36	6.40	6.50	5.60	7.00	0.45
Between 7 and 8	31	7.62	7.60	7.20	8.00	0.24
Over 8	24	9.17	8.80	8.50	10.00	0.46

to describe the uses of the available resources, without excluding the component variables from the analysis.

5. Results

The study employed both descriptive and inferential statistics. The descriptive analysis included the calculation of simple statistical measures such as means, standard deviations and correlations of all the variables investigated in the study with course grades. The inferential analysis was a forward stepwise multiple regression, run on SPSS 16.0 with level of significance of .05, to determine the relationship between the contextual variables and the student grade. The descriptive statistics of different quartiles of students' final grades, which was the independent variable in this study, are shown in Table 1, while the number of hits per week per student, a popular variable used in studies measuring student online effort or engagement, is illustrated in Fig. 2.

Simple scatterplots generated by SPSS were used to graphically present relationships between different variables and grade, determine whether a relationship is linear and detect outliers (see for example Fig. 3).

Correlation analysis of student actions via LMS found that there were 14 variables with a significant correlation ($p < 0.05$) to course grades (Table 2). Out of 29 actions/variables examined, one demonstrated a strong effect size ($R = 0.60$–0.79), two showed a moderate effect size ($R = 0.40$–0.59) and eight had a weak effect size ($R = 0.20$–0.39). Reading and posting messages, an imperative activity in collaborative blended learning environments, explained 37% of the variance in student course grade, while each one of the next five variables (*Wiki edit*, *Content creation contribution*, *Mail messages read*, *Assignments submitted*, *Quiz engagement*) explained from 10% to 27% of student course grade.

The analysis of the amount of the time spent online (Fig. 4), which is an important measurement of the involvement of a student in online learning, reveals that the average number of hours spent on the course web site was 40.73 ranging between 23 and 59 h, with a mode of 29 and standard deviation of 0.93. The correlation between time spent online and student final grade was weak, but similar to the correlation between total LMS hits and student grade.

Although correlation coefficients are of great value as sensitive indicators of relationship strength (effect size), they cannot explain why there is a relationship or which variable is affecting which. Being simply a way to describe how two variables vary together, correlation coefficients do not control for the other variables that affect the dependent variable and, therefore, can give false relationships. On the contrary, linear regression gives coefficients while controlling for the other variables, capturing in a better way the influence of independent variables on dependent variables.

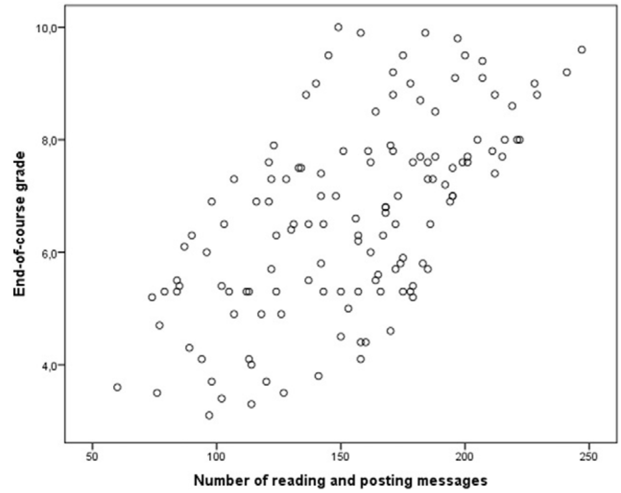

Fig. 3. Messages–grades correlation.

In previous studies, the researchers specified the set of predictor variables that make up the model and then assessed the success of this model in predicting the dependent variable. Since we had no reason to believe that one variable is likely to be more important than another in predicting student final grade, a stepwise regression was used to find the best set of independents. Stepwise is a sophisticated selection method that combines forward selection and backward elimination. Variables are entered in sequence in the regression and if they contribute to the model then they are retained. At the same time, all other variables are checked to see if they continue to contribute significantly to the predictive power of the model, and if not, they are removed. The process continues until no more variables are added or removed. Therefore, stepwise regression provides the minimum number of variables needed to build the predictive model. Stepwise regression model summary of all LMS variables on student course grade is shown in Table 3.

SPSS output in Table 3 shows that the variable *RePo messages* (reading and posting messages) was entered first in the regression analysis and was the best predictor, explaining 37.6% of the variation on final course grade. Three other variables entered progressively and remained as significant predictors: *CCC* (Content creation contribution), *Quiz efforts* (Quiz engagement), and *Files viewed*. The proportion of the variance of final course grade explained by the best fitting model is 52% (R^2 for Model 4 – the final model – in Table 3). The ANOVA test of significance shown in Table 4 reveals that the final regression model (Model 4) was significant ($F = 34.881$) at the 0.05 level.

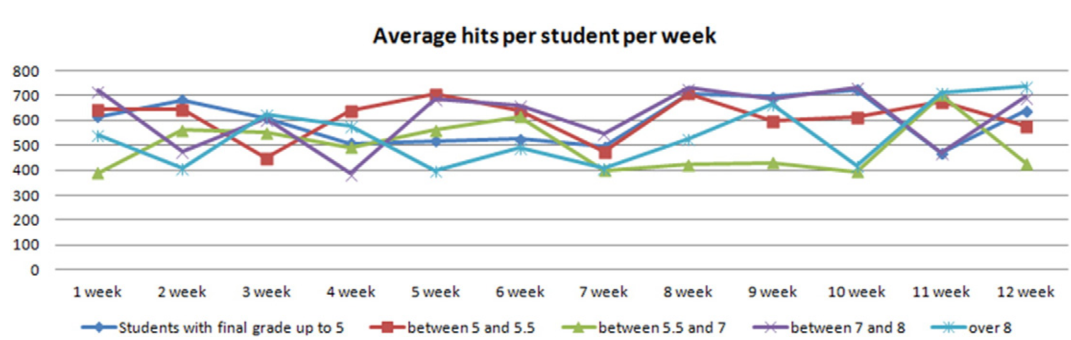

Fig. 2. Online student hits.

Table 2
Correlations with course grade.

Variable	R	R^2	p
Reading and posting messages (RePo messages)	0.61	0.37	0.00
Wiki edit	0.52	0.27	0.00
Content creation contribution (CCC)	0.49	0.24	0.00
Mail messages read (user view)	0.38	0.14	0.00
Assignments submitted (assign submit)	0.34	0.11	0.01
Quiz engagement (quiz efforts)	0.32	0.10	0.01
Blog update	0.26	0.07	0.02
Files viewed	0.24	0.06	0.02
Web links viewed (page view + url view)	0.23	0.05	0.02
Total time online	0.22	0.05	0.02
Total LMS hits	0.21	0.04	0.02
Wiki view	0.19	0.04	0.03
Online sessions	0.19	0.04	0.03
Forum add post	0.17	0.03	0.04
Assessments started	0.15	0.02	0.06
Blog view	0.14	0.02	0.09
Wiki add page	0.12	0.01	0.13
Quiz continue attempt	0.12	0.01	0.17
Chat view	0.11	0.01	0.19
Resource view	0.09	0.01	0.22
Quiz close attempt	0.06	0.00	0.23
Accesses to grades tool (assign view)	0.04	0.00	0.26
Course view	0.04	0.00	0.31
Quiz attempt	0.03	0.00	0.33
Forum view discussion	0.02	0.00	0.48
Quiz view	0.01	0.00	0.55
Quiz review	0.00	0.00	0.59
Chat talk	0.00	0.00	0.75
Assignment view	0.00	0.00	0.84

The coefficients for the final regression Model 4 and their standard errors are given in Table 5. From the values of the regression coefficients, the predicted value of student final grade for any case can be calculated as: course grade = 1.338 + 0.020 ∗ *RePo messages* + 0.49 ∗ *CCC* + 0.042 ∗ *Quiz efforts* + 0.012 ∗ *Files viewed*. Higher scores on all four variables predict higher scores on course grade (the coefficients are positive).

The values in the t and Sig columns determine whether or not the b-value differs significantly from zero, assuming that the other variables are included in the regression. If the t-test associated with a b-value is significant (if the value in the column labeled Sig. is less than .05) then the predictor is making a significant contribution to the model. We can see that all four predictors are significantly different from zero, contributing significantly to the predictive model.

Table 3
Model summary for stepwise regression.

Model	R	R square	Adjusted R square	Std. error of the estimate
1	0.613[a]	0.376	0.371	1.3563
2	0.693[b]	0.480	0.472	1.2430
3	0.711[c]	0.505	0.493	1.2172
4	0.721[d]	0.520	0.505	1.2036

[a] Predictors: (constant), *RePo messages*.
[b] Predictors: (constant), *RePo messages*, *CCC*.
[c] Predictors: (constant), *RePo messages*, *CCC*, *Quiz efforts*.
[d] Predictors: (constant), *RePo messages*, *CCC*, *Quiz efforts*, *Files viewed*.

The model was validated by a 10-fold cross-validation, with the root mean square error (RMSE) as an estimate for the standard deviation of the fit. 10-fold cross-validation works by splitting the data into a set of 10 segments or folds and performing 10 iterations of training and validation. In each iteration a different fold of the data is held out for validation, with the remaining 9 folds used for training. The WEKA 3.7 data mining software was used for creating the 10-fold cross-validation. The final result (Table 6) confirms the robustness of the model.

In order to evaluate the power of the model to discriminate between those students who are at risk of dropping from the course and those who are not, a binary logistic analysis was performed. All students with grades lower than 5.5 were counted as "Failed" and coded as "0", meaning that they were at risk of failure, while all others were coded as "1". As can be seen in Table 7, the logistic regression model correctly classified 30 students who were failed the course but misclassified 13 others (it correctly classified 69.8% of cases). The model also correctly classified 79 students who were not failed but misclassified 12 others (it correctly classified 86.8% of cases). The overall accuracy of classification is, therefore, the weighted average of these two values (81.3%).

6. Discussion of the findings

This study investigated the relationship of different LMS data variables with student achievement in the context of a blended learning programming course, in order to build a predictive model capable of identifying students at risk of failure. The literature review reveals that courses with online components provide students a flexible environment for learning while giving teachers extensible activity-monitoring capability. Several studies have been conducted so far to identify associations between selected LMS variables (online activities or tools used) and student academic performance. Time spent online, hit frequency and the number of discussion messages read or posted are some of

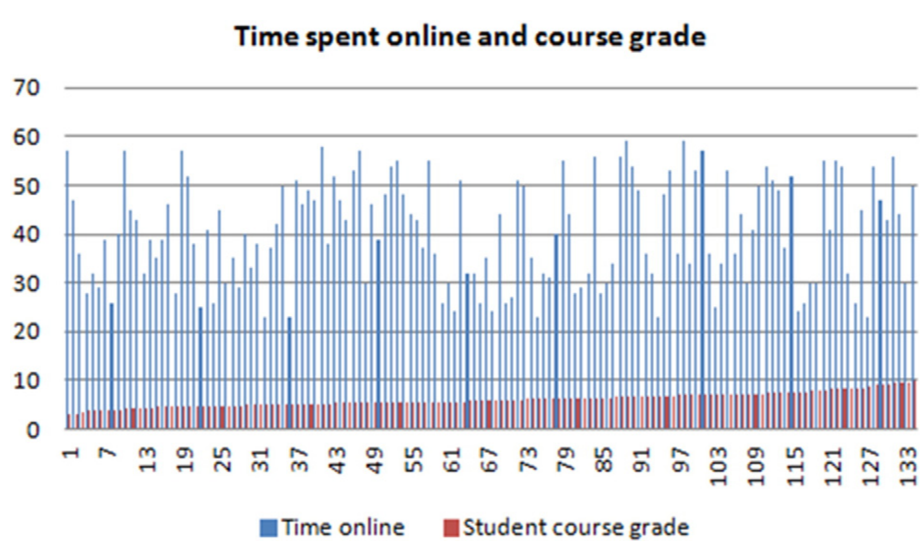

Fig. 4. Time spent online and course grade.

Table 4
Anova[a] SPSS output.

Model	Sum of squares	df	Mean square	F	Sig.
4 regression	202.119	4	50.530	34.881	0.000[b]
Residual	186.874	129	1.449		
Total	388.992	133			

[a] Dependent variable: course grade.
[b] Predictors: (constant) *RePo messages*, *CCC*, *Quiz efforts*, *Files viewed*.

the most frequently examined determinants towards student success. Although this univariate approach has the advantage of simplicity and provides valuable information on specific variables, a multivariate analysis can increase understanding of the complexity of student learning and achievement.

In this light, the study analyzed the log files generated by the Moodle LMS that hosted a blended learning course in order to find significant correlations between LMS data usage variables and final course grade. A bivariate correlation analysis of 29 course-related variables identified 14 of them to be significantly associated with course grade. A stepwise multiple regression analysis was then conducted on these 14 variables and revealed that 52% of the variance in the final student grade was predicted by just four variables: *Reading and posting messages*, *Content creation contribution*, *Quiz efforts* and number of *Files viewed*.

As expected, reading and posting messages on forum board, email and chat was found to be significantly correlated with course success, explaining 37.6% of the variation in the final student grade. This finding confirms the general belief that high levels of communication in blended learning settings are strongly associated with successful course completion. Previous research in blended learning environments has found that content creation via wiki and blog tools contributes to the development of collaboration components and enhances student comprehension. In this study, *CCC* (content creation contribution) variable accounted for an additional 10.4% variance, demonstrating itself as a strong predictor of final student grade.

A rather unexpected finding was that greater student engagement with online quizzes associated with higher final scores. Previous studies have shown that compulsory online quizzes (Macfadyen & Dawson, 2010) or graded optional online self-tests (Tempelaar et al., 2014), are strong predictors for indentifying underperforming students and academic success. In this study however, students were simply encouraged to use optional Moodle quizzes (designed as multiple choice or true/false or fill-in-the-blanks) to assess their understanding of lesson material. Quiz scores were not a factor in the course grading scheme and, thus, it was uncertain if students would make extensive use of the optional online quizzes. Throughout the academic term, students had unlimited access to 14 online quizzes and it seems that their small number and their non-obligatory nature made them attractive for students, who used them as a review and revision tool.

By contrast, the number of assignments submitted was not a significant predictor, although it was found to have a strong (univariate) correlation with final grade. The number of the files viewed by the students during the course was the fourth predictable variable in the best-fit model resulting from the stepwise regression. Student engagement with course content was proved, as expected, to be a decisive factor of student success. Both variables (*Quiz efforts* and *Files viewed*) accounted for 4% of the variation of final student grade.

Three variables that have been floating around in the online learning literature – *Total time online*, *Total LMS hits* and number of *Online sessions* – and are considered as measures of student effort, engagement and participation, were found to have a weak (univariate) correlation with performance grade. A possible reason for these weak associations is that, by its own specific design, this blended learning course encouraged students to take quizzes to assess their understanding, collaborate online on projects and create knowledge as a team. Students' online activity through LMS tools was logged and analyzed without any particular set of variables in mind and the stepwise regression demonstrated four variables, which correspond to interactions with the peers, the instructor and the teaching resources. The predictive model that contained these four variables was then tested for its ability to discriminate between students classified as at risk of failure and those not similarly labeled. The overall percentage of correct predictions by binary logistic regression was 81.3%, indicating the adequacy of the model.

Table 5
Regression coefficients.[a]

Model	Unstandardized coefficients		Standardized coefficients		
	B	Std. error	Beta	t	Sig.
4 (constant)	1.338	0.491		2.727	0.007
RePo messages	0.020	0.003	0.481	7.444	0.000
CCC	0.049	0.011	0.299	4.603	0.000
Quiz efforts	0.042	0.018	0.152	2.393	0.018
Files viewed	0.012	0.006	0.124	1.989	0.049

[a] Dependent variable: course grade.

Table 6
Weka 10-fold cross validation.

=== Run information ===	
weka.classifiers.functions.LinearRegression -S 0 -R	
Scheme:	1.0E-8
Relation:	Student_grade
Instances:	134
Attributes:	5
Test mode:	10-fold cross-validation
=== Classifier model (full training set) ===	
Linear regression model	
Grade = 0.0201 ∗ *RePo messages* + 0.0485 ∗ *CCC* + 0.0423 ∗ *Quiz_efforts* + 0.0116* *Files_viewed* + 1.3379	
Time taken to build model: 0.01 s	
=== Cross-validation ===	
=== Summary ===	
Correlation coefficient	0.6732
Mean absolute error	1.0097
Root mean squared error	1.2633
Total number of instances	70.3817%
Root relative squared error	73.5209%
Relative absolute error	134

7. Implications of results

Teaching in blended learning environments is a complex process which, among other duties, requires continuous monitoring of student engagement with and understanding of the online course content. Learning analytics provide instructors with the tools to improve educational outcomes through the analysis of data about how students are interacting with the online environment. Different blended learning approaches can be supported by different levels of autonomy, collaboration, and online technologies, but in any case, the point of learning analytics is to enable deeper understanding and better decision-

Table 7
Testing the model's classification accuracy.

	Observed		Predicted[a]		
			Failed?		Percentage correct
			Yes	No	
Step 1	Failed?	Yes	30	13	69.8
		No	12	79	86.8
	Overall percentage				81.3

[a] The cut value is 0.5.

making. Learning management systems, the main platforms for delivering online instruction today, provide teachers a vast amount of data about how often students log in, how often they post in discussion boards, what they view, how long they stay on the site, etc. Although the extraction of data from the LMS is easy, finding meaningful behavior patterns and relationships that inform effective learning is a time consuming task. Most tracking tools found in current LMSs organize and report data in a complex tabular format, while stand-alone data mining tools, such as SPSS, Weka, Excel with VB macros, and RapidMiner, are rather difficult to use. Data stored in an LMS database needs to be transformed before being analyzed, as well as cleaned from noisy data. Without experience in writing complex SQL queries and scripts that combine SQL commands, the only way out for the non-technical savvy users may be to extract data by manually performing tedious filtering loops.

Capturing and maintaining student engagement in blended learning courses are critical aspects of the learning process towards enhancing outcomes for all learners. Clear content structure, unambiguous instructions and goals, challenging tasks and timely feedback are indispensable constituents of different teaching scenarios that combine various learning strategies and technologies to motivate and engage students. Monitoring and early identification of students who are at risk of falling behind is a key factor to engagement and success, and LMSs can provide a wide range of predictors and indicators that have the potential to inform different types of decisions. Therefore, deciding what data should be collected and analyzed, and for what purpose, is central. The overall design of a blended learning course, and the particular online activities that support different aspects of the learning circle, such as dialog and communication, collaborative content creation and presentation, project-based group work, individual assessments and online quizzes, can guide the choice of the appropriate variables and the data analysis. Instructors can experiment with different combinations of metrics, such as IP addresses, content page views, time on task, average session length, total messages exchanged and content creation contributions. In line with the findings by Macfadyen and Dawson (2010) and Tempelaar et al. (2014), who found that some, but not all, variables are useful predictors of student achievement in an LMS-supported course, the present study highlighted the role of only four LMS variables, out of 29 examined, in predicting student course grades.

With the aim to identify outliers for early intervention, the study analyzed the frequency of use of LMS activities/tools, in an introductory programming course taught in blended learning format. This course was designed to promote engagement, self-directed learning and the creation of new artifacts through knowledge sharing and cooperation, and as the results show, the variables representing these aspects of blended learning effectively predicted student outcomes. The combined variable *RePo messages*, the most significant predictive variable in the model, highlights the importance of engagement with peers and instructor, while the *Files viewed* variable represents the level of engagement with online material. *CCC* (content creation contribution) variable, the second most dominant predictor, represents the learning potential of co-creating and sharing knowledge, while the *Quiz efforts* variable reflects the ability of online quizzes to support self-directed learning. Different blended learning strategies for different areas of knowledge or learning environments may highlight a different set of predictors. Although this study proposes an easy-to-use model to predict success, further research needs to be done to evaluate the accuracy of the model in different learning contexts and in larger, more heterogeneous student cohorts.

Since a large part of interaction today occurs through social networks or other digital platforms, a crucial amount of data is not stored in the LMS database and escapes the scrutiny of researchers. Therefore, there is a need for collecting and analyzing data from various media (blogs, facebook, twitter, wikis, skype, etc.) that students use outside of the LMS to complete their tasks, to get a clearer picture of the whole learning process and develop a finer understanding of the impact of various interaction modes on course grades. Content analysis of the written messages could also be used to determine the presence and frequency of certain words or phrases within texts, and identify the degree to which course related factors, such as complexity of content and design, workload and instructor's personality, influence success. Furthermore, data on students' learning dispositions, such as self-esteem, perseverance, social and communications skills and thinking flexibility, could be gathered through self-reported surveys to provide “higher order information about one's state or intentions, which are harder to infer from low-level system event logs” (Buckingham Shum & Deakin Crick, 2012, p. 3). Finally, since current LMSs report students' outcomes, by displaying data in the form of a visualization dashboard, without analyzing the learning process itself, there is a need for the development of easy-to-use visual analytic dashboards that provide learners and instructors useful insights through interactive visualizations of the various stages of the learning process.

8. Conclusion

Developing a successful blended learning course isn't always an easy task. Besides the need to address technology requirements, decide the best content to deliver or design an engaging set of classroom and online activities, there is also a great need for monitoring and tracking students' online activities. Extracting and analyzing LMS usage data stored in their log files provide instructors the means for monitoring student progress and planning interventions. This study proposes a mathematical model, consisting of four LMS data variables that can enable identification of students who are at risk of failure. Selecting the right variables to drive a predictive model is a decisive step towards an early warning system tied to student success in blended learning courses.

References

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M.Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE supported F2F and online learning. *Computers in Human Behavior*, *31*, 542–550. http://dx.doi.org/10.1016/j.chb.2013.05.031.

Allen, E., & Seaman, J. (2009). *Learning on demand: Online education in the United States.* Needham, Mass: Sloan-C (http://goo.gl/2OtLp6).

Ashford-Rowe, K., & Malfroy, J. (2009). *E-learning benchmark report: Learning management system (LMS) usage.* Sydney: Griffith University (http://goo.gl/zUBNS).

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*(2), 159–179. http://dx.doi.org/10.1111/j.1540-4609.2003.00016.x.

Beer, C., Clark, K., & Jones, D. (2010). Indicators of engagement. *Curriculum, technology and transformation for an unknown future. Proceedings ascilite, Sydney 2010* (http://goo.gl/YrDw5D).

Biktimirov, E. N., & Klassen, K. J. (2008). Relationship between use of online support materials and student performance in an introductory finance course. *Journal of Education for Business*, *83*(3), 40–48. http://dx.doi.org/10.3200/JOEB.83.3.153-158 (January/February).

Bowen, G. W., & Lack, A. K. (2012). Current status of research on online learning in post-secondary education. http://goo.gl/BlaJX

Buckingham Shum, S., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. *Paper presented at the 2nd International Conference on Learning Analytics & Knowledge, Vancouver, British Columbia* (http://oro.open.ac.uk/32823/).

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Journal of Educational Technology & Society*, *15*(3). http://dx.doi.org/10.1145/2330601.2330616.

Chanchary, F. H., & Haque, I. (2007). Preparedness of E-learners: A study on first year university students. *Proceedings of the 5th International Conference on ICT and Higher Education, Knowledge Management, Bangkok, Thailand* (http://goo.gl/wALoJk).

Coldwell, J., Craig, A., Paterson, T., & Mustard, J. (2008). Online students: Relationships between participation, demographics and academic performance. *Electronic Journal of E-Learning*, *6*(1), 19–30 (http://www.ejel.org/issue/download.html?idArticle=57).

Crampton, A., Ragusa, A., & Cavanagh, H. (2012). Cross-discipline investigation of the relationship between academic performance and online resource access by distance education students. *Research in Learning Technology*, *20*. http://dx.doi.org/10.3402/rlt.v20i0/14430.

Damianov, D., Kupczynski, L., Calafiore, P., Damianova, E., Soydemir, G., & Gonzalez, E. (2009). Time spent online and student performance in online business courses: A multinomial logit analysis. *Journal of Economics and Finance Education*, *8*(2), 11–19 (http://goo.gl/70ZOVo).

Dawson, S., Heathcote, L., & Poole, G. (2010). Harnessing ICT potential: The adoption and analysis of ICT systems for enhancing the student learning experience. *International Journal of Educational Management*, *24*(2), 116–128. http://dx.doi.org/10.1108/09513541011020936.

Dawson, S., McWilliam, E., & Tan, J. P. L. (2008). Teaching smarter: How mining ICT data can inform and improve learning and teaching practice. *Hello! Where are you in the landscape of educational technology? Proceedings ascilite, Melbourne* (http://goo.gl/BZkk5U).

Eduviews (2009). *Blended learning: Where online and face-to-face instruction intersect for 21st century teaching and learning*. Eduviews: A K-12 Leadership Series (http://goo.gl/nDxok).

Ellis, R. (2009). *A field guide to learning management systems*. ASTD Learning Circuits (http://goo.gl/yqHBGs).

Garrison, D. R., & Vaughan, N. (2008). *Blended learning in higher education: Framework, principles and guidelines*. San Francisco: Jossey-Bass. http://dx.doi.org/10.1002/9781118269558.

Glazer, F. (2012). *Blended learning: Across the disciplines, across the academy*. Sterling: VA Stylus Publishing (http://goo.gl/rShgFu).

Gradel, K., & Edson, A. J. (2010). Cooperative learning: Smart pedagogy and tools for online and hybrid courses. *Journal of Educational Technology Systems*, *39*(2), 193–212. http://dx.doi.org/10.2190/ET.39.2.i.

Graham, C. R. (2006). Blended learning systems: Definition, current trends, and future directions. In C. J. Bonk, & C. R. Graham (Eds.), *Handbook of blended learning: Global perspectives, local designs*. San Francisco: Pfeiffer Publishing (http://php.indiana.edu/~cjbonk/graham_intro.pdf).

Gray, G., McGuinness, C., Owende, P., & Carthy, A. (2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1) (http://goo.gl/fpjwMA).

Greenland, S. (2011). Using log data to investigate the impact of (a)synchronous learning tools on LMS interaction. *Changing demands, changing directions. Proceedings ascilite, Hobart* (http://www.editlib.org/p/43586/).

Hirumi, A. (2012). The design and sequencing of online and blended learning interactions: A framework for grounded design. *Canadian Learning Journal*, *16*(2), 21–25 (http://goo.gl/rOMDRn).

Ishtaiwa, F., & Abulibdeh, E. (2012). The impact of asynchronous e-learning tools on interaction and learning in a blended course. *International Journal of Instructional Media*, *39*(2), 141–159 (http://www.highbeam.com/doc/1G1-289721761.html).

Jaggars, S. S. (2011). Online learning: Does it help low-income and underprepared students? *CCRC Working Paper No. 26, Assessment of Evidence Series*. New York, NY: Columbia University, Teachers College, Community College Research Center (http://hdl.handle.net/10022/AC:P:13230).

Jenkins, R. (2011). Why are so many students still failing online? http://chronicle.com/article/Why-Are-So-Many-Students-Still/127584

Kats, Y. (2010). *Learning management system technologies and software solutions for online teaching: Tools and applications*. IGI Pub (http://goo.gl/lWtVAf).

Macdonald, J. (2008). *Blended learning and online tutoring* (2nd ed.). London: Gower (http://goo.gl/mBDHAi).

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599. http://dx.doi.org/10.1016/j.compedu.2009.09.008.

Mahle, M. (2011). Effects of interactivity on student achievement and motivation in distance education. *Quarterly Review of Distance Education*, *12*(3), 207–215 (http://goo.gl/bdMcn4).

Mattingly, K., Rice, C., & Berge, Z. (2012). Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management & E-Learning: An International Journal*, *4*(3), 236–247 (http://goo.gl/Iz2cq2).

Milne, J. D., Jeffrey, L. M., Suddaby, G., & Higgins, A. (2012). Early identification of students at risk of failing. In M. Brown, M. Hartnett, & T. Stewart (Eds.), *Future challenges, sustainable futures. Proceedings ascilite 2012 Wellington*. (pp. 657–661) (http://goo.gl/u3LVkb).

Moskal, P., Dziuban, C., & Hartman, J. (2013). Blended learning: A dangerous idea? *Internet and Higher Education*, *18*, 15–23. http://dx.doi.org/10.1016/j.iheduc.2012.12.001.

Nandi, D., Hamilton, M., Harland, J., & Warburton, G. (2011). How active are students in online discussion forums. *Proceedings of the Thirteenth Australasian Computing Education Conference. Conferences in Research and Practice in Information Technology*, *vol. 114*. (pp. 125–134) (http://crpit.com/confpapers/CRPITV114Nandi.pdf).

Naveh, G., Tubin, D., & Pliskin, N. (2010). Student LMS use and satisfaction in academic institutions: The organizational perspective. *The Internet and Higher Education*, *13*(3), 127–133. http://dx.doi.org/10.1016/j.iheduc.2010.02.004.

Nissen, E., & Tea, E. (2012). Going blended: New challenges for second generation L2 tutors. *Computer Assisted Language Learning*, *25*(2), 145–163. http://dx.doi.org/10.1080/09588221.2011.636052.

Picciano, A. (2009). Blending with purpose: The multimodal model. *Journal of Asynchronous Learning, Networks*, *13*(1), 7–18 (http://goo.gl/lic2G0).

Piña, A. A. (2010). An overview of learning management systems. In Y. Kats (Ed.), *Learning management systems technologies and software solutions for online teaching: Tools and applications* (pp. 314–331). Hershey, PA: IGI Global (http://goo.gl/d8USyv).

Reaburn, P., Muldoon, N., & Bookallil, C. (2009). Blended spaces, work based learning and constructive alignment: Impacts on student engagement. *Proceedings ascilite Auckland* (pp. 820–831) (http://goo.gl/SqrCo).

Rubin, B. (2013). University business models and online practices: A third way. *Online Journal of Distance Education Administration*, *16*(1) (http://goo.gl/w8X0s).

Salmon, G. (2013). *E-tivities: The key to active online learning* (2nd ed.). Routledge (http://goo.gl/xBqenV).

Sheridan, K., & Kelly, M. (2010). The indicators of instructor presence that are important to students in online courses. *Journal of Online Learning and Teaching*, *6*(4), 767–779 (http://goo.gl/RneYuH).

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. http://dx.doi.org/10.1177/0002764213498851.

So, H., & Brush, T. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education*, *51*(1), 318–336. http://dx.doi.org/10.1016/j.compedu.2007.05.009.

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2014). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*. http://dx.doi.org/10.1016/j.chb.2014.05.038 (Advance online publication).

The Horizon Report (2013). http://www.nmc.org/pdf/2013-horizon-report-HE.pdf

U.S. Department of Education (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development (http://goo.gl/gBZi8B).

Vengroff, R., & Bourbeau, J. (2006). In-class vs. on-line and hybrid class participation and outcomes:Teaching the introduction to comparative politics class. *Paper presented at the Annual Meeting of the APSA Teaching and Learning Conference, Washington, D.C* (http://goo.gl/yb68xm).

Ward, M. E., Peters, G., & Shelley, K. (2010). Student and faculty perceptions of the quality of online learning experiences. *International Review of Research in Open and Distance Learning*, *11*(3), 57–77 (http://www.irrodl.org/index.php/irrodl/article/view/867).

Xu, D., & Jaggars, S. S. (2011). Online and hybrid course enrollment and performance in Washington State community and technical colleges. *CCRC Working Paper No. 31*. New York, NY: Columbia University, Teachers College, Community College Research Center (http://files.eric.ed.gov/fulltext/ED517746.pdf).

Zacharis, N. Z. (2010). The impact of learning styles on student achievement in web-based versus an equivalent face-to-face course. *College Student Journal*, *44*(3), 591–597 (http://goo.gl/BctAwR).

Zacharis, N. Z. (2011). The effect of learning style on preference for web-based courses and learning outcomes. *British Journal of Educational Technology*, *42*(5), 790–800. http://dx.doi.org/10.1111/j.1467-8535.2010.01104.x.