Using Learning Analytics to Predict Students Performance in Moodle LMS

https://doi.org/10.3991/ijet.v15i20.15915

Yaqun Zhang (✉)  
Shenyang Normal University, Shenyang, China  
Tomsk State University, Tomsk, Russian Federation  
334812333@qq.com

Ahmad Ghandour  
Al Ain University, Abu Dhabi, United Arab Emirates

Viktor Shestak  
MGIMO University, Moscow, Russian Federation  
Military University, Moscow, Russian Federation

Abstract—Today, it is almost impossible to implement teaching processes without using information and communication technologies (ICT), especially in higher education. Education institutions often use learning management systems (LMS), such as Moodle, Edmodo, Canvas, Schoology, Blackboard Learn, and others. When accessing these systems with their personal account, each student’s activity is recorded in a log file. Moodle system allows not only information saving. The plugins of this LMS provide a fast and accurate analysis of training statistics. Within the study, the capabilities of several Moodle plugins providing the assessment of students’ activity and success are reviewed. The research is aimed at discovering possibilities to improve the learning process and reduce the number of underperforming students. The activity logs of 124 participants are analyzed to identify the relations between the number of logs during the e-course and the final grades. In the study, a correlation analysis is performed to determine the impact of students’ educational activity in the Moodle system on the final assessment. The results reveal that gender affiliation correlates with the overall performance but does not affect the selection of training materials. Furthermore, it is shown that students who got the highest grades performed at least 210 logs during the course. It is noted that the prevailing part of students prefers to complete the tasks before the deadline. The study concludes that LMSs can be used to predict students’ success and stimulate better results during the study. The findings are proposed to be used in higher education institutions for early detection of students experiencing difficulties in a course.

Keywords—Learning management systems, Moodle, electronic journal file, Moodle plugin, student success, student behavior.
1 Introduction

Since the advent of personal computers and the Internet, almost every aspect of life, including the educational system, has changed [1]. E-learning systems, in particular, Edmodo, Canvas, Schoology, Blackboard LearnSakai, Moodle, ATutor, Chisimba [2] are actively introduced into the higher education institutions [3]. Thus, the matter of e-learning has attracted the attention of many researchers. Currently, various high-quality publications that encourage online learning methods exist [4]. The importance of e-learning technology is improving steadily. Many universities use it as a key tool in educational programs [5].

The learning management system (LMS) is usually focused on organizing courses by teachers and includes managing learners in a modern online learning environment [6]. To succeed in higher education, as well as in future life and career, students should have a number of so-called “21st century skills” (for example, critical thinking, creativity, communication). Therefore, to enhance these skills, online learning systems should be developed. E-learning has several significant advantages over traditional classes in the classroom, primarily due to accessibility and flexibility [7]. It is easier to search for information in online resources for learning anytime, anywhere [8].

Initially, LMSs are simple and can be compared to regular web pages with information about disciplines, lecturers, teaching, and monitoring methods [9]. Gradually, LMSs became more complex and multi-tasking [10]. These days, LMSs provide the opportunity not only to exchange educational materials but also to interact with teachers and other students. Besides, LMSs allow objective assessment and statistical interpretation of the results of mastering the course by students.

Modern e-learning systems have advanced technologies and functions to support all forms of educational activity, including face-to-face interaction during the study [11]. Some training programs allow applicants to become students and get a university diploma without even leaving home. LMS is used to create study groups, conduct lectures, seminars, practical classes, and pass exams [12].

Over the past decade, many universities have acquired or developed LMS for curriculum management, and the creation of training materials, as well as a student assessment tool. Since 2012, global spending on LMS grew by 52% (21% only in 2014), totaling more than $2.5 billion per year. Nine out of ten US schools use one of the top five LMS providers. In the US, Blackboard holds the largest market share with 42%. The value of using educational analytics is that it changes the quality of administration, research, teaching and learning and allows countries such as the EAU to implement the best modern practices in education and expand the presence of the world's largest universities in the country [13].

Educational institutions spend thousands of dollars to implement LMSs, seeking to improve the quality of education, as well as increase the number of students through distance and blended learning. The impact of this system on improving students' performance has been a popular subject of research in recent years. Studies have been relying on data from users' opinions and subjective interpretation through surveys to determine the effectiveness of LMS usage on students' learning performance [14].
Even though the full distance learning course remains an advanced approach to education, additional opportunities that expand the use of the e-learning also exist [15]. Leading educational institutions are actively introducing learning analytics for monitoring and evaluating educational activities. Literature is replete with research on the benefits of e-learning and its impact on students' academic performance. For example, a significant positive correlation is found between the use of online tools and student exam results. It is discovered that using technology can improve academic performance and helps students with special needs to learn the course [16].

In Russia, distance learning is offered by more than 30 universities. The relevance of introducing e-learning approach is due to the growing tendency to reduce in-class learning and increase the share of students' independent educational activity [17]. LMS ensures the integrity of the learning process, its optimization, and objective control. Furthermore, students consider online learning useful since the availability of access to educational files, participation in the content formation, the possibility of real-time monitoring, and knowledge self-control [18].

In 2000, the Ministry of Education in the People's Republic of China (PRC) approved over 60 educational institutions for distance learning [19]. Today, additional Moodle plugins that expand the analytical capabilities of the system are actively introduced into the educational process [20].

Most universities in UAE offer distance education. This form of training is widely developed due to the presence in the country of departments of such famous foreign universities as the University of New York, Paris Sorbonne University and others. Most of these distance learning courses rely on Moodle and its modules in their work [21]. This choice is also due to easier opportunities for the implementation of training analytics in Moodle. The use of educational analytics in recent years has been one of the areas of constant interest of Arab researchers, because it opens up the possibility of predicting the quality of student learning. Based on the prediction, the shortcomings of training courses can be eliminated or an influence can be made on those groups of students who have higher risks of not completing the course or lowering academic performance [22].

However, the opportunities offered by educational analytics to predict students' performance require additional examination. The current study is focused on LMS Moodle. Thus, this LMS is described in detail, paying particular attention to plugins responsible for educational analytics. A visual analysis of student behavior and the results of the examination of log files are also presented. The paper intends to analyze the data obtained using the LMS Moodle, improve the learning process, and reduce the number of underperforming students.

2 Materials and Methods

LMS components offer various opportunities for improving student learning performance and can influence their final grades [23]. A Moodle log consists of the time and date it was accessed, the Internet Protocol (IP) address from which it was accessed, the name of the student, each action completed (i.e., view, add, update, or
delete), the activities performed in different modules (e.g., the forum, resources, or assignment sections), and additional information about the action [24]. All these stored data are beneficial and can be applied for data mining algorithms. Preidys and Sakalauskas [25] notice a trend toward the combined use of data mining learning techniques for the analysis of activity data. Having the data in an LMS provides various opportunities for the use of data mining methods to examine them. Data mining can be useful to explore, visualize, and analyze data with the aim of identifying useful patterns in order to understand students’ learning behavior and feedback [26]. Data mining remains a promising field for the exploration of data from educational settings. A number of institutions have also developed bespoke systems for learning analytics [27]. For example, such developments include commercial business intelligence tools that predict students at risk based on indicator variables within the system. Besides, a popular area of development is applications and plugins that combine the analysis of the training module's databases and the identification of students [28]. These applications and plugins combine data from various sources and allow teachers to communicate with students electronically.

To improve the prediction of students' performance, it is proposed to apply the clustering rules of the developed module of LMS Moodle to data mining [29]. The bulk of data is accumulated in accordance with the quantitative, qualitative, and social activities of students. As a result, it allows defining how the sample of students and the use of various classification algorithms affect the accuracy and intelligibility of predictions on academic performance.

This work presents developments in the field of setting up a highly accessible LMS Moodle for the technical implementation of fully automated virtual lessons. The students' knowledge is assessed automatically basing on compulsory tests [30]. Besides, the study proposes to conduct the process of mining e-learning data step by step, with the guidelines about how to use data mining techniques for mining Moodle data [31]. It is considered relevant since the authors report highly accurate predictions using the Bayesian classifier and the support vector machine.

Within the research, it is also proposed to apply [32] a classification model that predicts the student's ability to achieve excellent results during the study. The model is based on the following data:

- Number of passed or failed quizzes
- Number of messages sent or read on the forum
- Total time spent on assignments, quizzes, and forum
- Final grade received by the student in the course

Based on a sample of 124 first-year students aged 19 to 22 years, an initial and correlation analysis of the Moodle logs data is performed. The analytical calculations are made on logs of students registered in the disciplines of the Department of Physical Education of Tomsk State University (Tomsk, Russian Federation) and the College of Sport Science of Shenyang Normal University (Shenyang, Liaoning Province, PRC), and College of Business of Al Ain University (Dubai, UAE). All study participants remain anonymous and are asked for permission to use the information only about their final grades to build a correlation model. The main limitation of the study is a
relatively small sample of participants since it was possible to involve only three groups of students in each university training course.

3 Results

Moodle (Modular Object-Oriented Dynamic Learning Environment) is an open source online learning management platform known for its ease of use, intuitive controls, and many features offered [33]. Moodle is becoming widely used by numerous schools, universities, and companies wishing to offer distance education to their employees and clients [34]. Moodle provides educational or communicative functions to create an online learning environment: it is an application or creating interactive courses, through the network of interactions between educators, learners, and learning resources. Moodle presents multiple advantages; therefore, it has become a staple platform [35].

The most frequently used Moodle plugins provided with a basic installation are Assignment, Attendance, Choice, Lesson, Page, Quiz, Link, Seminar, Folder, File, Glossary, SCORM Package, Feedback, and Database. The extended version of Moodle 3.4 also includes the Inspire Analytics plugin. One of the distinguishing features of this plugin is a model that predicts students at risk of failure to complete a course based on low engagement in the study process. Besides, it is possible to develop a new plugin or customize an existing one.

Figure 1 illustrates a model where elements and connections supported in the educational process via the mechanisms of the distance learning system are highlighted. Based on the use of LMS Moodle components, the study of the logs in training groups is presented.

![Fig. 1. The Components of the LMS Moodle](http://www.i-jet.org)
The study is aimed at identifying the extent to which individual data obtained from activity logs is a reliable parameter of students' academic success. Moreover, it is of particular interest to determine how the gender of students registered in the disciplines of the Department of Physical Education correlates with the final results.

Figure 2 shows the distribution of logs by the overall grade students achieved in the course. The highest number of logs is achieved by the students with the highest grades, 4 and 5. To investigate if there is any correlation between specific activities on the LMS and student grades, we have performed correlation analysis.

Table 1 examines correlations between achievement over the span of the course (measured by grades) and effort in files, forums, and link usage, as well as the assignments uploaded.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Grade</th>
<th>File usage</th>
<th>Forum usage</th>
<th>Links usage</th>
<th>Assignment uploads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>-</td>
<td>0.361</td>
<td>0.182</td>
<td>-0.01</td>
<td>0.082</td>
</tr>
<tr>
<td>File usage</td>
<td>0.361</td>
<td>-</td>
<td>0.411</td>
<td>0.053</td>
<td>0.19</td>
</tr>
<tr>
<td>Forum usage</td>
<td>0.182</td>
<td>0.411</td>
<td>-</td>
<td>0.132</td>
<td>0.381</td>
</tr>
<tr>
<td>Links usage</td>
<td>0.012</td>
<td>0.053</td>
<td>0.132</td>
<td>-</td>
<td>0.093</td>
</tr>
<tr>
<td>Assignment uploads</td>
<td>0.082</td>
<td>0.19</td>
<td>0.381</td>
<td>0.093</td>
<td>-</td>
</tr>
</tbody>
</table>

The results indicate a statistically significant correlation among students’ grades and the opening of files. The correlation is positive, which indicates that students with a higher frequency of file openings have higher grades. There is a lack of association between grades and other logs in the course. File opening is correlated with activities on the forum, demonstrating that students who are active in forum discussions opened files more often.

In order to analyze the activity dependence on gender, three groups of students (Russian, Chinese and Arabian) are formed. The examination reveals that in both groups female students have a higher number of logs than their male colleagues. Differences between genders is also visible in the average grade received. Female students have a higher average grade than male students (Figure 3, 4).

---

**Fig. 2.** The Frequency of logs by grade [36]
The experience of the teachers of the training course demonstrates that Russian students are much more likely to complete their assignments in the last moments before the deadline. In turn, Chinese students mostly complete the assignments on time and seldom hand in their works at the last moment. Arab students also predominantly completed assignments in advance and passed the work as close to the last moment as possible. However, in general, the increase in the number of logs with the approach of the deadline is confirmed for both groups. For this reason, it is decided to combine the sample. Figure 5 presents the data on the frequency of students' logs before the deadline day.
The diagram demonstrates the logs from the first to the twenty-first academic weeks. The dates for midterm and final assessment are indicated during the eighth and sixteenth weeks of the academic semester. Thus, weekly analysis showed that the highest number of logs appear on the day before test days.

Figure 6 indicates course logs related to the days before test days. Page views occurred mostly between 16:00 and 20:00 on the day before the midterm assessment and after 14:00 on the day before the final test. In both cases, there is a significant number of logs in the late hours of the day. During these times, the most students downloaded test materials and started to study.

Similarly, the time-focused analysis results for the whole course period level are presented in Figure 6. It shows the hours of the day during which students were logged into the course, and the opening is concentrated from 11:30 on. During the afternoon, a number of logs persist, but this decreases in the evening. In the period after midnight, there is no activity on the e-course. However, this does not mean that students were not active at all—some may have performed an offline assignment after downloading it earlier.

Figure 8 presents an analysis of students’ activities with respect to the grade they achieved and the day in the week. Surprisingly, for the students with the highest
grade, most of the course activity was done on the day before lectures, seminars, and tests.

![Graph showing distribution of logs per day in week with respect to achieved grade](http://www.i-jet.org)

**Fig. 8.** Distribution of logs per day in week with respect to achieved grade [33]

## 4 Discussion

Nowadays, the research of e-learning is becoming increasingly large-scale with the tools and methods used to explain student behavior in LMS as fundamental components. Online learning systems are essential for improving thinking skills and introducing innovative forms of mastering courses by students of higher educational institutions. However, at the same time, it is crucial to implement platforms that support mobile technologies so that students can build their learning plans with minimal time and space restrictions [37]. This will positively affect academic performance.

Reusable and flexible learning materials will also improve students' academic performance. For this reason, it is proposed to focus on intuitive and visually appealing tasks. For better tracking and forecasting student performance, it is recommended to use special JavaScript-based applications that significantly save time when processing large amounts of data on the midterm and final assessments [38].

Nowadays, the classical approach of teaching with a steady learning rhythm has lost its relevance. Thus, it is necessary to create a progress-oriented training course that can activate students' interest not only in grades but also in increasing personal potential [39]. Teachers should assign practical tasks via a virtual learning platform to develop student competencies [40]. Due to the fact that e-learning systems can accumulate a large amount of information, analyze student behavior, and help the teacher to identify possible student errors during the training course, the popularity of e-learning continues to grow exponentially.

Even though no correlation between gender and preferences in the selection of educational tools is found, the role of gender and the influence of online tools on performance remains an interesting question for further research [41]. Since the analyzed sample is characterized by a predominance of female students, this matter can be investigated in groups with the same gender composition. Nevertheless, in terms of student performance, the current examination is consistent with existing research, determining factors that correlate with activity and overall success [42].
Now more and more research is being conducted on the empirical results of applying various data mining algorithms aimed at studying learning processes and predicting academic performance. These studies show that a large number of attributes that in one way or another can influence student performance are relevantly inclined to be attracted to a small number of defining categories. Among UAE students, such categories, for example, were demographic data, information about previous student performance indicators, information about courses and teachers, and general student information [43]. These factors can be considered independent of the personality of the student. Equally important are the factors of personal motivation.

In an e-learning environment, student motivation largely influences the quality of the educational process. Data mining techniques provide valuable information for assessing student motivation and the improvement of the learning process [44]. One of the drivers of student motivation is participation in joint projects. It provides team training and enhances the interest in completing complex tasks [45], therefore encouraging students to communicate within the framework of the LMS. Future research on student performance will consider a wider range of factors and review an increased sample of students engaged in various courses.

5 Conclusion

The results of the study prove that learning management systems enable producing new information about student behavior based on their digital profile. First of all, this information provides an opportunity to explore the successes of students in order to generate and implement new types of activities that stimulate positive results. For example, an examination revealed that students who have taken full advantage of the Moodle platform achieve higher grades. The achieved results are potentially beneficial in the early detection of students experiencing difficulties in a course. Both teachers and students benefit from this kind of research, as teachers can identify excellent students for collaboration and students find out how to give greater effort to obtain good results.

In the conducted research, the female students are more active and successful in the course than are the male students. There is a correlation between the number of logs in the e-course and the final grades. The students were most active in the test weeks and, specifically, on the day before the tests. Students can be characterized as “last-minute” students, as they perform their obligations as late as possible in terms of the deadline and are active in the late hours. However, this cannot be generalized because the research was conducted in only one course. Also, the research covered only informatics students. In future research, the analysis will be performed across several courses. Additionally, students from other disciplines will be included.

6 References


7 Authors

Yaquon Zhang has PhD degree and works as an Assistant Researcher at the Shenyang Normal University (China) and Tomsk State University (Russian Federation). yaquz@yahoo.com

Ahmad Ghandour has PhD degree and works at the College of Business, Al Ain University, Abu Dhabi, United Arab Emirates.

Viktor Shestak is a Doctor of Juridical Science. He works as a Professor at the Department of Criminal Law, Criminal Procedure and Criminology in MGIMO University, Russian Federation and also in Military University, Moscow, Russian Federation.

Article submitted 2020-06-01. Resubmitted 2020-07-06. Final acceptance 2020-07-07. Final version published as submitted by the authors.