

ASSESSING THE IMPACT OF STUDENTS' ACTIVITIES IN E-CLASSES ON LEARNING OUTCOMES: A DATA MINING APPROACH

Lan Umek, Nina Tomaževič, Aleksander Aristovnik and Damijana Keržič
Faculty of Administration University of Ljubljana, Gosarjeva ul. 005, 1000 Ljubljana, Slovenia

ABSTRACT

In the paper, we present the results of a case study conducted at Faculty of Administration, University of Ljubljana among 1st year undergraduate students. We investigated the correlations between students' activities in the e-classroom and grades at the final exam. The sample included 92 participants who took part at the final exam in the course Basic Statistics. In the e-classroom, students learn new content for individual self-study is prepared and their knowledge is checked with quizzes. In the empirical study, we used data mining software Orange for two tasks of predictive modelling: The research question was: based on the student's performance on quizzes is it possible to predict if (1) a student will pass an exam, and (2) a student's grade at the exam will be good. The empirical results indicate very strong connection between student's performance on quizzes and their grade at final exam in the course. Moreover, the results pointed out which quizzes, in other words topics, are most important for passing an exam or obtaining better grade.

KEYWORDS

Blended Learning, Moodle, Student's Activities, Quizzes, Students' Performance, Predictive Modelling

1. INTRODUCTION

Higher education institutions all over the world are increasingly adopting blended learning, which combines face-to-face and technology-mediated instruction (Porter, Graham, Spring, & Welch, 2014) with the aim of complementing each other (Graham, Woodfield, & Harrison, 2013). Not only to higher education, blended learning has been introduced to many other sectors, e.g. business (Arbaugh et al., 2009, Bersin, 2004), military (Wisher, 2006), healthcare (Story & Boyd, 2008) etc. The use of Learning Management Systems (LMS) has grown exponentially in the last several years and has come to have a strong effect on the teaching and learning process (Cerezo, Sánchez-Santillán, Paule-Ruiz, & Núñez, 2016; Romero, Espejo, Zafra, Romero, & Ventura, 2013). Moodle is one of the most popular open source LMS. It has a full range of functionalities that other similar programs have, including tools for posting and sharing course information, conducting online discussion, and administration of online quizzes. Moodle is also an environment that facilitates "social constructionist pedagogy"; it provides avenues for students to collaboratively engage in learning and other academic activities (Zhang, 2008). Because learner activities are crucial for an effective online teaching-learning process, it is necessary to search for empirical methods to better observe patterns in the online environment (Neuhauser, 2002).

As online higher education programs began their rapid growth, they created a dynamic tension, spawning ambivalence in some sectors of higher education. A positive side effect of that tension included new learning environments that offered potential for maximizing the effectiveness of contemporary teaching and learning. That movement assumed various labels such as mixed mode, hybrid, and combined, but blended learning emerged as the dominant label for an educational platform that represents some combination of face-to-face and online learning (Moskal, Dziuban, & Hartman, 2013). There are different proportions of both types of learning implemented, e.g. 50-50, 70-30, 60-40. In the case of the faculty, the ratio is 80-20, i.e. for each undergraduate course 20% of its content is held in an e-classroom, including both lectures and tutorials.

While the definition of blended learning is clear and simplistic, its implementation is complex and rather challenging since virtually limitless designs are possible depending on how much or how little online instruction is inherent in blended learning (Garrison & Kanuka, 2004). Diverse instructional models and best practices of blended learning have been reported from simple use of computer or online mediated technologies to full usages of them for a complete course (Park, Yu, & Jo, 2016).

The key stakeholders of blended learning system are institution's management, teachers and students. Each of them tries to attain their goals. The management wishes to improve the efficiency of classroom resources and improve teaching through faculty development. The teachers aim to adopt innovative, student-centred teaching practices. Student' goals are increased flexibility (in time and space) and expanded access, better academic success and enhanced information literacy (Moskal, Dziuban, & Hartman, 2013). However, learners often do not successfully adapt their behaviour to the demands of advanced learning environments, such as LMS (Azevedo & Feyzi-Behnagh, 2011), because it requires more independency by the students (deciding what, how, how much to learn, how much time to invest, when to increase effort etc.) (Azevedo, Cromley, Winters, Moos, & Greene, 2005).

Whatever the motivation to blend, the strategy works best when clearly aligned with the institution's mission and goals and the needs of students, faculty, and institution are simultaneously addressed. A clear vision and strong support are necessities when moving to the blended environment. Only then can this modality not just succeed but become a transformational force for the university (Dziuban, Hartman, Cavanagh, & Moskal, 2011).

When preparing e-classrooms, there are many possible online activities, in which students can be engaged and with which they can be motivated to learn efficiently. These are announcements, links, lecture notes, resources, Q&As, discussion forums, quiz items, group works, Wikis and assignment submissions (Park, Yu, & Jo, 2016). Many researches have already been made investigating the impact of students' involvement in these activities. Romero, López, Luna and Ventura (2013) investigated how different data mining approaches can be used to improve the prediction of first-year computer science university students' final performance based on their participation in an on-line discussion forum which may not only inform the students about their peers' doubts and problems but can also inform instructors about their students' knowledge of the course contents. According to Owston and York (2018) a consensus has emerged in the literature that students, on average, perform modestly better in blended courses when compared to those in completely online or face-to-face courses across a broad range of subject areas and institutional offerings. Furthermore, there is evidence that suggests that the proportion of time devoted to online activities in a blended course is related to course performance.

The above issues became the challenge and the basis for research. In our study, we focused on quizzes – trying to find out whether the performance in solving the quizzes affects students' performance at final exams. It was there not a question how much time the students spend being active in e-classrooms but whether the effort out in studying for quizzes helps them gaining more knowledge and being more successful at final exams.

At the Faculty of Administration, University of Ljubljana, Slovenia, blended learning, implemented in LMS Moodle, has a long tradition – it has been used for over a decade. In order to improve the satisfaction of key stakeholders, i.e. students, teachers and faculty management, regular analyses are performed on a half-year basis by an internal team of researchers. These results enable management and teachers to acquire an insight into the contemporary situation and provide the opportunities for improvements. The purpose of presented study was to examine the correlation between the active involvements of student in prepared activities in e-classroom, specifically in quizzes, and the final exam results. The objective of our study was to find an answer to the following research questions: how the total score achieved at quizzes is related to the final score at the exam and below, whether they affect the value of the final grade. We analysed the data from the undergraduate course Basic Statistics.

2. EMPIRICAL RESEARCH

2.1 Data

Our data sample consisted of students from the 1st year of professional study programme at the Faculty of Administration, University of Ljubljana. This group of students is each year the largest group of students at the faculty. For our analysis, we chose the course Basic Statistics which is held in the first study year and has plenty of activities held in a Moodle e-classroom. Each week students have to study additional content, which is not talked at face-to-face lectures or tutorials. The knowledge of the additional content is then examined on two occasions: solving quizzes in e-classrooms, and at the final written exam.

The course Basic Statistics includes 25 topics for individual self-study. Therefore, student optional answers 25 quizzes during a 15-week semester. To stimulate self-study, the scores from quizzes represent 20% of the final grade and the remaining 80% student gathers from the final exam. Each quiz has the same score, therefore for simplicity; we can assume that student can get from 0 to 100 points on each quiz. The final score at the quizzes is calculated as weighted sum of points achieved, again from 0 to 100 points.

In the study, we investigated if students' performance at quizzes is related to their knowledge at the final exam. We took data of 92 students who participated at final exam. We collected their scores for all 25 quizzes and their performance at the final exam.

Both scores (at the exam and at the quizzes) can take the values from 0 to 100. The mean score at quizzes is higher (73.77) than at the exam (59.89). Standard deviation is also higher for scores at quizzes: some students did not participate seriously at the quizzes (take part just in the beginning of the course) while some of them (contrary to the scores at exam) achieved all 100 points out of quizzes.

A student passed the exam if two conditions were satisfied: (1) at least 51 points at the final exam achieved and (2) at least 51 points from the weighted sum of score from quizzes (20%) and final exam (80%), i. e. $0.8 * \text{final exam} + 0.2 * \text{score on quizzes} \geq 51$. Out of 92 participating students, 58 (63%) passed the exam.

Students who passed the exam obtained a positive grade, from 6 to 10. Since final grade is computed both scores, from quizzes and final exam, the grade can be predicted from the performance on quizzes. However, such prediction would be biased. For the analysis, we "re-graded" the students, where the new positive grade was based on the performance solely at the final exam. Later in the paper we use term "grade" to estimate the performance at the final exam only.

2.2 Methodology

The two research questions were: based on their performance on quizzes, is it possible to predict, if student (1) will pass the final exam, and (2) if student's grade will be good grade on the final exam. These two questions can be answered using methods for statistical classification (Mitchell, 1997; Hastie, Tibshirani, & Friedman, 2017). In terminology of machine learning, both problems belong to class of supervised learning tasks. In such case, each student is a data instance represented with several features (variables), in our case their performance on each quiz (25 features altogether), and a "label", describing membership of group. For the first question student belongs to one of two groups depending on the performance at the exam, either "pass" or "fail". For the second question, we limited the analysis on students who passed the exam, therefore we divided them into groups of "good students" (with grades 8, 9 or 10) and "worse students" (with grades 6 or 7).

In statistical classification, plenty of methods are suitable to tackle such problems. They differ in the interpretability of the results, performance, and speed. For our empirical study we chose naïve Bayesian classifier, logistic regression, k-nearest-neighbours (kNN), support vector machines (SVM) with linear kernel, and random forest (Mitchell, 1997; Hastie, Tibshirani, & Friedman, 2017). The results of naïve Bayesian classifier and logistic regression are usually more interpretable while the interpretability of the other three methods is harder. In the results section, we will see that naïve Bayesian classifier outperformed other methods, so we will later present its performance using nomograms (Možina, Demšar, Kattan, & Zupan, 2004).

In machine learning, quite of measures can be used to estimate the performance of methods (statistical models). Most of them can be defined using a confusion matrix (Table 1). In the confusion matrix the abbreviation TP (True Positive) stands for the number of students who passed the exam and for which the statistical model also predicted passing the exam. Similarly, the TN (True Negative) represents the number of students who failed the exam and the prediction of the model was negative, i.e. student fails the exam. Analogically, we can define quantities FN and FP.

Table 1. Confusion matrix for pass/fail an exam

		True condition	
		Passed	Failed
Predicted by statistical mode	Passed	TP (True positive)	FP (False positive)
	Failed	FN (False negative)	TN (True negative)

The proportion of correctly classified students is therefore $(TP+TN)/n$, where n stands for the total number of students, in our case $n = 92$. This measure is known as classification accuracy and is denoted as CA. The other two commonly used measure are precision and recall (sensitivity). Precision is defined as $TP / (TP + FP)$. In other words, to compute precision we have to limit ourselves to students, for which the statistical model predicted that they will pass an exam ($TP + FP$). The precision is therefore the probability that they actually passed an exam. The other measure, frequently used, is recall (sensitivity). It is defined as $TP / (TP + FN)$. To compute recall we have to limit only on students who passed the exam ($TP + FN$). Recall is therefore the probability that statistical model will predict that they will actually pass the exam.

There is, however, one important measure of performance, which cannot be defined using a confusion matrix. It is called Area under Receiver-operating-characteristic (ROC) curve. It is usually denoted as AUC (standing for area under curve). For our case, it represents the probability that a statistical model will correctly distinguish between a student, who will pass an exam from the student who will fail. More precisely. We deal with two students: one will pass an exam, the other will fail. A statistical model will distinguish between them if the estimated probability of passing will be higher for the student who passed the exam. The AUC measure is therefore computed using all pairs of students where exactly one student in the pair passed an exam. The AUC measure above 0.75 is generally considered as high (Murphy-Filkins, Teres, Lemeshow, & Hosmer, 1996).

All measures mentioned above were estimated using 10-fold cross-validation to prevent overfitting. That means that the data set was split to 10 subsets of students of approximately equal size. In each fold of cross-validation a statistical model fitted the parameters on 9 subsets and tested predictive accuracy on the remaining one. The procedure was repeated for each all subsets, and the performance measure was averaged among 10 folds. The whole analysis was done in the open source data mining software Orange (Demšar et al., 2013).

2.3 Results

2.3.1 Passing vs. Failing the Exam

The results of the study for the first question (Is it possible to distinguish between a student who passed the exam from a student who failed?) are shown in the Table 2. With an exception of logistic regression, all other classification models produced AUC score above 0.75. The other measures of their performance (CA, precision, recall) are high as well which means that the answer to our research question is positive. The method that outperformed other methods was naïve Bayesian classifier.

Table 2. Performance of five classification models for the first research question – distinction between students who passed and who failed the exam

Method	AUC	CA	Precision	Recall
Naive Bayes	0.81	0.77	0.88	0.74
Logistic Regression	0.64	0.62	0.71	0.67
kNN	0.78	0.73	0.75	0.86
SVM	0.77	0.67	0.73	0.76
Random Forest	0.75	0.70	0.75	0.78

To get more information about the impact of quizzes on the passing the exam we use the nomogram representation of naïve Bayesian classifier. Top five (for prediction) quizzes (out of 25) are show in Figure 1. These quizzes cover topics on hypotheses testing, statistical terminology, correlation and regression and probability. The last quiz covers different topics from the first part of the course and is the first review quiz in e-classroom. The nomogram additionally explains how the prediction works. If we do not know anything about a student, their prior probability of passing the exam is 63% (58 out of 92). This situation is marked with white dot at the bottom. If we know their performance on quizzes, the probability of passing may change. For illustration: for a selected student whose performance on the top five ranked quizzes is marked with solid dots, the probability of passing increases up to 95% (solid dot at bottom).

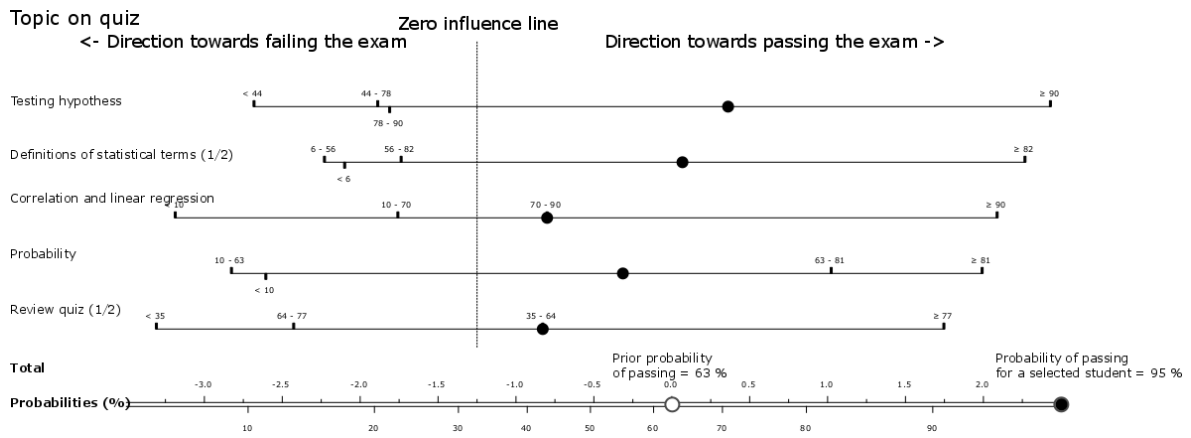


Figure 1. Performance of naïve Bayesian classifier for the first research question – distinction between students who passed and who failed the exam

The top five ranked quizzes play an important role at the course Basic Statistics. The topic “Testing hypotheses” is one of the most difficult topics in the entire course. Students require knowledge about probability, sampling, and estimation of statistical parameters. The other difficulty in this topic is formalizing the research questions to statistical hypotheses where the real-world cases have to be represented in a strict mathematical way. Therefore, it is not surprising that the performance on this quiz plays a vital role for passing the exam.

The second most influential quiz covers statistical terminology (“Definition of statistical terms (1/2)¹”). This quiz is also related to several topics which cover the first part of the course. Students require knowledge about different terms and have to compute several quantities to show that they understand them. Since this quiz covers various topics, it requires skills of recalling and connecting the knowledge, which students had to acquire by then.

The other three quizzes presented in Figure 1 are also more complex and require combining knowledge from various topic. With exception of the “Review quiz”, they occur at the end of the course.

2.3.2 Good vs. Worse Students

The second research question which we have asked is if it is possible to distinguish between a student with good grade (8, 9 or 10) and a student with worse grade (6 or 7). The performance of five methods of classification is presented in Table 3.

¹ The »1/2« means that there are two quizzes on the topic »Definitions of statistical terms« and that the presented one is the first of them (covers the first part of the course).

Table 3. Performance of five classification models for the second research question – distinction between “good” and “worse” students

Method	AUC	CA	Precision	Recall
Naïve Bayes	0.75	0.64	0.65	0.69
Logistic Regression	0.64	0.57	0.58	0.58
kNN	0.73	0.62	0.63	0.65
SVM	0.70	0.74	0.73	0.73
Random Forest	0.70	0.72	0.71	0.71

From Table 3 we see that AUC measure is mostly above 0.7, again with an exception of logistic regression, but above 0.75 only for naïve Bayesian classifier. Other measures of performance are also high. Since the naïve Bayesian classifier again outperformed other methods we will present the prediction using nomogram (Figure 2).

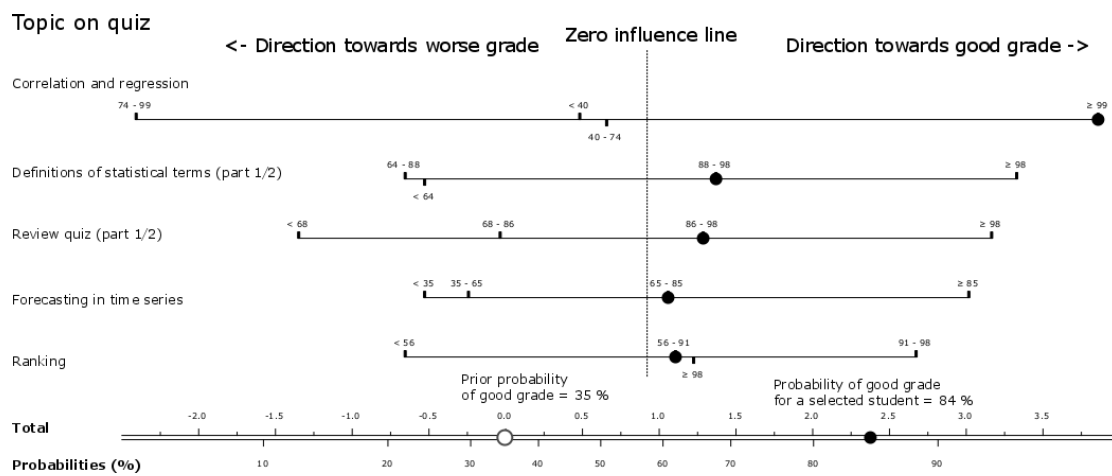


Figure 2. Performance of naïve Bayesian classifier for the second research question – distinction between “good” and “worse” students

The top five ranked quizzes in this case are very similar to the previous case. The most influential quiz covers correlation and regression analysis. It contains several questions where computation has to be done by hand. Students at the course Basic Statistics are used to analyse data in Excel so computation without concrete data by hand is considered as a more advanced skill. Although most of the exercises at the final exam can be solved using Excel, some of them require computation by hand. Only the best students can handle such exercises and one of the possibilities to prepare for them is by taking the quiz “Correlation and regression”. The same principle holds for the fourth most influential quiz, which covers the topic “Forecasting in time series”.

The other influential quizzes have been discussed previously, so the same arguments hold for them. The only exception among the top five quizzes is a topic on ranking. We cannot find reasonable explanation why this quiz occurs on the list so high.

3. CONCLUSION

The results of the empirical study confirmed that the prompt self-study in e-classrooms in terms of prepared content and quizzes has a positive impact on student’s performance. It increases their chances of passing the final written exam at the course Basic Statistics and gives the chances of getting better grades. The empirical results pointed out quizzes that covered topics, which require combination and linking knowledge among

various themes from the course. Good performance on such topics has the strongest impact on increasing chances of passing the final exam and obtaining better grade.

The main limitation of our research is that it is a case study and therefore cannot be generalized, which can also be found at similar case study (Romero, López, Luna, & Ventura, 2013). In the future, we are planning to perform the same study for other courses with similar obligatory activities in e-classroom. Such analysis will enable identification of courses where the online learning part of the blended learning is beneficial. As we observed, course Basic Statistics is one such example. We are also planning to track the students' behaviour in e-classroom, such as time spending on different resources and activities, and try to link patterns of their behaviour to their performance.

In the future, we will establish a framework for analysing several courses by investigating the students' behaviour at different e-classrooms and linking that to their performance at final exams.

REFERENCES

- Arbaugh, J. B., Godfrey, M. R., Johnson, M., Pollack, B. L., Niendorf, B., & Wresch, W., 2009. Research in online and blended learning in the business disciplines: Key findings and possible future directions. *The Internet and Higher Education*, Vol. 12, No. 2, pp. 71–87.
- Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C., & Greene, J. A., 2005. Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science*, Vol. 33, No. 5-6, pp. 381–412.
- Azevedo, R. & Feyzi-Behnagh, R., 2011. Dysregulated learning with advanced learning technologies. *Journal of e-Learning and Knowledge Society*, Vol. 7, No. 2, pp. 9–18.
- Bersin, J., 2004. *The Blended Learning Book*. San Francisco, John Wiley & Sons.
- Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C., 2016. Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education*, Vol. 96, pp. 42–54.
- Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., & Zupan, B., 2013. Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, Vol. 14, pp. 2349–2353.
- Dziuban, C., Hartman, J., Cavanagh, T., & Moskal, P., 2011. Blended courses as drivers of institutional transformation. In A. Kitchenham (Ed.), *Blended learning across disciplines: Models for implementation*. Hershey (pp. 17–37). PA: IGI Global.
- Garrison, D. R. & Kanuka, H., 2004. Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, Vol. 7, pp. 95–105.
- Graham, C. R., Woodfield, W., & Harrison, J. B., 2013. A framework for institutional adoption and implementation of blended learning in higher education. *The Internet and Higher Education*, Vol. 18, pp. 4–14.
- Hastie, T., Tibshirani, R., & Friedman, J., 2017. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd Edition). Springer Series in Statistics.
- Mitchell, T. M., 1997. *Machine Learning*. McGraw-Hill Science/Engineering/Math.
- Moskal, P., Dziuban, C., & Hartman, J., 2013. Blended learning: A dangerous idea? *The Internet and Higher Education*, Vol. 18, pp. 15–23.
- Možina, M., Demšar, J., Kattan, M., & Zupan, B., 2004. Nomograms for visualization of naïve Bayesian classifier. In Jean-F. Boulicaut, F. Esposito, F. Giannotti, & D. Pedreschi (Eds.), *Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases* (PKDD '04) (pp. 337–348). Springer-Verlag New York, Inc., New York, NY, USA.
- Murphy-Filkins, R., Teres, D., Lemeshow, S., and Hosmer, D.W. 1996. Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model: How to distinguish a general from a specialty intensive care unit. *Critical Care Medicine*, Vol. 24, No. 12, pp.1968–1973,
- .Neuhauser, C., 2002. Learning style and effectiveness of online and face-to-face instruction. *American Journal of Distance Education*, 16(2), pp. 99–113.
- Owston, R. & York, D. N., 2018. The nagging question when designing blended courses: Does the proportion of time devoted to online activities matter? *The Internet and Higher Education*, Vol. 36, pp. 22–32.
- Park, Y., Yu, J. H., & Jo, I-H., 2016. Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute. *The Internet and Higher Education*, Vol. 29, pp. 1–11.
- Porter, W. W., Graham, C. R., Spring, K. A., & Welch, K. R., 2014. Blended learning in higher education: Institutional adoption and implementation. *Computers & Education*, Vol. 75, pp. 185–195.

- Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S., 2013. Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, Vol. 21, No.1, pp. 135–146.
- Romero, C., López, M.-I., Luna, J.-M., & Ventura, S., 2013. Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, Vol. 68, pp. 458–472.
- Story, A. & Boyd, T., 2008. LMS and eLearning Assessments for Healthcare Compliance. In C. Bonk, M. Lee, & T. Reynolds (Eds.), *Proceedings of E-Learn 2008--World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (p. 1289). Las Vegas, Nevada, USA: Association for the Advancement of Computing in Education (AACE).
- Wisher, R. A., 2006. *Blended learning in military training. Handbook of blended learning: Global perspectives, local designs*, pp. 519–532.
- Zhang, L. Z., 2008. MOODLE for Integrative Learning. In C. Bonk, M. Lee, & T. Reynolds (Eds.), *Proceedings of E-Learn 2008--World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (p. 2070). Las Vegas, Nevada, USA: Association for the Advancement of Computing in Education (AACE).