Expanding Learning Scenarios
Opening Out the Educational Landscape
Proceedings of the European Distance and E-Learning Network 2015 Annual Conference
Barcelona, 9-12 June, 2015
ISBN 978-615-5511-04-2
ISSN: 2707-2819
doi: https://doi.org/10.38069/edenconf-2015-ac-0029

WEB-BASED ANALYSIS OF STUDENT ACTIVITY FOR PREDICTING DROPOUT

Anat Cohen, Tel Aviv University, Israel

Introduction

Persistence in learning processes is perceived as a central value in education (Horowitz, 1992); therefore, dropout from studies is a prime concern for educators (Barefoot, 2004). Since the increase in student usage of online learning materials on course websites, as well as online courses (Allen & Seaman, 2014; Parker, Lenhart, & Moore, 2011), it is essential to address the dropout issue in a wide array of configurations from web-supported learning to fully online courses. Additional tools and strategies must be developed to allow instructors or other educational decision makers to quickly identify at-risk students and find ways to support their learning in the early stages, before they actually drop out. The ability to detect these students during the semester, and not at the end of the course, can also serve as a basis for the development of appropriate assistance mechanisms which will enable those students to complete the course and even to fulfil the curriculum for their degree.

In order to meet the challenge of identifying dropouts early, it is possible to use the large databases that are created automatically in the Learning Management Systems (LMSs) servers. These databases contain enormous amounts of data relating to learning processes and learner behaviours on course websites. This data can be analyzed in order to evaluate the learning processes (Horizon Report, 2014); improve teaching and learning; optimize the construction of learning systems and their operation (Ai & Laffey, 2007; Romero & Ventura, 2007; Lu, Yu & Lin, 2003); and can even be used to predict potential dropouts and failures (Macfadyen & Dawson, 2010; Lykourentzou et al., 2009; Nistor & Neubauer, 2010).

The presented study focuses on identifying at-risk learners who might drop out from specific courses or from degree studies in general, based on the analysis of student activity in the course website data which is accumulated in the log files of the LMS, Moodle. Student data from six courses in the disciplines of exact science were analyzed. In these courses, the dropout rate is usually very high. Furthermore, students who fail these courses occasionally drop out of their degree study as well. The information obtained from the presented analysis may assist teaching staff and other institutional mechanisms in supporting and retaining their students. This analysis enables the instructors to monitor student activity on the website throughout the learning process during a course, not only at the end, in order to detect students who are not using the website, or students with unexpected behaviour. The main purpose of getting this information is to allow the instructor to contact students who could potentially abandon their studies, and to understand the reasons why. Additionally, this analysis will allow instructors to understand the scope of course material usage and patterns so that they can make improvements. This analysis can be used by educational decision makers too, according to ethical guidelines, since it will allow them to see the information on a campus level; to identify students with potential for dropout in relation to different faculties, instructors, types of courses, or other chosen criteria. Thus, intervention programs for potential dropouts can be initiated and action can be taken for instructors that are observed to have high dropout rates from their courses.

Methodology

Research aims and questions

The aim of the study is to identify learners who are at risk for dropping out from specific courses or from degree studies through analyzing the large web log files accumulated in the LMS by the specific courses. The study addressed the examination of student activity on course websites that may be associated with dropout. Following that variables and measurements that may alert to dropout were developed and the correlations between them and the completion of the course and student status regarding continued studies the following year were tested. Thus, the research questions were: i) What are the variables and measurements regarding student activity on course websites that may alert to dropout? ii) Is there a correlation between the defined variables and measurements that may alert to dropout and successful completion of a course? ii) Is there a correlation between the variables and measurements that may alert to dropout and the termination of studies the following academic year?

Research field

The study was conducted on six courses in the exact science faculty, in the fields of mathematics and statistics taught at a large university during one semester in 2013 (N = 718 students; 6 instructors). All courses were using the Moodle LMS and their websites contained varied contents related to the learning issues, especially exercises. These courses were chosen due to the fact that in these courses the dropout rate is very high. Furthermore, students who fail these courses occasionally drop out from their degree study as well.

Method and procedures

Our main hypothesis is that academic students heading towards dropout will first become absent from course websites. Essentially, early traces of dropouts will be manifested first on course websites; hence, they can be traced in the LMS log files.

During the study, different variables that define students' activities on the website were calculated, and the log-files of six courses in the fields of mathematics and statistics were retrieved and analyzed. Different variables that may alert to dropout in real-time were calculated and a set of alert variables were developed for identifying these at-risk learners during preliminary stages of the course. Then, the correlation among the student's activity in the system, successful completion of the course, and the continuation of studies in the following year were tested.

The study was conducted in three stages.

Stage 1

Data collection – While using the course websites Moodle automatically accumulates a vast amount of data regarding student activity (such as the number of actions, their types, timing, and frequency) in its server web logs. Through web-mining techniques, access to the server database was enabled and data regarding 718 student activities in six selected courses was retrieved. The data was organized in a file in which each action is represented by: Student ID, the date and time of the activity, and its type (such as viewing content, uploading an item, sending a post to a forum, and submitting a quiz or an assignment, etc.). 170,445 actions were demonstrated in these log files. Notably, we were committed to protecting student privacy. Data was handled very carefully according to ethical guidelines.

Stage 2

Data analysis was performed on the six courses in order to identify the students at risk for dropping out – student activity variables were calculated and the index, which represents alerts for an exception activity for each student, was constructed.

Three groups of variables were defined (Table 1): the first group measures the student usage of the course website. The variables in this group relate to course level (V_1-V_4); the second group measures the intensity of the student activity on the website, aiming to identify any unexpected or extraordinary changes in activity. The variables in this group are related to student level (V_5-V_8); the third group of variables refer to student grades and academic status – student final grades and learning status the following year (V_9-V_{10}), which shows if the student did or did not drop out.

Table 1: Groups of variables measuring student activities

Variable name	**Variable description**
student usage of course website variables (course level)	
V_1: No. of all student actions during the semester	Total number of student hits (total number of records for all students in the log file).
V_2: No. of all student actions during specified period of time (e.g., per month)	Total number of student hits recorded each month during the semester (total number of records for all students in the log file for each month).
V_3: Average of all student activity days	Average number of days in which the actions were performed by all students.
V_4: Average of all student activity (hits) in the course in a specific month	Average student hits each month during the semester (Average of records for all students for each calendar month).
Variables for measuring the intensity of student activity on a course website (student level)	
V_5: No. of actions for the entire semester	Total number of student hits (total number of records for a specific student).
V_6: No. of actions in specified period of time (e.g., per month)	Total number of student hits every month during the semester (total number of records for a specific student in the log file for each month).
V_7: Relative No. of actions in relation to other students in a specified period of time (e.g., per month)	The ratio between the number of activities for a specific student and the average of all students' activities (hits) for a course. Each month during the semester, a relative score was calculated for each student.
V_8: No. of activity days	Total number of days in which the actions were performed.
Student status variables (student level)	
V_9: Student final grade	Results at the end of the course: passed, failed, or did not complete (meaning studies were abandoned before the course ended).
V_{10}: Learning status of student the following academic year	The distinction between an active student who continued his/her studies, and an inactive student who abandoned his/her degree studies.

Using variables from the first three groups, a fourth group of variables was calculated. This fourth group contains alert variables for unexpected student activity that is represented in the alert index (Table 2). A set of measures (V_{11}-V_{14}) were defined that reflect student inactivity or an unexpected change in student activity intensity over time, for example, a month with inactivity or a month with low activity (in this case, lower than 70% of the classroom activity average that month). Consequently, student activity intensity on course websites each month was reflected in three main variables: amount of activity during this period of time (V_2), the ratio between the student activity and the average of all student activity during the course (V_3), and the number of days during which the activity occurred (V_4).

Table 2 Variables for alerting inactivity or unexpected change in student activity

Variable name	**Variable description**
Variables for alerting inactivity or unexpected change	
V_{11}: No. of alerts for monthly inactivity	Each month in which inactivity was identified a warning flag for monthly inactivity was turned on. For each student, the number of warning flag months was calculated.
V_{12}: No. of alerts for monthly decreases in activity compared to the class average (low average of actions)	When the number of student actions was at least 70% lower than the class average that month, an alert flag was turned on for low average activity. For each student, the number of months in which this warning flag was turned on was calculated.
V_{13}: The calendar month in which the inactivity alert appears	The name of the first calendar month during the course in which the alert flag for inactivity appeared the first time.
V_{14}: The calendar month in which the low activity alert appears	The name of the first calendar month during the course in which the alert flag for low activity appeared the first time.

Our working hypotheses regarding the identification of at-risk students before dropout through the students' activity in the course website is summarized in Figure 1.

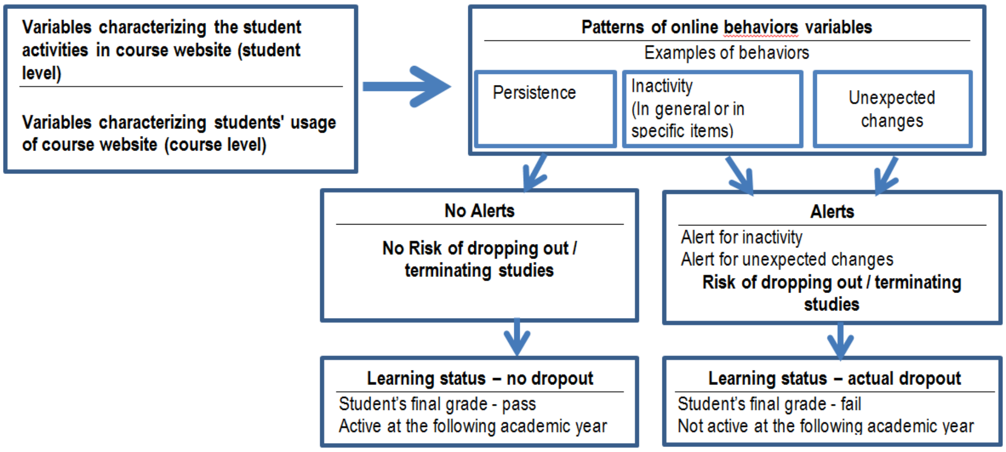
Variables characterizing the student activities in course website (student level)
Variables characterizing students' usage of course website (course level)
Patterns of online behaviors variables
Examples of behaviors
Persistence
Inactivity (In general or in specific items)
Unexpected changes
No Alerts
No Risk of dropping out / terminating studies
Alerts
Alert for inactivity
Alert for unexpected changes
Risk of dropping out / terminating studies
Learning status – no dropout
Student's final grade - pass
Active at the following academic year
Learning status – actual dropout
Student's final grade - fail
Not active at the following academic year

Figure 1. Working hypotheses for predicting dropout through the students' activity in the course website

At the end of this stage, the analysis for predicting dropout was performed. Three components are included in the analysis: the first component is the input file which contains student activity variables retrieved from the LMS log files; the second component is a file containing the measurable variables which may provide alerts for unusual activity; and the third component is the output file of students who were marked with flags as potential dropouts, generated based on monthly changes in student activity, when unexpected or unusual activity was shown regarding specific students (Figure 2).

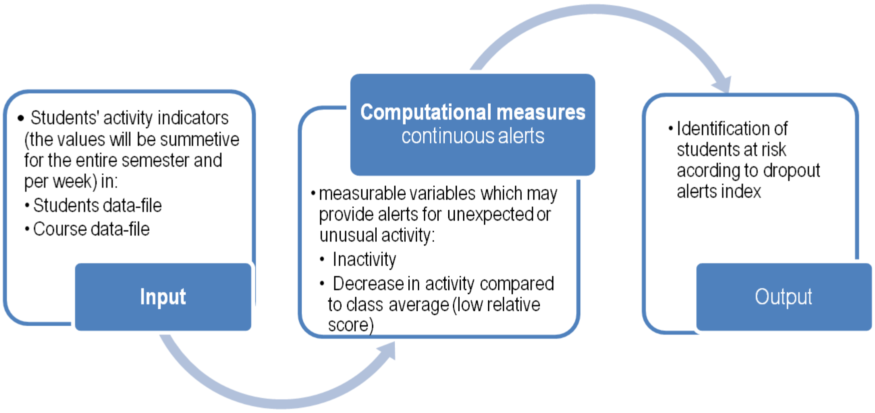
• Students' activity indicators (the values will be summetive for the entire semester and per week) in:
• Students data-file
• Course data-file
Input
Computational measures
continuous alerts
• measurable variables which may provide alerts for unexpected or unusual activity:
• Inactivity
• Decrease in activity compared to class average (low relative score)
• Identification of students at risk acording to dropout alerts index
Output

Figure 2. Web-based analytics tool for predicting dropout

Stage 3

Examination of correlations – The alert variables for each student were analyzed and tested against two different variables: the completion status of the course – whether the student passed, failed, or did not complete the course; and the student's academic status the subsequent academic year – if he/she continued studies at the university. Correlation analysis between alerts and student status in regards to course completion and continued studies the following year were performed.

Results

Variables and measurements regarding student activity on course websites that may alert to dropout

Variables and measurements regarding student activity on course websites that may alert to dropout based on student activity intensity were identified. Student activity intensity can be analyzed relating to different features available to students through the course website. Looking at one student's activity alone is not enough. A comparison between the activity of a specific student compared to the rest of a class is required, in order to understand whether the observed activity is expected or not, according to the course requirements.

While exploring the student activity on the websites, different kinds of patterns were found. Figures 3 and 4 show examples of different behaviours that may indicate a student is at risk for dropping out. These figures show changes in the intensity of a student's activity in relation to the average intensity of the students in the course.

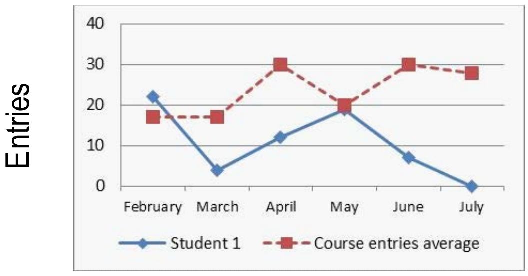

Figure 3. Activity intensity of Student1 compared to average intensity

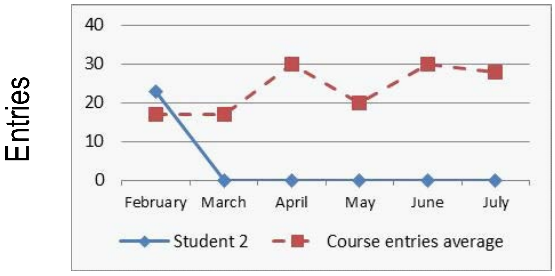

Figure 4. Activity intensity of Student2 compared to average intensity

Testing the defined variables and measurements that may alert to dropout

Most of the students performed between 100 and 300 activities on the website during their course. For each student, a relative score was calculated, which was the ratio between the number of the student's activities performed and the average activities of all students in the course for a period of one month. Examining the distribution of the relative scores by quartiles shows that most students are located in the 4th quartile with the score of between 0.83-1.77. Most students are active for 30-60 days during the semester.

In order to discover valuable information concerning identification of dropout's early traces on course websites, the websites should be meaningfully integrated in the course learning processes. In this study, no significant learning processes were conducted on three course websites, thus, no valuable information was provided concerning the potential drop out of students from those courses. For this reason, the findings of only three courses are presented in this paper. Analysis of the three courses shows that 41 students out of all students in those courses (n = 362) were flagged as potential dropouts (Table 3). Eventually, a large percentage of the students whom our analysis had flagged as at-risk did not finish the course and/or degree. 25 students actually dropped out from the courses or were not active the following academic year, meaning that the prediction was 66% accurate. Meaning that, the changes in a student's activity during the course period could identify a learner at risk in real time, before he/she drops out.

Table 3: Summative results regarding at-risk students and actual dropout

Course ID	**No. of students in courses**	**No. of students identified as at-risk**	**No. of students who dropped out/were inactive the following year**	**Prediction %**
Course 1	124	10	16	63%
Course 2	120	4	5	80%
Course 3	118	11	20	55%
Total	**362**	**41**	**25**	**66%**

Correlation between the alert variables for student dropout, completing the course successfully, and learning status the following year

While examining the correlations between the intensity and student activity volume for the entire semester (using variables such as total number of student activities for the semester; the average number of student activities in relation to the average number of all student activities for the semester; and number of days in which students were active) and completing the course successfully, no significant correlations were found. Similarly, no significant correlations were found between those variables and continuing studies the following year. However, while examining the student activity every month and testing the correlations between the alerts for student dropout, completing the course successfully, and the status of the student's studies the following academic year, significant strong and positive correlation was found (Table 4). Significant correlations were found between completing the course successfully, the number of alerts for inactivity ($r = .533$), and the alerts for low relative activity ($r = .433$). Significant correlations were found as well between the two kinds of alerts and the status of the student's studies in the following academic year. Positive significant correlations were found between continued study status, the number of alerts for monthly inactivity ($r = 0.483$), and the alerts for low relative activity ($r = 0.466$).

Table 4: Correlation analyses between the alerts for and completing the course successfully ($N_{students}$=718)

	No. of alerts for inactivity	**No. of alerts for low relative activity**
Completing the course successfully	.533**	.433**
Learning status	.483**	.466**

** p < 0.01

The results of this study show that our main hypothesis is supported. Students most likely to drop out of a course will first become absent from the course website. Essentially, a dropout's early traces will be manifested first on course websites; hence, they can be traced in the LMS log files. These findings are preliminary and may be used as a basis for developing a web-based analyzing tool for predicting potential dropouts.

Discussion

Students continuously leave hidden traces of their learning activity in the LMS course log files. Some of these traces might be identified as initial stages in the dropout process. A web-based analysis for predicting student dropout was presented in this study using data from courses considered to be difficult and characterized by high dropout rates; in many cases these dropouts also quit their degree program entirely. By using the presented analysis, crucial information regarding student activity was provided for identifying at-risk students before they drop out. Unlike previous studies, which dealt with data analysis of fully online courses (Nistor & Neubauer, 2010; Macfadyen & Dawson, 2010), this study analyzed activities of web-supported learning in which the academic instruction is accompanied by a course website. The findings of this study show that certain characteristics of student web activity can indicate potential risk of course dropout or even degree study termination. At-risk students might show changes in their behaviour, unexpectedly, such as inactivity on the course website or a reduced amount of activity in relation to the rest of the class. Furthermore, a high percentage of predicted dropout was shown and there was a strong correlation between the dropout alerts, course completion status, and continued study status. Students who were flagged using the analysis did not complete their course and/or discontinued their studies the following year.

From these findings, several preliminary conclusions can be drawn: firstly, the analysis of student activity in course web logs can serve as a tool for predicting student dropout from the course or even from degree studies. However, analyzing the web logs at the end of the learning process with summative measures is insufficient (Lykourentzou, et al., 2009). Analysis of overall usage data with measures such as the monthly average of student activity during the semester is also not enough to create timely flags and point to the potential for dropout. When considering the changes in student activity during the course study period, it is possible to perceive the potential for dropout from the course and studies in general. When the data was examined on a monthly basis during the semester and potential dropouts were flagged, a large percentage of these students did not finish the course and/or degree. Consequently there is significance for analyzing the student activity data over time and measuring the changes revealed in their behaviour. One explanation for this finding may be related to the fact that at the beginning of a course many students enter the course website frequently, perform a reasonable amount of actions or even above average, and at more advanced stages of the course, for various reasons, they stop or greatly reduce their volume of activity; and this change indicates a potential for not completing the course (Hwang & Wang, 2004; Hershkovitz & Nachmias, 2011).

The proposed analysis focuses on predicting learner dropout during the preliminary stages of academic course study by using an early warning system. Future research will extend the analysis to examine significant additional variables that may affect the predictability and potential for abandonment of studies: variables related to course characteristics, such as course type (compulsory or elective); types of material on the course website; and variables related to different types of student activity on the website, such as the nature of student activity (active or passive). In so doing, a dashboard-like, web-based visualization tool will be developed for educators. It will be easy to use and will provide information at various levels: course, department and campus wide.

References

1. Ai, J. and Laffey, J. (2007). Web mining as a tool for understanding online learning. In *Merlot Journal of Online Learning and Teaching, 3(2),* (pp. 160-169).

2. Allen, E. and Seaman, J. (2014). *Grade Change: Tracking Online Education in the United States.* Babson Survey Research Group and Quahog Research Group, LLC.

3. Barefoot, B. (2004). Higher education's revolving door: Confronting the problem of student drop out in US colleges and universities. In *Open Learning: The Journal of Open, Distance and e-Learning, 19(1),* (pp. 9-18).

4. Hershkovitz, A. and Nachmias, R. (2011). Online Persistence in Higher Education Web-Supported Courses. In *The Internet and Higher Education, 14(2),* (pp. 98-106).

5. Horowitz, T. (1992). Dropout-Mertonian or reproduction scheme? In *Adolescence, 27(106),* (pp. 355-451).

6. Hwang, W.-Y. and Wang, C.-Y. (2004). A study of learning time patterns in asynchronous learning environments. In *Journal of Computer Assisted Learning, 20(4),* (pp. 292–304).

7. Johnson, L.; Adams Becker, S.; Estrada, V. and Freeman, A. (2014). *NMC Horizon Report: 2014 Higher Education Edition.* Austin, Texas: The New Media Consortium.

8. Lu, J.; Yu, C.-S. and Liu, C. (2003). Learning style, learning patterns, and learning performance in a WebCT-based MIS course. In *Information & Management, 40(6),* (pp. 497-507).

9. Lykourentzou, I.; Giannoukos, I.; Nikolopoulos, V.; Mpardis, G. and Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. In *Computers & Education, 53,* (pp. 950–965).

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

11. Nistor, N. and Neubauer, K. (2010). From participation to dropout: Quantitative participation patterns in online university courses. In *Computers & Education, 55,* (pp. 663-672).

12. Parker, K.; Lenhart, A. and Moore, K. (2011). *The digital revolution and higher education: College presidents, public differ on value of online learning.* Washington, DC: Pew Research Center Social & Demographic Trends.

13. Romero, C. and Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. In *Expert Systems with Applications, 33(1),* (pp. 135-146).