

Analyzing students online learning behavior in blended courses using Moodle

Rosalina Rebucas Estacio and Rodolfo Callanta Raga Jr
*College of Computer Studies and Engineering, Jose Rizal University,
Mandaluyong City, Philippines*

Abstract

Purpose – The purpose of this paper is to describe a proposal for a data-driven investigation aimed at determining whether students' learning behavior can be extracted and visualized from action logs recorded by Moodle. The paper also tried to show whether there is a correlation between the activity level of students in online environments and their academic performance with respect to final grade.

Design/methodology/approach – The analysis was carried out using log data obtained from various courses dispensed in a university using a Moodle platform. The study also collected demographic profiles of students and compared them with their activity level in order to analyze how these attributes affect students' level of activity in the online environment.

Findings – This work has shown that data mining algorithm like vector space model can be used to aggregate the action logs of students and quantify it into a single numeric value that can be used to generate visualizations of students' level of activity. The current investigation indicates that there is a lot of variability in terms of the correlation between these two variables.

Practical implications – The value presented in the study can help instructors monitor course progression and enable them to rapidly identify which students are not performing well and adjust their pedagogical strategies accordingly.

Originality/value – A plan to continue the work by developing a complete dashboard style interface that instructors can use is already underway. More data need to be collected and more advanced processing tools are necessary in order to obtain a better perspective on this issue.

Keywords Moodle, Online behaviour, Activity logs, LMS access, Vector space model

Paper type Research paper

1. Introduction

Many higher educational institutions in the Philippines have started to implement web-based learning environments capable of delivering online education in a blended learning academic setting. Blended learning, also called hybrid learning or mixed method learning involves both face-to-face classroom style instruction as well as the use of online methods (Prasad, 2015). Researchers are unanimous in stating that the blended learning strategy enables educational institutions to implement a more learner-centered approach to teaching where learners are given space and flexibility to indulge with effective learning activities (Alonso *et al.*, 2005; Hughes, 2007; Roby *et al.*, 2013). To implement blended learning, a web-enabled tool or learning management system (LMS) is often utilized to design a particular course in asynchronous mode. Moodle, a free open-source software package used by educators to create online courses (Borromeo, 2013; Maila *et al.*, 2014;

© Rosalina Rebucas Estacio and Rodolfo Callanta Raga Jr. Published in the *Asian Association of Open Universities Journal*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The authors would like to express their sincerest gratitude to the Office of the Research Director of Jose Rizal University for all the support extended for the accomplishment of this work.



López *et al.*, 2016). It provides a modular design that makes it easy to add contents that will engage learners and supports a social constructionist pedagogy style of teaching (Romero *et al.*, 2008).

The use of Moodle has been cited by several literatures as an effective tool for teacher course administrative tasks (Perkins and Pfaffman, 2006); improving student inquiry and critical analysis skills (Regueras *et al.*, 2011); inducing self-directed learning (Woltering *et al.*, 2009); as well as promoting collaborative activities (McLuckie *et al.*, 2009). However, even with many cited benefits of using Moodle particularly in higher education institutions, there are still factors that need to be looked upon to ensure its effective implementation. One of the most difficult factors has to do with assessing how the utilization of the various features of Moodle within the online environment affects the overall course performance of students. Are there patterns of utilization that can lead to better success in learning and higher course grade? This aspect can be analyzed by looking into the sort of activities that students often engage with. Due to the nature of the design of Moodle, it is able to routinely collect detailed activity data on students through its log files. Unfortunately, because of the inherent difficulties in handling these enormous log data files generated online by students; teachers would not agree to analyze them manually. Traditional assessment techniques, on the other hand, do not provide appropriate measures on the kind of skills that students develop while interacting with the features of the Moodle environment (Macfadyen and Dawson, 2010).

Fortunately, in the last few years, data mining technologies have been making a lot of headway in capturing and analyzing massive amounts of data (Romero *et al.*, 2008).

These technologies utilized techniques adopted from machine learning and text mining which have enabled researcher to gain unique insights from huge amount of data with minimal effort (Blikstein, 2011).

This paper presents part of an on-going research focused on analyzing students' behavior in a blended learning environment. Hundreds of activity logs for each student were collected, filtered and analyzed using a machine learning technique known as vector space model (VSM). The paper also describes some prototypical coding trajectories generated using these logs, look on probable relationship to student's overall course performance and finally on effects teaching and learning in blended environments.

2. Previous work

VSM has traditionally been used to search and process important information in large collections of unstructured texts (Raghavan and Wong, 1986; Mikolov *et al.*, 2013; Farid *et al.*, 2016). Recently, however, there has been some progress on utilizing VSM for purposes outside the domain of information retrieval. Sreeja and Mahalakshmi (2016), for example, explored the use of VSM to automatically detect emotions in English poems. They compared the performance of VSM with a probabilistic corpus-based method and found that VSM performs better in recognizing emotions in poems mined from public websites.

Fraser and Hirst (2016) investigated using VSM to detect language impairments among people with Alzheimer's disease and compare it with those from healthy controls. Initial findings showed changes in word usage in Alzheimer patients after analyzing their words when mapped in the VSM semantic space. Younge and Kuhn (2016) used VSM as a measure to detect patent similarity and concluded that VSM is a better measure to use for this purpose. Li and Zeng (2016) also used VSM as a foundational technology to develop a system that can be used to filter spams in mobile text messages.

The dynamic explosion of information in web-based educational system in recent years has required additional efforts in finding appropriate learning materials suitable for learners. Salehi *et al.* (2013) developed a hybrid recommender system that can overcome this problem by finding appropriate learning materials based on some specific attributes of each learner.

In the same manner, in this study, we attempt to apply VSM to data generated within web-based environments in the context of blended learning courses to enable instructors to overcome the voluminous amount of activity data generated by students as they interact with resources and with each other within the Moodle system.

Student activity logs are a key resource for gaining insight into student behavior in online environments. Behavior patterns observation, in turn, is a necessary step in detecting students' learning style.

Govaerts *et al.* (2011), for example, developed a tool called student activity meter (SAM) which can visualize the amount of time spent by students on learning activities and resources used in online learning environments. They found that visualizations generated through SAM contribute to creating awareness for teachers and that this awareness enables them to develop various teaching strategies. Ateia and Hamtini (2016), on the other hand, connected students' behavioral patterns to specific features of an online environment then used this to define the effect each visual, auditory, and kinesthetic (VAK) learning style will have on each pattern using the VAK. They claim that web-enabled learning systems "that are supported by a dynamic approach to detect the learners learning styles are better and more effective than traditional ones that extract learner learning style using traditional questionnaires."

Similarly, the work by Romero *et al.* (2013) mines the web usage data from Moodle to predict the student performance. They use features such as assignment, quizzes, and forum activity to predict students final ratings based on four categories – fail, pass, good, and excellent. The paper also presents a mining tool to extract data from Moodle. The results of the paper compare multiple algorithms and show that the fuzzy rule learning algorithms and decision trees perform well with an accuracy of 65 percent.

Agnihotri *et al.* (2015) focused on studying login data extensively and used it to cluster the students using the data they generate while interacting with a tool called Connect. The work used machine learning-based clustering techniques to group students based on their attempts, scores, and logins. Their results identified three distinct student clusters: "high-achieving students," "low-achieving students," and "persistent students" (Luik and Mikk, 2008). They also found a non-linear relationship between logins and performance based on the cluster results.

Wen and Rosé (2014) also attempted to look at the varying patterns in the behavior of students relative to their grades. Utilizing clickstream data from MOOC courses to characterize the sessions, they were able to mine student behavior within individual sessions. Results of their experiments show distinctive behaviors among students who pass, fail, and receive a distinction. This provides an indication on how different students with varying levels of course performances distribute their activities differently in online environments.

While many works in the literature describe methods to identify patterns in the student behavior, Champaign *et al.* (2014) posited that there is a strong negative correlation between student's skill and the time they spend doing online tasks. They likewise observed a negative correlation between the improvement in skill and the time on task. This finding provided added motivation for the current work in terms of verifying whether similar correlations also exist in and among students exposed in blended courses.

3. Research questions

The main goal of this research is to find effective ways to sift through the vast quantity of data generated by web-based learning environments. In particular, it aims to look into the action log data maintained by Moodle to determine whether processing models can be developed that can extract useful information that instructors can use in monitoring class activity. This involved the extraction of sample log data for selected blended learning

courses offered at Jose Rizal University (JRU). Specifically, the study addresses the following research questions:

- RQ1.* How can students' learning behavior be extracted and visualized from activity logs recorded by Moodle?
- RQ2.* Can Moodle's action log of student's online activity offer meaningful insight into students' course performance?
- RQ3.* Does the demographic profile of students have any effect on their level of activity in an online learning environment such as Moodle?

4. Data set

The data analyzed and processed in this exploratory research was extracted from various blended learning courses offered at JRU during the second semester of SY 2015-2016. "A blended learning course is defined as a formal education program in which a student learns at least in part through online learning, with some element of student control over time, place, path, and/or pace" (Blended Learning Definitions, 2017; Horn, 2013). The courses under study include Elementary Statistics (MAT22), Human Behavior Organization (MGT26), Engineering Management (EGR36), and Ethics in Information Technology (ITC56). These courses are being offered to undergraduate students taking up BSA, BSCpE, and BSIT, respectively. These blended courses were chosen because they all served as pilot implementations of the Course Redesign Program (CRP) of JRU where extensive use of the Moodle environment was introduced to enable instructors to deliver course contents – learning materials, supplemental links; promote student engagement thru online forums and chats; assessment tasks – quizzes and assignments more effectively.

As such, the online structure of each course consists of courses – readings, assignments, exercises, lecture quizzes, and a final exam, which the students are required to complete with a minimum grade of 3.5 to pass the course.

In this setup, instructors are still required to spend an one hour lecture time with the students each week, afterward, their tasks consist mainly of monitoring students' online activities while students, aside from attending class lecture will need to spend a minimum of two hours laboratory time with Moodle every week at their own discretion.

To support this setup, a pre-designed course content template is already placed in the Moodle course resources before the start of each class; however, individual teachers can customize this template by uploading additional learning materials like PowerPoint presentations, video clips, and web resources. They can also require forum participations and add additional exercises, assignments, and quizzes as they deem fit. Students can browse the contents independently through individual accounts. Some student accounts can only be used locally in the laboratories but there are also experimental accounts which are cloud-based and can be used online and accessed conveniently anywhere. Many students even access the online courses using mobile devices.

The predefined course template divides each course into several modules. For each module, students were asked to complete topic-related readings and perform the prescribed exercises, on scheduled dates, and take the online quizzes. Students can also download content materials and exercises and work on them offline.

Relative to this setup, Moodle has built-in features that can produce several types of reports that can be used to track student activity. One of these reports, called action logs, enable instructors to keep track of which resources and activities in a course have been accessed, when, and by which student. For the purposes of this study, logs of students' action for the entire semester for each of the courses were collected then cleaned up. This resulted in a data set with a total of $n = 199$ students.

4.1 Action logs

Each event record in the raw action log has six attributes (see Table I): course name, time of the event, IP address, username, action, and information. In this study, we only focused on the username and action attributes, the other attributes from the raw data were reserved for future use. The action attribute represents actions initiated by students on various items that can be accessed from Moodle such as assignment, quiz or assessment, course content, forum discussion, resource, and URLs. The actions that can be performed on these items include:

- view individual and view all – opening the items on Moodle;
- view forum – opening the forums;
- forum add discussion – add or post a forum topic;
- submit – upload completed assignments or quiz; and
- submit for grading – submit the uploaded assignments or quiz for grading.

4.2 Extracted log records

Table II provides the total number of action log records extracted for each course as well as the average number of actions per student. These logs constitute actions initiated for Moodle tools identified previously.

4.3 Student demographics

Finally, student demographics were taken by means of a structured survey using purposive sampling, participants in the survey conducted were the same students whose activity logs were extracted and processed. These data as shown in Table III are essential in identifying possible focal determinants of students’ online behavior exhibited by Moodle recorded action logs.

These attributes would be checked later to determine whether students online activity is affected by gender, year level, enrollment status or the number of CRP (online) course they are taking for second semester.

The two latter attributes (device ownership and access mode) would help describe how students took advantage of the mobility factor of an online course relative to the blended

Table I.
Moodle action log
data attributes

Data dimension	Description
Course	Identification string of the course in which the action is related
Time	Date and time stamp of when the action was executed
IP address	Unique numerical label assigned to the device used by the user
User full name	The user who initiated the action
Action	Type of action initiated
Information	General information on learning activities

Table II.
Total number of
actions per course

Course	Number of students	Total number of action logs	Average actions per student	SD
MAT22	60	10,370	173	43.76
EGR36	80	12,009	150	17.04
MGT26	22	2,687	128	20.32
ITC56	37	3,520	95	16.61
Total	199	28,586		

Demographic characteristics	Course Respondents (N)	MAT22 n = 60		MGT26 n = 21		EGR36 n = 80		ITC56 n = 37	
		F	%	F	%	F	%	F	%
Gender	Female	32	53.33	12	57.14	24	30.00	13	35.14
	Male	28	46.67	9	42.86	56	70.00	24	64.86
Year level	1st Year	50	83.33	2	9.52	0	0.00	1	2.70
	2nd Year	5	8.33	17	80.95	0	0.00	31	83.78
	3rd Year	3	5.00	2	9.52	77	96.25	4	10.81
	4th Year	2	3.33	0	0.00	3	3.75	1	2.70
Enrollment status	Regular	51	85.00	11	52.38	65	81.25	20	54.05
	Irregular	9	15.00	10	47.62	15	18.75	17	45.95
No. CRP course taken	1 CRP course	17	28.33	15	71.43	32	40.00	7	18.92
	2 CRP course	18	30.00	3	14.29	34	42.50	18	48.65
	3 CRP course	25	41.67	1	4.76	11	13.75	9	24.32
	> 3 CRP course	0	0.00	0	0.00	3	3.75	3	8.11
Device ownership	PC with Internet	35	58.33	13	61.90	50	62.50	29	78.38
	No PC	20	33.33	2	9.52	18	22.50	5	13.51
	PC without Internet	5	8.33	6	28.57	2	2.50	3	8.11
Access mode	JRU Open Lab	20	33.33	4	19.05	20	25.00	7	18.92
	Use own device home/mobile	28	46.67	16	76.19	16	20.00	12	32.43
	JRU Lab Session	0	0.00	0	0.00	9	11.25	4	10.81
	Use of mobile device	3	5.00	0	0.00	20	25.00	3	8.11
	Combinational access mode	9	15.00	1	4.76	15	18.75	11	29.73

Table III.
Student demographic characteristics

mode design (Section 4.3). Device ownership, in particular, defines how many students have their own desktop or mobile provisions, while access mode tells us if students are using these for LMS access relative to specific types of Moodle activity, e.g. content-, assessment- or engagement-related tasks.

As shown, the majority of students own a computer with internet access; 58.33, 61.90, 62.50, and 78.38 percent for MAT22, MGT26, EGR36, and ITC56, respectively, and they are those who access Moodle using their own provisions as reflected by 46.67, 76.19, 20.00, and 32.43 percent.

Respondents with no PC or have a computer but no internet access relied on using JRU Open Lab (33.33, 19.05, 25.00, and 18.92 percent). This implies a positive observation because it implies that students who lack personal devices can still access course contents and perform online tasks through the university infrastructure as provided in the open lab.

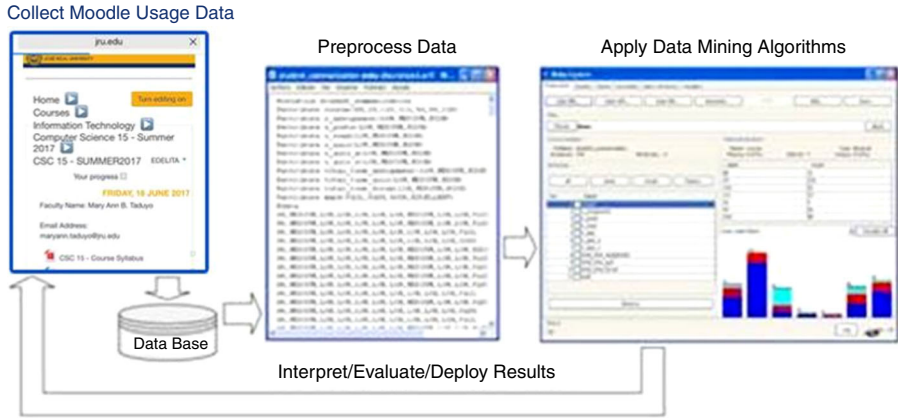
Use of mobile device was differentiated from use mobile device home/mobile by verifying IP address stamped in each action log (the IP address is a set of numeric values that specifically identifies the device being used by the student). Another notable item in the survey shows that MAT22 students have taken more CRP (Moodle online) courses during the semester compared to the other three courses.

5. Analysis of activity data

5.1 Extracting and visualizing learning behavior

After data collection, the first question that was addressed is how to process the data set and extract patterns of activity that can be used to visualize students learning behavior. Following the mining process described by Romero *et al.* (2008), as shown in Figure 1, a two-phase process was used which include initial preprocessing of data then afterwards applying data mining algorithms that transform the data into a form suitable for interpretation and evaluation. In the context of this study, Moodle log data were collected from JRU LMS for a particular CRP course as depicted in Figure 1.

Figure 1.
Data processing model



5.1.1 Data preprocessing. Data pre-processing is one crucial step in data mining (Mohamed, 2014). In this phase, the raw log files were first processed to clean and prepare it for further processing. This is critical because many of the data sets extracted in Moodle can have missing values, noisy data, and/or irrelevant and redundant information. For this purpose, the raw log files were first imported into an Excel worksheet. Here, the actions logged by instructors and course administrators were selectively removed and the data set was anonymized by removing each student’s name and replacing it with a unique identification number. Processing then started by filtering the data set by course, user identification, and action. Then, two-dimensional tables for each course were built containing the list of student identifiers as row headers and specific types of actions as column headers. Table IV presents the set of action types used in this study for analyzing the students’ online behavior. The key aspect of these actions is that collectively, they can be used to represent the different types of activities that students can engage with inside Moodle, that is: accessing course content, engaging with peers, and taking assessment tests. The key assumption here is that student’s actions indicate intentionality which in turn provide clues, as to their learning preferences. Thus, when categorized based on class activities, the actions helps to infer whether the student prefers to study by accessing learning materials, by engaging with peers and/or the instructor or simply by taking assessment tests.

Each cell in the two-dimensional table was filled with values representing the total number of times each action type was initiated by each student. Figure 2 shows a sample table generated after pre-processing the raw data files. The process of counting this value

Class activity	Action type
Content access	Course view
	Resource view
	URL view
Engagement	Forum add disc
	Forum view disc
	Forum view forum
Assessment	Quiz view
	Quiz attempt
	Assign view
	Assign submit

Table IV.
Moodle action
identifiers

Students	Course view	Resource view	URL view	Forum add disc	Forum view disc	Forum view forum	Quiz view	Quiz attempt	Assign view	Assign submit	Final Grade
1	1001	116	29	36	0	0	3	47	0	0	1.8
2	1002	79	44	7	2	12	10	36	7	1	1.9
3	1003	46	7	0	0	0	0	36	7	0	2.2
4	1004	50	34	8	3	8	13	35	7	0	2.3
5	1005	32	15	0	0	0	0	24	6	0	2.3
6	1006	20	2	6	0	6	2	21	7	0	2.4
7	1007	51	45	28	3	15	23	25	7	0	2.4
8	1008	39	23	15	1	2	7	19	8	0	2.4
9	1009	93	56	30	0	0	0	28	7	0	2.5
10	1010	69	18	4	0	0	0	40	7	0	2.5
11	1011	98	27	8	0	2	3	37	8	0	2.5
12	1012	109	15	21	1	17	13	33	8	0	2.6
13	1013	63	17	0	1	5	8	25	7	0	2.7
14	1014	48	12	20	0	0	5	20	6	0	2.8
15	1015	41	6	0	0	0	1	39	8	0	2.9
16	1016	30	1	0	0	0	2	30	7	0	2.9
17	1017	35	8	5	1	4	0	24	7	0	2.9
18	1018	63	35	2	0	0	0	26	7	0	3.2
19	1019	41	25	8	1	1	6	21	7	0	3.4
20	1020	27	1	2	0	0	1	23	6	0	3.4
21		0	0	0	0	0	0	0	0	0	5
	Total	1,150	420	200	13	66	97	589	142	0	2,7143

Figure 2.
Preprocessed raw data
(activity logs)

was automatically done using a customized Excel macro. The total counts extracted for each action type are shown in Table V wherein course views, view forum discussions, view forum, and assignment views have the highest occurrence while quiz view, quiz attempt, add forum discussion, and URL view are relatively low and assignment submit, resource view, and assignment view actions are the fewest actions initiated by the students.

5.1.2 Data mining algorithm – VSM. Data mining algorithms enable extraction and visualization of patterns of activity that can be used to infer students' behavior.

VSM, a statistical model representation often used in processing documents in information retrieval (Raghavan and Wong, 1986). The main idea behind VSM is to construct vector representation for documents and use these vectors to analyze and compare the contents of each document. A vector is simply a labeled set of values arranged in a specific order. In the case of VSM, the labels are the unique words that occur in the document and the values refer to the number of times each unique word occurred in that document. So for example, if there is k number of documents to be represented and these documents contain n number of unique words, a $k \times n$ matrix can be built as shown in Figure 3. In this matrix, D_1 to D_k represent the set of documents while W_1 to W_n represent the set of unique words. The values in each cell represent the number of times a specific word W occurred in a particular document D . Each row in this matrix is considered a vector representation for its corresponding document.

The analogy used in VSM is that the vector representation acts as a sort of coordinate that can be used to plot the position of the document in an n -dimensional semantic space where n corresponds to the number of values in the vector. Figure 4 depicts what a three-dimensional semantic space looks like along with the documents plotted in this space using vector representation.

Using this analogy, to compare the contents of documents VSM simply determines how far the location of their vector representations is within the semantic space. For instance, to determine how closely related the topic of document D_1 is to the topic of D_2 , VSM simply

Action	Count
Course view	11,265
View forum disc	6,491
View forum	3,047
Quiz view	2,041
Quiz attempt	1,870
Add forum disc	1,817
URL view	1,193
Assign submit	444
Resource view	347
Assign view	71

Table V.
Statistics on logged
student actions

	w_1	w_2	w_3	\dots	w_n
D_1	14	6	1	\dots	0
D_2	0	1	3	\dots	1
D_3	0	1	0	\dots	2
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
D_K	4	7	0	\dots	5

Figure 3.
 $k \times n$ matrix

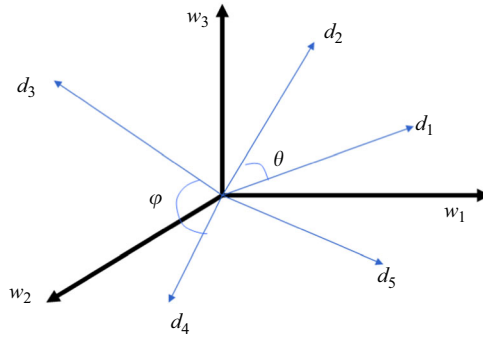


Figure 4.
3D semantic space

measures the distance of D_2 relative to the position of D_1 . The most common method for measuring this distance is by calculating the cosine of the angle formed between the two locations (represented by the symbol Θ). The formula for computing the cosine angle is as follows:

$$s(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} = \frac{\sum_{i=0}^{n-1} x_i y_i}{\sqrt{\sum_{i=0}^{n-1} (x_i)^2} \times \sqrt{\sum_{i=0}^{n-1} (y_i)^2}} \quad (1)$$

Basically, for two vectors with n values, this formula simply computes the scalar product of the two vectors for the numerator; computes the product of the length or norm of the two vectors for the denominator. So for example, if vector x is represented by the values (1, 1, 1, 3, 0) and vector y is represented by the values (0, 0, 0, 1, 1) the computation for the resulting cosine angle is as follows:

$$\begin{aligned} x \cdot y &= 1 \times 0 + 1 \times 0 + 1 \times 0 + 3 \times 1 + 0 \times 1 = 3 \\ \|\mathbf{x}\| &= \sqrt{(1^2 + 1^2 + 1^2 + 3^2 + 0^2)} = \sqrt{12} = 3.4641 \\ \|\mathbf{y}\| &= \sqrt{(0^2 + 0^2 + 0^2 + 1^2 + 1^2)} = \sqrt{2} = 1.4142 \\ \text{cosine} &= 3 / (1.4142 \times 3.4641) = 0.61 \end{aligned}$$

The cosine formula returns a value between 0 and 1. The rule in VSM is that the more similar the contents of two documents are, the higher their cosine value will be. So a cosine value of 1 for two documents means that the documents are completely identical and a value of 0 means they are totally unrelated. Any value in between reflects the degree of comparison between documents, the higher value means documents are highly related.

5.1.3 *Representing student activity using VSM.* Given the previous discussion, representing student activity using VSM requires the construction of activity vectors for each student. An activity vector can be defined as simply a list of action types with their corresponding values depicting how many times each action was initiated by the student. Here, a value of 0 means that the action type was not initiated at all. For instance, in Figure 3, the level of activity of student 1001 can be represented by the vector:

116	29	36	0	0	3	47	8	0	0
-----	----	----	---	---	---	----	---	---	---

There are two ways by which this vector representation can be used. First, it can be used to compare students' activity to each other in order to group them based on how similar their level of activity is. Second, it can be used to assign students to a predefined set of categories based on how close their activity level *vis-à-vis* defined activity level for a specific category. Both cases will enable the identification of similar characteristics that occur within each group of students. In this paper, the latter approach is explored.

The color-coded header in Figure 3 indicates the type of class activity to which the action type belongs, such as content access, engagement related, and assessment activity. These sets of activities can be used to classify students to determine which type of activity they implicitly prefer. To do this, an archetypal activity vector for each activity class needs to be constructed. This can be done by setting the corresponding action types for each activity to a non-zero value while the rest of the action type values are set to 0 as shown in Table VI. Thus, the archetypal vector for each activity class would be as follows.

To classify, each student's activity vector will be compared to the archetypal vector of each activity. The student will be grouped accordingly as per archetypal vector which generated the highest cosine or similarity value. Student activity vector can also be analyzed and grouped and compared on a per course/class basis.

5.2 Analysis of correlation between activity level and course grades

Prior to delving further in analyzing activity logs, it is necessary to first determine whether there is a relationship that exists between the action types initiated by the students and the students' course achievements, what the direction of the relationship is and its strength of magnitude. For this purpose, the student's final course grade is treated as an indicator of course achievement; which can reflect both student knowledge and level of engagement. The goal is to gain insight into how students' actions in the online environment correlate with their course grades. Pearson coefficient correlation (r) was used to investigate the significance and computations were done by importing the excel data worksheet to SPSS.

5.3 Analysis of the effect of students demographic profile with level of activity

Correlation and descriptive statistics were conducted to examine whether student demographic attributes, namely, gender, year level, enrollment status, and device ownership could affect the level of LMS utilization as exhibited by total activity logs. Descriptive statistics used to determine mean activity logs of students while Pearson coefficient (r) was also used to establish possible relationships.

6. Results and discussions

The development of easily interpretable graphic that can depict trends in student activity based on action logs is a useful tool for instructors to constantly visualize and monitor course progress with minimal effort. Each line point in the graph (Figure 5) shows representative visualizations of the cosine values generated by each student in their respective course. Although students are anonymously depicted, the graph depicts the degree of activity among the participants, and can possibly be even be refined to drill down to each individual student's level of activity. The visualizations clearly depict some patterns of online behavior relative to three different activities: content access, engagement, and assessment. It indicates that different classes vary widely in how they utilize the tools

Table VI.
Vector model for
activity class

Content	1	1	1	0	0	0	0	0	0	0
Engagement	0	0	0	1	1	1	0	0	0	0
Assessment	0	0	0	0	0	0	1	1	1	1

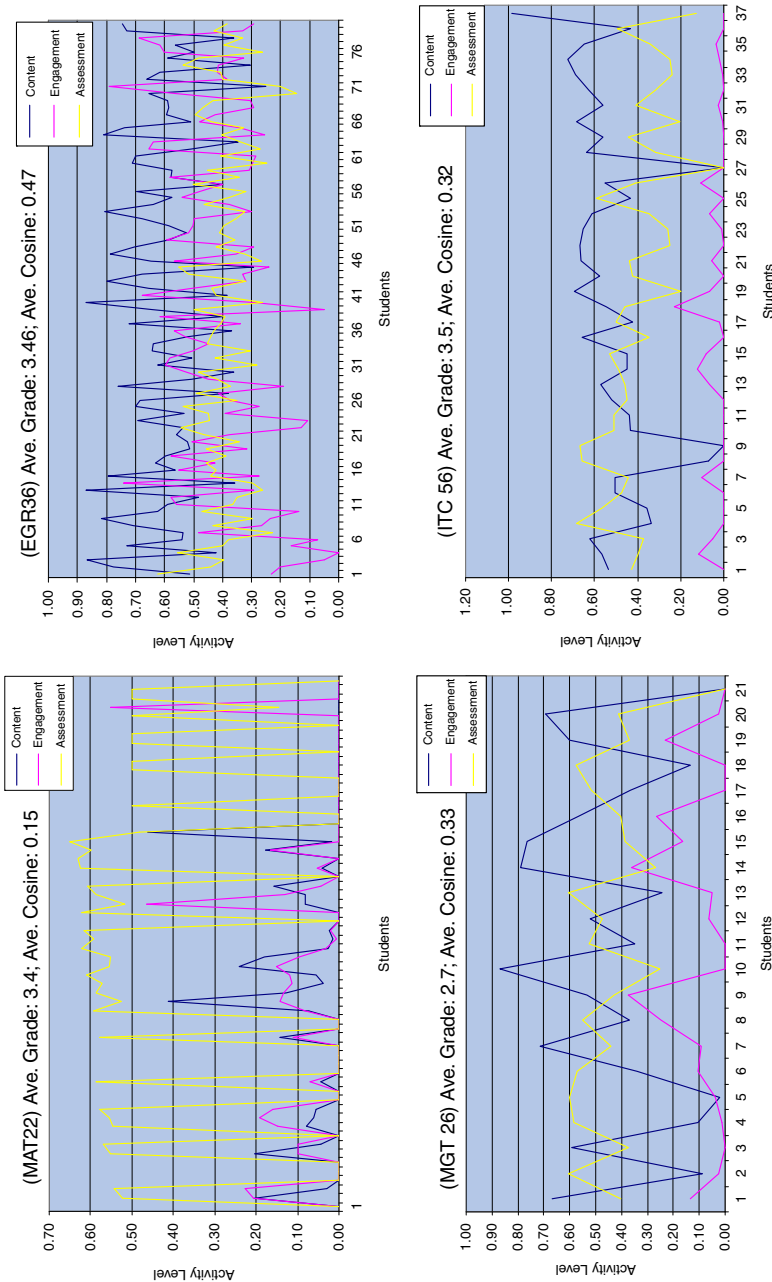


Figure 5.
Visualization of class
activity based on
action logs analyzed
thru VSM

provided within Moodle. And that, even within a certain class, students undertake complex behaviors in allotting time between different tools and activities. Students from MAT22, for example, generally login into Moodle mainly to do assessment tasks with little access to educational resources and engagement. Whereas EGR36 students seem to prioritize content access and engagement over access to assessment tasks; students from MGT26 and ITC56, on the other hand, while more or less showing equal interest on content access and assessment show very little interest in engagement. These visualizations can help course administrators to determine the type of strategic interventions that each course would need to ensure that student's activities are kept within intended learning outcomes. Unfortunately, while some of these online activities aim to provide effective teaching strategies, the visualizations (along with the cosine ratings) does not seem to correlate with the students' course accomplishments. This suggestion comes from observing that some course with a low level of online activity (MGT26 e.g. average cosine: 0.33) have higher average grades (e.g. 2.7) than a course with higher levels of activity (e.g. EGR36, average grade: 3.46, average cosine 0.47). This seems to suggest that longer time spent on Moodle may not result in higher course achievement. An implication of these observations is a need to redesign the online component more effectively in order to achieve quality instruction.

Another issue that the visualizations reveal is the lack of a standard teaching approach. Since students' activities are often governed in part by the teacher requirements, what the visualizations indicate is that teaching methods among different classes seem to vary. Some instructors mainly focus on uploading lectures, while others focus more on assessment tasks or activities; some require a certain level of engagement among their students. More studies are needed to determine which pattern of teaching approach would be most beneficial to the students.

A simple correlation analysis was conducted, in order to determine whether there are relationships between the action types initiated by students and their course accomplishment represented by the final grade given to them by their instructors. As shown in Table VII, the results seem to suggest that there is some variability in terms of positive and significant correlation in the final grade between courses. In Mat22, for instance, the only action which shows correlation with the final grade is the URLView ($r = 0.266$, $p = 0.01$). Whereas, in the EGR36 course, all the actions correlated positively except AssignmentSubmit and AssignmentView ($p = 0.01$). For MGT26, it is the QuizView and QuizAttempt actions that correlated positively ($p = 0.01$) and for ITC56 it is ResourceView and QuizView shows medium to high correlation at $p = 0.01$ and small correlation for AssignView at $p = 0.05$. This variability in terms of correlation, to some degree, seems to agree with our previous observation regarding the lack of correlation between the students' online activity level and final grades. In other words, in some cases, it may correlate but in others, it may not. The magnitude of correlation also varies from activity type which reflects how students have prioritized tool or activity performed in the online environment. The best explanation for this observation is that instructors are considering other factors in assigning grades to the students which may not be present in the online environment.

Course	Res. view	URL view	ForAdd disc	ForView disc	ForView forum	Quiz view	Quiz attempt	Assign submit	Assign view
MAT22 ($n = 60$)	0.245	0.266*	0.183	0.027	0.117	0.145	0.161		
MGT26 ($n = 21$)	0.431	0.390	0.258	0.272	0.282	0.742**	0.788**		
EGR36 ($n = 80$)	0.344**	0.288**	0.520**	0.334**	0.485**	0.415**	0.694**	0.202	0.194
ITC56 ($n = 37$)	0.604**	0.207	0.045	0.235	0.002	0.470**	0.502		0.373*

Notes: * $p < 0.05$; ** $p < 0.01$

Table VII.
Pearson coefficient
of correlation (r)
of final grade vs
Moodle activity

The analysis of correlation coefficients for student demographics and total action logs (TAL) obtained as depicted in Table VIII can be observed per course wherein MAT22 has relatively negative, positive small correlation on gender, enrollment status, and device ownership while statistically significant relationship between TAL and CRP courses taken (0.356, $p < 0.01$, 58). This was contrasted by results for MGT26, with negative/positive medium to high correlation was significantly established for gender and enrollment status (-0.564 , $p < 0.01$, 0.506 at $p < 0.05$, 19) while EGR36 got a negatively small coefficient for gender (-0.305^{**} , $p < 0.01$, 78). ITC56 results did not show any significant relationship among any factors considered. Closer analysis of the IP address stamped on the log hits indicated that students still displayed a high level of reliance on computer units provided in the open laboratories. In effect, this nullifies the mobile access advantages of the online course. It was also observed that majority of the task performed online are assessment related. This inference could be attributed to the environment of blended learning implementation of the university where students despite mobility and availability of learning resources online still rely on classroom discussions performed during the face-face session which is one primary structure of a blended mode. Likewise, prior experience in LMS tool indicated by the number of CRP course taken is not a factor on students' online LMS activity level nor students' achievement.

7. Limitations and future work

The paper showed how students of various courses utilized the LMS system, it may not be indicative of the overall effectiveness of the system but a structured analytical study of actual online activity thru a data-driven approach gave important highlights. In summary, this work has shown that VSM can be used to aggregate the action logs of students and quantify it into a single numeric value that can be used to generate visualizations of students' level of activity. While the visualizations do not seem to depict course performance, the value it presents is in terms of helping instructors monitor course progression and enable them to rapidly identify which students are not performing well and adjust their pedagogical strategies accordingly. Since the VSM visualizations generated in this study, for the most part, relied on the structure and contents of the activity reports generated by Moodle, adopting the same methodology to other LMS will necessarily involve minimal adjustments, specifically, in the construction and generation of the activity vectors. Nonetheless, once the necessary adjustments have been made, VSM can be applied to other courses even with varying designs and structure.

A plan to continue the work by developing a complete dashboard style interface that instructors can use is already underway. The study also looked into whether or not various action types can be used as indicators of student's class performance. The current investigation indicates that there is a lot of variability in terms of the correlation between these two variables. It is highly likely that the design and nature of the course as well as the individual teaching strategy of instructors are introducing other factors that are not present

Course	Mean TAL	SD	Gender	Year level	Enrollment status	No. of CRP course taken	Device ownership profile
MAT22 ($n = 60$)	98.72	112.02	-0.032	0.125	-0.230	0.356**	0.085
MGT26 ($n = 21$)	73.19	35.81	-0.564**	-0.219	0.506*	-0.138	0.132
EGR36 ($n = 80$)	150.10	20.52	-0.305**	-0.173	0.167	0.015	0.191
ITC56 ($n = 37$)	54.86	29.28	-0.302	-0.176	-0.049	-0.224	-0.125

Notes: * $p < 0.05$; ** $p < 0.01$

Table VIII.
Correlation analysis of
total activity log
(TAL) and student
demographics

in the online environment. A comparative inquiry on a per subject basis can also be done to explore the effectiveness of a particular course module taken by students in the different disciplinary area. More data need to be collected, and more advanced processing tools are needed in order to obtain a better perspective on this issue.

References

- Agnihotri, L., Aghababayan, A., Mojarad, S., Riedesel, M. and Essa, A. (2015), "Mining login data for actionable student insight", *International Educational Data Mining Society*, pp. 472-474, available at: www.educationaldatamining.org/EDM2015/proceedings/edm2015_proceedings.pdf (accessed April 13, 2016).
- Alonso, F., López, G., Manrique, D. and Viñes, J.M. (2005), "An instructional model for web-based e-learning education with a blended learning process approach", *British Journal of Educational Technology*, Vol. 36 No. 2, pp. 217-235, available at: www.fisme.science.uu.nl/publicaties/literatuur/2005_modelforwebbasedelearning.pdf (accessed January 2, 2017).
- Ateia, H. and Hamtini, T. (2016), "Designing and implementing of dynamic technique for detecting learning style using literature-based approach", *International Journal of Database Theory and Application*, Vol. 9 No. 6, pp. 9-20, available at: www.sersc.org/journals/IJDTA/vol9_no6/2.pdf (accessed January 2, 2017).
- Blended Learning Definitions (2017), "Clayton Christensen Institute", available at: www.christenseninstitute.org/blended-learning-definitions-and-models/ (accessed March 15, 2017).
- Blikstein, P. (2011), "Using learning analytics to assess students' behaviour in open-ended programming tasks", *Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, Alberta*, pp. 110-116, available at: <http://dl.acm.org/citation.cfm?id=2090132> (accessed April 12, 2017).
- Borromeo, R.M.H. (2013), "Online exam for distance educators using moodle", *2013 IEEE 63rd Annual Conference International Council for Educational Media, Institute of Electrical and Electronics Engineering IEEE, Singapore*, pp. 1-4.
- Champaign, J., Colvin, K.F., Liu, A., Fredericks, C., Seaton, D. and Pritchard, D.E. (2014), "Correlating skill and improvement in 2 MOOCs with a student's time on tasks", *Proceedings of the First ACM Conference on Learning@ Scale Conference, ACM, Georgia*, pp. 11-20, available at: <https://edocs.uis.edu/jcham4/www/p11-champaign.pdf> (accessed June 4, 2016).
- Farid, M., Métais, E., Saraee, M., Sugumaran, V. and Vadera, S. (2016), "Natural language processing and information systems", *21st International Conference on Applications of Natural Language to Information Systems, Salford, June 22-24*, available at: <https://link.springer.com/book/10.1007/978-3-319-41754-7> (accessed April 12, 2017).
- Fraser, K.C. and Hirst, G. (2016), "Detecting semantic changes in Alzheimer's disease with vector space models", *Proceedings of LREC 2016 Workshop, Resources and Processing of Linguistic and Extra-Linguistic Data from People with Various Forms of Cognitive/Psychiatric Impairments, Linköping University Electronic Press, Portorož, No. 128*, pp. 1-8, available at: www.ep.liu.se/ecp/128/ecp16128.pdf (accessed March 15, 2017).
- Govaerts, S., Verbert, K. and Duval, E. (2011), "Evaluating the student activity meter: two case studies", *International Conference on Web-Based Learning, Springer, Berlin and Heidelberg*, pp. 188-197, doi: 10.1007/978-3-642-25813-8_20.
- Horn, M. (2013), "Disrupting the classroom", *The Freeman*, Vol. 63 No. 2, pp. 6-7, available at: https://admin.fee.org/files/doclib/20130318_march2013freemanweb.pdf (accessed April 12, 2017).
- Hughes, G. (2007), "Using blended learning to increase learner support and improve retention", *Teaching in Higher Education*, Vol. 12 No. 3, pp. 349-363, available at: <http://eprints.ioe.ac.uk/2022/> (accessed August 8, 2016).
- Li, W. and Zeng, S. (2016), "A vector space model-based spam SMS filter", *2016 11th International Conference on Computer Science and Education IEEE*, pp. 553-557, available at: <http://ieeexplore.ieee.org/document/7581640/> (accessed March 15, 2017).

- López, G.A., Sáenz, J., Leonardo, A. and Gurtubay, I.G. (2016), "Use of the Moodle platform to promote an ongoing learning when lecturing general physics in the physics, mathematics and electronic engineering programmes at the University of the Basque Country UPV/EHU", *Journal of Science Education and Technology*, Vol. 25 No. 4, pp. 575-589, available at: <http://link.springer.com/article/10.1007/s10956-016-9614-8> (accessed August 8, 2016).
- Luik, P. and Mikk, J. (2008), "What is important in electronic textbooks for students of different achievement levels?", *Computers & Education*, Vol. 50 No. 4, pp. 1483-1494, available at: www.sciencedirect.com/science/article/pii/S0360131507000164 (accessed April 12, 2017).
- McLuckie, J.A., Naulty, M., Luchoomun, D. and Wahl, H. (2009), "Scottish and Austrian perspectives on delivering a master's: from paper to virtual and from individual to collaborative", *Industry and Higher Education*, Vol. 23 No. 4, pp. 311-318.
- Macfadyen, L.P. and Dawson, S. (2010), "Mining LMS data to develop an 'early warning system' for educators: a proof of concept", *Computers & Education*, Vol. 54 No. 2, pp. 588-599, available at: <https://pdfs.semanticscholar.org/66d9/efad07ea90dd1796f35ed439016ab0f5a389.pdf> (accessed August 8, 2016).
- Maila, R., Custer, C.D., Celso, B. and Shearyl, U.A. (2014), "Make-It-ECE', a mathematics learning management system (LMS) for engineering students in the Philippines", *International Journal of Education and Research*, Vol. 2 No. 9, pp. 109-118, available at: www.ijern.com/journal/2014/September-2014/10.pdf (accessed May 7, 2016).
- Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013), "Efficient estimation of word representations in vector space", available at: <https://arxiv.org/abs/1301.3781> (accessed March 15, 2017).
- Mohamed, E. (2014), "Predicting causes of traffic road accidents using multiclass support vector machines", *International Conference in Data Mining, Las Vegas, NV, July 21-24*, available at: http://worldcomp-proceedings.com/proc/proc2014/dmin/DMIN_Papers.pdf (accessed April 13, 2017).
- Perkins, M. and Pfaffman, J. (2006), "Using a course management system to improve classroom communication", *Science Teacher*, Vol. 73 No. 7, pp. 33-37, available at: <http://search.proquest.com/docview/62029987?accountid=27575> (accessed August 8, 2016).
- Prasad, R.K. (2015), "Hybrid, mixed-mode, or blended learning: better results with elearning", *Learning Solutions Magazine*, available at: www.moodlerooms.com (accessed April 10, 2016).
- Raghavan, V.V. and Wong, S.M. (1986), "A critical analysis of vector space model for information retrieval", *Journal of the American Society for Information Science*, Vol. 37 No. 5, pp. 279-288, available at: [http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-4571\(198609\)37:5%3C279::AID-ASH1%3E3.0.CO;2-Q/full](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(198609)37:5%3C279::AID-ASH1%3E3.0.CO;2-Q/full) (accessed August 8, 2016).
- Regueras, L.M., Verdu, E., Verdu, M.J. and de Castro, J.P. (2011), "Design of a competitive and collaborative learning strategy in a communication networks course", *IEEE Transactions on Education*, Vol. 54 No. 2, pp. 302-307.
- Roby, T., Ashe, S., Singh, N. and Clark, C. (2013), "Shaping the online experience: how administrators can influence student and instructor perceptions through policy and practice", *The Internet and Higher Education*, Vol. 17, pp. 29-37, available at: <http://faculty.washington.edu/rvanderp/DLData/RobyAshe2012.pdf> (accessed April 10, 2016).
- Romero, C., Ventura, S. and Garcia, E. (2008), "Data mining in course management systems: Moodle case study and tutorial", *Computers & Education*, Vol. 51 No. 1, pp. 368-384, available at: [http://research.moodle.net/95/1/Romero%20\(2008\)%20Data%20mining%20in%20course%20management%20systems-%20Moodle%20case%20study%20and%20tutorial.pdf](http://research.moodle.net/95/1/Romero%20(2008)%20Data%20mining%20in%20course%20management%20systems-%20Moodle%20case%20study%20and%20tutorial.pdf) (accessed April 12, 2016).
- Romero, C., Espejo, P.G., Zafra, A., Romero, J.R. and Ventura, S. (2013), "Web usage mining for predicting final marks of students that use Moodle courses", *Computer Applications in Engineering Education*, Vol. 21 No. 5, pp. 135-146.
- Salehi, M., Pourzaferani, M. and Razavi, S. (2013), "Hybrid attribute-based recommender system for learning material using genetic algorithm and a multidimensional information model", *Egyptian Informatics Journal*, Vol. 14 No. 1, pp. 67-78.

- Sreeja, P.S. and Mahalakshmi, G.S. (2016), "Comparison of probabilistic corpus-based method and vector space model for emotion recognition from Poems", *Asian Journal of Information Technology*, Vol. 15 No. 5, pp. 908-915.
- Wen, M. and Rosé, C.P. (2014), "Identifying latent study habits by mining learner behavior patterns in massive open online courses", *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, ACM*, pp. 1983-1986.
- Woltering, V., Herrler, A., Spitzer, K. and Spreckelsen, C. (2009), "Blended learning positively affects students' satisfaction and the role of the tutor in the problem-based learning process: results of a mixed-method evaluation", *Advances in Health Sciences Education*, Vol. 14 No. 5, pp. 725-738, available at: <http://link.springer.com/article/10.1007%2Fs10459-009-9154-6> (accessed April 12, 2016).
- Younge, K. and Kuhn, J. (2016), "Patent-to-patent similarity: a vector space model", *SSRN Electronic Journal*, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2709238 (accessed December 30, 2015).

Corresponding author

Rosalina Rebuscas Estacio can be contacted at: rosalina.estacio@jru.edu