Visualizing Student Activity in Blended Learning Classroom by Mining Course Log Data

Rodolfo C. Raga Jr., Jennifer D. Raga, and Israel V. Cariño
College of Computer Studies and Engineering, Jose Rizal University, Mandaluyong City, Phillippines
Email: {rodolfo.raga, jenny.raga, israel.carino}@jru.edu

***Abstract*—Many higher educational institutions (HEIs) in the Philippines have started to implement web-based learning environments capable of delivering online education in an academic setting. These environments are often based on the functions and features of Learning Management Systems (LMS) such as Moodle. Due to the nature of the design of these environments which focuses on allowing students access to educational resources at their own discretion, HEIs are now also capable of routinely collecting vast quantities of data on student's online activity. Taking advantage of this data, however, is neither simple nor straightforward due to its massive volume and high rate of velocity. Based on experience, this keeps instructors from making meaningful sense and use of this data. This paper describes a proposal for a novel data driven approach in analyzing student action logs recorded by Moodle in order to generate graphical representations that can be used by instructors to monitor students' activities at any stage of course progression. The analysis was carried out using log data obtained from several blended courses dispensed in one University using a Moodle platform. The initial findings indicate a strong indication that the approach can be used to reveal variations in behavioral aspects of students in terms of patterns in resource access, assessment tasks, and degree of engagement. The concluding section of this paper presents some implications for further improvements and for future research plans.**

***Index Terms*—Blended learning, learning management system, moodle**

I. Introduction

Blended Learning (BL) has become popular in the last few years, spurred by the widespread use of the web and the opportunities and conveniences that it provides [1]. By combining face-to-face and online learning, it is virtually considered as being capable of accommodating students with distinct learning needs. Because of many benefits to learners, BL technologies have been adopted by many Higher Education Institutions (HEIs) in the Philippines in recent years. Wikipedia defines BL as "a formal education program in which a student learns at least in part through delivery of content and instruction via digital and online media with some element of student control over time, place, path, or pace". As one study indicates [2], the benefits of BL include the ability to learn outside the classroom and the potential for students to take control of learning. By allowing students access to learning resources "anytime, anywhere", BL creates an advantageous environment that removes many barriers to enhancing student performance, whilst enabling high quality interactions between faculty and students [3].

Learning Management Systems (LMS) is a key feature of Blended Learning. Specifically, in the Philippines, LMS is often used to distribute learning contents and assignment to students, deliver tests, promote engagement through online discussions, and, as a whole, to enable faculty to manage their blended learning classes [4]. In particular, at Jose Rizal University, the Moodle (Modular Object Oriented Developmental Learning Environment) LMS was adopted. Moodle is an open-source learning course management system deployed for the purpose of enabling educators to create online learning communities. It serves as an effective alternative to proprietary commercial online learning solutions because it is distributed free under open source licensing. It has been installed at universities and institutions all over the world and has established a large and diverse user community.

In a blended education context, due to the nature of its purpose, the Moodle system can accumulate vast amount of data related to students; such as when students are accessing learning contents, submitting their assignments, taking tests, and even communicating with each other [5]. Previous studies conducted in developed countries provide evidence that these data provide a gold mine of valuable information that can be used for analyzing students' learning behavior [5]. However, due to its vast quantities, these data can be very difficult to manage manually. Most often than not, assistance from specialized tools is needed to extract useful information for tracking and assessing the activities performed by students, especially in cases where faculty is handling multiple classes. At the same time, although the Moodle system provides some reporting tools, it does not provide specific features which can enable educators to evaluate the structure and contents of the course and its effectiveness in the learning process by monitoring and keeping track of students' activities. This paper describes a pilot study that presents that use of an approach called Vector Space Model (VSM), adopted from the field of Information Retrieval, in order to analyze and mine the log data generated by Moodle in a blended learning context. The proposed techniques can be used to

Manuscript received May 2, 2017; revised August 1, 2017.

doi: 10.18178/ijlt.4.1.1-6

breakdown multidimensional tracking data collected by the LMS in order to generate various graphical representations that provide a profile of students' activities, both individually and within a group.

The rest of the paper is organized as follows. Section 2 provides a brief discussion on the rationale of the study. Section 3 describes the course log report that Moodle generates and discusses the dataset used in this paper. Section 4 presents the proposed data processing approach. Experiments and results are presented in Section 5 followed by a discussion on the findings of this study and future work in Section 6 and references in Section 7.

II. Research Rationale

In general, a Learning Management System is a software infrastructure that allows relatively easy creation of online course content and coursework and the subsequent delivery and sharing of these resources among students in an online learning environment, such as that of blended learning [6]. Moodle, for example, includes a set of tools that allows faculty to monitor student participation as well as assess students' performance online. These functionalities enable both faculty and students to continue interacting even outside the classroom. Beyond delivering content, the LMS can also be used to handle course registration and course administration and also allows instructors to conduct skills gap analysis to some extent.

However, how LMS promotes learning activities among students and how this affects the students' performance remains unclear. Debates with regards the pedagogic value of LMS is still on-going; because, while there is evidence to suggest its high potential, there is also considerable evidence that attempts at utilizing this technology fail to fulfill this potential [7], [8]. Some studies, for example, indicate that the use of LMS have had a limited impact on pedagogy [9], [6]. Others argue that personal tools are more suitable in supporting students' independent activity. This argument suggests an approach that moves the focus away from the integrated facilities of LMS and into providing several separate tools that can provide different learning opportunities to students [10], [11]. Still, others simply claim that the use of technology has no beneficial effect on learning and is even instrumental in maintaining students in a state of semi-disengagement [12]. The same study reported concern by teachers that technology could decrease student interaction and result in greater social isolation for the student. Moreover, Ref. [13] suggested that, from an academic perspective, specifically in coursework, the use of internet-based LMS is not necessarily correlated with student's satisfaction. Reference [14], on the other hand, found that being convinced of the effectiveness of technology was necessary before teachers would fully engage with it.

Obviously, there is a need to further examine the way tools within LMS promotes activities and/or affects the performance of students. Such effort could shed some light on specific issues concerning the strengths and weaknesses of LMS and will enable more thorough enhancement of the design and utilization of BL to suit the needs of its users, i.e., students, teachers. This sets the basic rationale for this study. It serves as an invitation to investigate how LMS like Moodle support and promote learning activities among students in blended learning courses and whether or not it promotes higher student academic achievement? A major prerequisite of this investigation is the development of tools that can be used to visualize and monitor students' activities within a given online environment.

III. Course Log Report of the Moodle System

Moodle uses a modular approach, making it easy for teachers to create new courses as well as adding content that will engage students. It claims to support social constructionist pedagogy [15], an approach to learning that promotes students' interaction with the learning material, allowing them to construct new materials and to interact with other students about these materials. To achieve this, Moodle has a flexible array of module activities and resources for creating five types of static course material: (i) a text page, (ii) a web page, (iii) a link to anything on the Web, (iv) a view into one of the course's directories and (v) a label that displays any text or image. Moodle also defines six types of interactive course materials (assignments, choice, journal, lesson, quiz, and survey) and provides five kinds of activities where students can interact with each other (chat, forum, glossary, wiki and workshop).

With all these features available for use, logging of activities performed by students is an essential component of the Moodle interface [15]. Logging involves the process of keeping records of what activities students are doing. The logged data, in turn, allows instructors to request reports telling which resources and activities in a course have been accessed, when, and by whom. One such report is the course log report which provides a list of time-stamped actions showing all class members, with details about which resources or activities have been accessed and where. Table I enumerates and describes the six data dimensions of the course log reports provided by Moodle.

This data can provide rich information related to students' activity in the system. A disadvantage, however, is that the course log data is usually provided in tabular format with poor logical organization (The log data can be downloaded in text-only, ODS, or Excel format, See Fig. 1).

TABLE I. Descriptions of the Dimensions of Moodle Course Log Report

Data Dimension	Description
Course	Course in which the action is related
Time	Date and time stamp of when the action was executed
IP Address	Unique numerical label used 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

In addition, its voluminous size often makes it incomprehensible and difficult to follow. This means that, although the data can be useful in checking whether everyone has done a certain task, it requires some method of processing that will convert it into a friendlier format before any faculty can extract useful statistics that can be used for evaluating student activities and use this to interpret course outcomes.

Although specific data mining tools are already available that can be used to process this type of data, most are too powerful and too complex for use by faculty without technical skills and their output often exceeds the ability of faculty to interpret them. One of our objectives is to devise simple measures of analyzing the log data that is more intuitive and faculty-friendly, with as little parameter as possible in order to simplify the resulting visualization of information and make it more meaningful to the faculty. In achieving this, a novel approach is proposed, as explained in details in the following section, whereby the log data of Moodle is processed to gain some perspectives of the students' activity profile per class and analyze how the opportunities provided by Moodle affects their learning strategy.

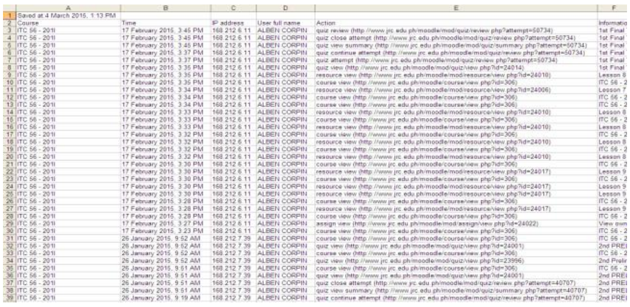

Figure 1. Moodle log report

IV. Data Processing

After data collection, the first question that was addressed is how to process the dataset in order to extract patterns that can be used to visualize students activity. Following the data mining process described by Reference [4], as shown in Fig. 2, a two-phase process was used which include initial pre-processing of data then afterward applying data mining algorithms that transform the data into a form suitable for interpretation and evaluation.

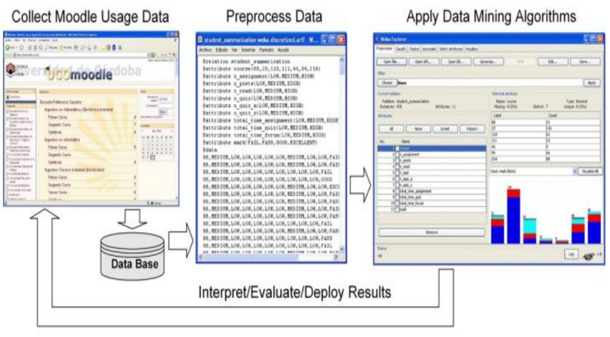

Figure 2. Data processing model

A. Data Preprocessing

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 dataset extracted in Moodle can have missing values, noisy data, and/or irrelevant and redundant information. Here, the actions logged by instructors and course administrators were selectively removed and the dataset was anonymized by removing each student's name and replacing it with a unique identification number. Processing then started by filtering the dataset 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 II shows the initial set of action types examined in this study for analyzing students' activity. 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 behavior. Thus, when categorized based on class activities, the actions help 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. Fig. 3 shows a sample table generated after pre-processing the raw data files. The process of counting this value was automatically done using a customized Excel macro.

TABLE II. Moodle Action Identifiers

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

	Students	course viev	resource vi	url view	forum add (	forum view	forum view	quiz view	quiz attemp	assign subi	assign view	Final Grade	
1	1001	116	29	36	0	0	3	47	8	0	0	1.8	1
2	1002	79	44	7	2	12	10	36	7	0	1	1.9	1
3	1003	46	7	0	0	0	0	36	7	0	0	2.2	1
4	1004	50	34	8	3	8	13	35	7	0	0	2.3	1
5	1005	32	15	0	0	0	0	24	6	0	0	2.3	1
6	1006	20	2	6	0	0	2	21	7	0	0	2.4	1
7	1007	51	45	28	3	15	23	25	7	0	3	2.4	1
8	1008	39	23	15	1	2	7	19	8	0	0	2.4	1
9	1009	93	56	30	0	0	0	28	7	0	0	2.5	1
10	1010	69	18	4	0	0	0	40	7	0	0	2.5	1
11	1011	98	27	8	0	2	3	37	8	0	0	2.5	1
12	1012	109	15	21	1	17	13	33	8	0	0	2.6	1
13	1013	63	17	0	1	5	8	25	7	0	4	2.7	1
14	1014	48	12	20	0	0	5	20	6	0	2	2.8	1
15	1015	41	6	0	0	0	1	39	8	0	0	2.9	1
16	1016	30	1	0	0	0	2	30	7	0	0	2.9	1
17	1017	35	8	5	1	4	0	24	7	0	0	2.9	1
18	1018	63	35	2	0	0	0	26	7	0	0	3.2	1
19	1019	41	25	8	1	1	6	21	7	0	0	3.4	1
20	1020	27	1	2	0	0	1	23	6	0	0	3.4	1
21		0	0	0	0	0	0	0	0	0	0	5	0
	Total	1150	420	200	13	66	97	589	142	0	10	2.7143	

Figure 3. Preprocessed raw data (activity logs)

B. Data Mining Algorithm and Vector Space Model

Data mining algorithms enable extraction and visualization of patterns of activity that can be used to infer students' behaviour. For this purpose, this study makes use of a technique known as Vector Space Model.

Vector Space Model (henceforth VSM) is a statistical model of representation often used in processing documents in Information Retrieval [16]. The main idea behind VSM is to construct vector representation for documents and use these vectors to analyse 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 x n matrix can be built as shown in Fig. 4. In this matrix, D1 to Dk represent the set of documents while W1 to Wn 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.

	w_1	w_2	w_3	...	w_n
D_1	14	6	1	...	0
D_2	0	1	3	...	1
D_3	0	1	0	...	2
⋮	⋮	⋮	⋮	⋱	⋮
D_K	4	7	0	...	5

Figure 4. A k x n document by words matrix

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. Fig. 5 depicts what a 3-dimensional semantic space looks like along with the documents plotted in this space using vector representation.

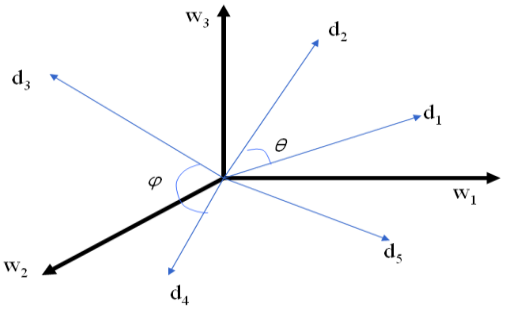

Figure 5. 3D Semantic space

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 D1 is to the topic of D2, VSM simply measures the distance of D2 relative to the position of D1. 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:

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

The cosine formula returns a value between 0 and 1. The assumption 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 similarity between the two documents, the higher the value, the more related they are considered.

C. Representing Student Activity Using VSM Representation

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 zero means that the action type was not initiated at all. For instance, in Fig. 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 determine how similar their level of activity is. Second, it can be used to identify activity patterns that occur within each group of students. In this paper, the latter approach is explored.

The color-coded headers for the elements of the vectors depicted in Fig. 3 indicates the type of class activity to which each action type belongs such as: content access, forum engagement, and assessment activity. This coding can be used to construct an archetypal activity vector for each type of activity class. This can be done by simply setting the corresponding action types for each activity to a non-zero value while the rest of the action type values are set to zero. Thus, the archetypal vector for each activity class would be as shown in Table III:

TABLE III. ARCHETYPAL ACTIVITY VECTOR FOR EACH 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

Given this, each student's activity vector can be processed by comparing them to the archetypal vector of each activity using the cosine distance formula. This will enable student's activity to be analyzed and compared on a per course/class basis.

The development of easily interpretable graphic that can depict trends in student activity based on action logs is a useful tool for instructors and course administrators to constantly visualize and monitor course progress with minimal effort in order to determine whether and how the environment actually affects student performance. The graphs in Fig. 6 (A-C) shows representative visualizations of the cosine values generated for students in each course. Each point in the graph represents a student action generated by students while participating in the course. Although students are anonymously depicted, the graph clearly indicates the overall degree of activity among the cohort, and can possibly be even be refined to drill down to each individual student's level of activity.

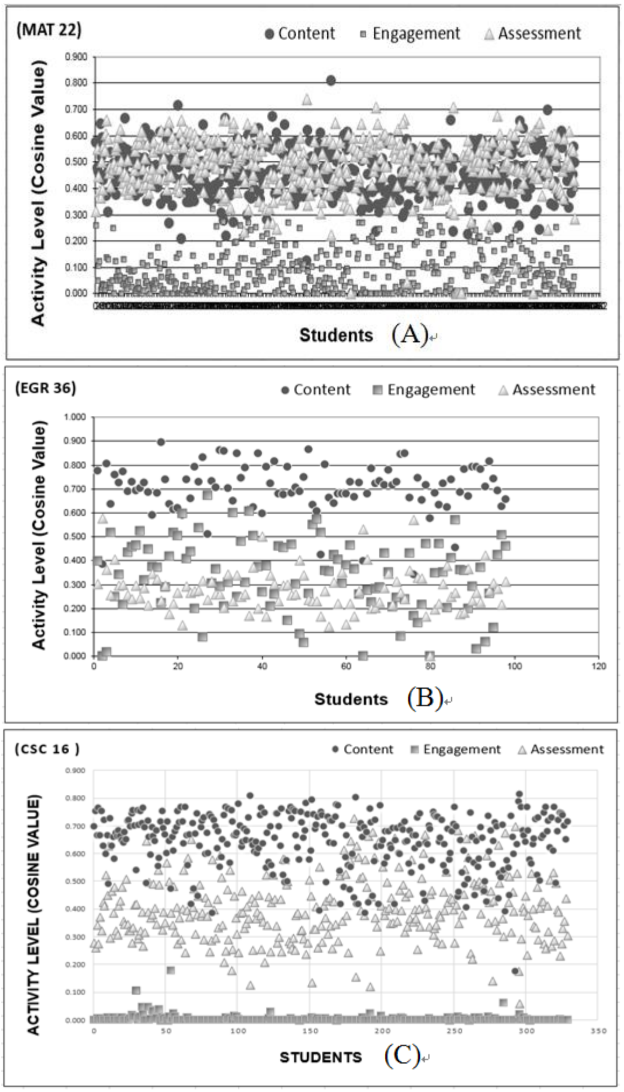

Figure 6. (A-C): Graphic visualizations of actions logs per course generated using VSM

The visualizations clearly depict some patterns of students' 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 provided within Moodle. However, within a certain class, students take on a similar set of behavior in allotting time and effort between different tools and activities.

Students from CSC16, for example, generally log-in into Moodle to access lecture materials with little regard for forum engagement. MAT22 students display an equal level of preference for accessing lecture materials and taking assessment activities with some forum discussions initiated. Whereas EGR16 students seem to prioritize content access and forum engagement over access to assessment tasks. These visualizations can help course administrators in determining the type of strategic interventions that each class/course would need to ensure that student's activities are kept in line with the intended pedagogical outcomes.

Unfortunately, while some of these online activities depicted are likely to translate into effective teaching strategies, the visualizations (along with the cosine ratings) doesn't seem to correlate with the students' course accomplishments. This suggestion comes from observing that some class with a low level of online activity, e.g. CSC16 (average grade: 3.01, average cosine content access: 0.644, assessment: 0.392, engagement: 0.003) have higher average grades than a course with higher levels of activity, e.g. EGR36 (average grade: 3.55, average cosine for content access: 0.700, assessment: 0.276, engagement: 0.335). This seems to suggest that more time spent on online activities' does not convert to higher course achievement. An implication of these observations is that there is a need to redesign the online component more effectively in order to achieve quality instruction.

In particular, the 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. But of course, more studies are needed to determine which pattern of teaching approach would be most beneficial to the students.

V. Discussion and Future Work

This paper proposes a novel approach for processing course log data obtained from blended courses using Moodle in order to mine and visualize patterns of student activity. Action types are processed and represented using VSM techniques borrowed from the IR field. The results indicate that the VSM-based representation can visualize the differences in level and type of activity preferences of students per class. In the long run, these type of visualizations could be used to monitor which class requires immediate and specific pedagogical adjustments. At the moment, however, it indicates that there is a need to promote a more standard approach in utilizing the various tools provided by Moodle as there seems to be a high level of variability on how these tools are used on a per class basis. It is highly likely that the design and nature of the course as well as the individual teaching strategy of instructors are introducing factors that promote this variability. Further inquiry comparing the action profile of instructors with those of students can help provide a better perspective on this issue.

References

[1] A. Norberg, C. D. Dziuban, and P. D. Moskal, "A time-based blended learning model," *On the Horizon*, vol. 19, no. 3, pp. 207-16, August 2011.

[2] J. M. Nicdao, "Teaching literature through blended learning in the Philippines," in *Proc. EDULEARN13*, IATED, 2013, pp. 2817-2822.

[3] H. Kanuka, C. Brooks, and N. Saranchuck, "Flexible learning and cost effective mass offerings," presented at the Improving University Teaching (IUT) Conference, Vancouver, CA, 2009.

[4] C. Romero, S. Ventura, and E. García. "Data mining in course management systems: Moodle case study and tutorial," *Computers & Education*, vol. 51, no. 1, pp. 368-384, August 2008.

[5] J. Mostow, J. Beck, H. Cen, A. Cuneo, E. Gouvea, and C. Heiner. "An educational data mining tool to browse tutor-student interactions: Time will tell," in *Proc. Workshop on Educational Data Mining, National Conference on Artificial Intelligence*, AAAI Press, 2005, pp. 15-22.

[6] E. W. Black, D. Beck, K. Dawson, S. Jinks, and M. DiPietro, "Considering Implementation and use in the adoption of an LMS in online and blended learning environments," *TechTrends*, vol. 51, no. 2, pp. 35-53, March 2007.

[7] M. Driscoll, "Blended learning: Let's get beyond the hype," *E-learning*, vol. 1, no. 4, pp. 1-4, March 2002.

[8] J. Hofmann, "Why blended learning hasn't (yet) fulfilled its promises: Answers to those questions that keep you up at night," in *Handbook of Blended Learning: Global Perspectives, Local Designs*, C. J. Bonk and C. R. Graham, Eds., 2006, pp. 27-40.

[9] OECD. "E-learning in tertiary education: Where do we stand?" Paris, 2005.

[10] M. Bullen, "Reconsidering the learning management system," *Issues in E-Learning & Distance Education Video Series*, July 2014.

[11] C. Dalsgaard, "Social software: E-learning beyond learning management systems," *European Journal of Open, Distance and E-Learning*, vol. 9, no. 2, July 2006.

[12] C. Heaton-Shrestha, S. May, and L. Burke, "Student retention in higher education: What role for virtual learning environments?" *Journal of Further and Higher Education*, vol. 33, no. 1, pp. 83-92, February 2009.

[13] S. Jones, C. Johnson-Yale, S. Millermaier, and F. S. Pérez. "Academic work, the internet and US college students," *The Internet and Higher Education*, vol. 11, no. 3, pp. 165-177, December 2008.

[14] M. Christie and R. G. Jurado, "Barriers to innovation in online pedagogy," *European Journal of Engineering Education*, vol. 34, no. 3, pp. 273-279, June 2009.

[15] H. William, *Moodle: E-learning Course Development: A Complete Guide to Successful Learning Using Moodle*, Packt Publishing, 2006.

[16] V. V. Raghavan and S. M. Wong "A critical analysis of vector space model for information retrieval," *Journal of the American Society for Information Science*, vol. 37, no. 5, p. 279, Sept. 1986.

Rodolfo C. Raga Jr. received his PhD in Computer Science degree from the Dela Salle University, Manila, Philippines in 2013. He's an Associate Professor in the College of Computer Studies and Engineering of Jose Rizal University, Philippines. His research interests lie in educational data mining, learning analytics, and academic e-learning.

Jennifer D. Raga received her Master in Information Technology degree from the University of LaSalette, Santiago City, Philippines in 2009. She is currently the MIS Manager at Western Marketing Corporation and a part-time IT lecturer. Her research interests lie in knowledge management, business analytics, and corporate e-learning.

Israel V. Carino received his Master in Information Technology degree from the University of the Cordilleras in Baguio City, Philippines in 2009. He is currently the Chair of IT Program at the College of Computer Studies and Engineering of Jose Rizal University, Philippines. He is a former Software Engineer at Infor Philippine Solutions and Service Center in Bonifacio Global City, Philippines.