Issues in Information Systems
Volume 16, Issue IV, pp. 102-110, 2015

USING *K*-MEANS CLUSTERING TO MODEL STUDENTS' LMS PARTICIPATION IN TRADITIONAL COURSES

Klara Nelson, The University of Tampa, krnelson@ut.edu

ABSTRACT

The focus of this research is on the relationship between student participation in a learning management system(LMS) in traditional courses and course grades using Blackboard Learn tracking data from two undergraduate courses taught by the author from January to May 2015. The results are consistent with prior research that found a positive relationship between LMS participation and student achievement. Correlation analysis showed significant and positive relationships between the students' course grade and their frequency of access overall as well as frequency of access to course materials. In addition, detailed LMS participation profiles were obtained from using k-means clustering, an unsupervised data mining method. The significant correlations between course grade and frequency of access variables are also evident in the 5-cluster solution that emerged. Despite the small sample size, the present study shows the usefulness of k-means clustering, a data mining method, for better understanding students' LMS participation.

Keywords: Learning Management Systems, LMS Participation, Academic Performance, Educational Data Mining, *k*-Means Clustering

INTRODUCTION

Learning management systems (LMS) such as Blackboard are not only used extensively for online and hybrid course delivery but also as a supplement to traditional course instruction. Most students in traditional campus settings view an LMS a good addition to face-to-face instruction, particularly the ability to access grades, assignments, course materials/resources and the syllabus [14].

Active student participation and engagement in a traditional classroom tends to have a positive effect on achievement with students earning higher grades as their participation increases [11]. Using the large amount of user activity and interaction data tracked by learning management systems, a substantial body of research has examined the relationship between online activity and performance in online and hybrid course environments. Online activity has been used as a measure of participation and an indicator of performance as it reflects both attendance and study time to some degree [22]. Gillingham [9] suggests that 'participation' of students in an LMS rather than 'interaction' is important, with participation being defined as both passive (reading) and active (posting messages).

The focus of this research is on examining LMS participation in traditional face-to-face courses where the LMS is used extensively to supplement instruction. Following Dawson et al. [7] the intent of this study is to examine the actual LMS behavior trends "with a view to identify potential differences between high academically performing students and those requiring early learning support interventions" (p. 226).

This study uses correlation analysis and cluster analysis, an unsupervised data mining technique to uncover natural groupings in a data set based on common characteristics they share. Educational data mining (EDM) is considered an emerging "area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand students and the settings which they learn in" [1].

Using Blackboard Learn tracking data from two undergraduate courses taught by the author during Spring 2015, this exploratory study aims to address the following research questions:

1. Which LMS participation variables (if any) correlate significantly with student achievement as measured by the course grade?
2. Are there different participation profiles that characterize students' LMS activity? If so, do they relate to course performance?

LITERATURE: LMS PARTICIPATION AND STUDENT PERFORMANCE

Online Courses

Several studies have shown a positive relationship between LMS participation and student achievement in online course environments. A study of students in 13 sections of three lower division online courses concluded that students who are more engaged with the content and discussions in an online course will persist and complete successfully. Specifically, number of discussion posts viewed, number of content pages viewed, and seconds on viewing discussion pages were statistically significant predictors of final grades [18]. A study of student engagement in an online ethics course found a significant relationship between overall site usage and the students' final course grade [15]. Both attendance and the volume of participation as identified by the students' activity within the Blackboard LMS were found to have a positive influence on student achievement in a study of 548 undergraduate online students [20]. Davies and Graff [6] found that students achieving high or medium passing grades engaged more actively with the course in terms of Blackboard accesses than students achieving low passing grades. Similarly, Pratt-Philips [19] reports a significantly positive relationship between online activity (number of online sessions, files viewed, total time online) and student final grade among students who completed the course.

Hybrid/Traditional Courses

Findings from studies of student LMS usage in hybrid courses are similar to those of online courses. Whitmer [23], however, suggests that the frequency of student LMS use "is more predictive of students success in the fully online environment than in a hybrid environment, in which some learning activities are conducted outside of the LMS" (p. 116). Biktimirov and Klassen [3] examined the value of online support materials in a hybrid course and found that students' access to homework solutions was the only significant predictor that had a positive relation with the course grade. Examining Hit Consistency and Total Hits as different aspects of online activity, Baugher, Varanelli and Weisbord [2] found that only Hit Consistency related to course success which was measured as the average of graded course components. Campbell [4] as well as Lauría et al. [13] found significant, albeit low correlations between course grades and LMS event frequencies for the following areas: Sessions Opened, Content Viewed, Discussions Read, Discussions Posted, Assignments Submitted, and Assessments Submitted. Whitmer [23] examined LMS usage frequency and its effect on performance in terms of the four areas of the LMS tool design framework described by Dawson et al. [7]. Content activity frequency and Administrative activity frequency emerged as significant predictors, while Assessment activity frequency and Engagement activity frequency did not. Both Campbell [4] and Whitmer [23] also included student characteric variables in their studies. Campbell found that student characteristic variables had a considerable greater impact on the prediction of student achievement than LMS use while Whitmer found the opposite: LMS variables were over four times as strongly related to achievement as demographic variables. A study of Serbian students showed a strong significant correlation between a student's interactive usage of the LMS and his/her active participation in class. Furthermore, both students' active participation in class and interactive usage of an LMS had a high predictive capacity on student achievements in a blended learning environment [17]. Aside from event frequencies, Dawson et al. [7] also compared session times and academic performance of low versus high performing students. Findings from the Dawson study indicate significantly higher frequency of online session times as well as greater total time online. However in terms of total time per session, no significant differences were found between low and high performing students. The authors argue that low-performing students may not be optimizing their learning time online in the same way their higher-performing counterparts are able to" (p. 227).

RESEARCH METHODOLOGY

Sample

Data analyzed in this study was extracted from tracking data captured in the institutional LMS Blackboard Learn for two traditional 400-level courses taught by the author that together enrolled 48 senior-level students. Demographic data obtained from the class rosters showed that only 19% of the students across the two courses were female, and that 73.8% were MIS students (see Table 1). Age and GPA data shown in Table 1 were only available in the aggregate at the course level. Students were on average 24 years old which is consistent with the their senior class standing.

Table 1. Sample Characteristics

Variable	M	Minimum	Maximum
Age	24	20	47
GPA	3.15	2.06	4.00
	n		%
Gender			
Male	34		81.0
Female	8		19.0
Academic Major			
Economics	1		2.4
Entrepreneurship	2		4.8
Finance	2		4.8
International Business	1		2.4
Management	3		7.1
Marketing	2		4.8
MIS	31		73.8

Measures of Online Course Activity and Performance

To supplement the traditional face-to-face delivery format that consisted of lectures and many in-class learning activities (discussions, hands-on exercises, in-class exams and quizzes), both courses made extensive use of Blackboard to post information/announcement, support materials (Power Point slides, data files), assignments, study guides, sample exams/quizzes, and all grades. While students were required to use Blackboard's Assignments feature to access and submit all assignments, support materials were non-obligatory. All students had previously used Blackboard in many other classes given their senior class standing and Blackboard being the required institutional LMS.

Two evaluation reports, the *All User Activity inside Content Areas* report and the *Course Activity Overview* report, were generated for the time period January 18, 2015 to May 11, 2015 and downloaded as Excel spreadsheets. Data collected on each student included counts of accesses to the different LMS areas, the total number of accesses, and the total course activity in hours. Because of differences in the courses with respect to LMS content provided to students, only comparable metrics were used specifically the frequency of accesses to Assignments, Syllabus, and Course Materials. These three features address two of the four areas of the LMS tool design framework described by Dawson et al. (2008): Assessment (Assignments), and Content (Syllabus and Course Materials). To make metrics comparable across courses, access counts for the different LMS areas were expressed as percent of the total count for a specific content area in each course. Similarly, total number of accesses and hours of activity per student were measured in terms of percent of the overall total in each course. Measures of student performance included the weighted course average of the graded course components, which ranged from 0 to 100, as well as the student's letter grade for the course.

ANALYSIS METHODS

IBM SPSS Statistics 23 was used to generate descriptive statistics (Table 2), and explore the correlation between students' LMS activity and their course grades. Cluster analysis was conducted with IBM's data mining software SPSS Modeler 17 to investigate similarities and differences between students regarding their use of Blackboard Learn throughout the semester. A cluster represents a collection of records that are similar to one another, and dissimilar to records in other clusters. Membership in a cluster is determined based on some measure of distance between cases. As an unsupervised data mining technique, the clustering task does not have a known dependent variable, and does not try to classify, estimate or predict the value of a dependent variable [12]. Cobo et al. [5] for example used agglomerative hierarchical clustering to model learner participation profiles in online discussion forums.

Table 2. Descriptive Statistics (n = 42)

Variable	Mean	Std. Dev.		Standardized Variables	Mean	Std. Dev.
Course Activity Hours	26.40	11.67		%HoursCourse	4.06	1.79
Assignments	29.81	14.68		%AssignmentsCourse	3.94	1.17
Class Materials	59.93	25.40		%MaterialsCourse	3.96	1.50
Syllabus/Schedule	14.38	10.15		%SyllCourse	3.83	2.59
Total Hits	150.60	65.71		%TotalHits	3.96	1.21
WeightedGrade	82.42	13.30				

For this study, the k-means clustering method was chosen because all the standardized input variables shown in Table 2 are continuous. Because k-means has trouble clustering data that contains outliers [21], two students were eliminated from the analysis. Four students were in both courses. They were eliminated from one of the courses to avoid duplicates after confirming that each of them exhibited similar LMS usage behavior in both courses. A first cluster analysis showed that each of these four students loaded into the same cluster twice, i.e., had similar access patterns. The algorithm requires the user to specify the number of clusters k that are desired. When there is no a priori knowledge of the number of underlying clusters in the data set, Larose and Larose [12] suggest to cycle through various promising values of k and compare clustering solutions for each value of k using some measure of cluster validity as well as domain expertise. Consequently, the Auto-Cluster modeling node was used to execute four models with k = 2, 3, 4, and 5, and to display the top 3 models ranked by Silhouette coefficient in descending order. Ranging between -1 and 1, the silhouette coefficient is an intrinsic measure of clustering quality based on both cohesion within a cluster and separation between the clusters [10]. Cohesion refers to how tightly related the records are within the individual clusters, while separation addresses how distant the clusters are from each other [12].

RESULTS

Research Question 1: Which LMS participation variables, (if any), correlate significantly with student achievement as measured by the course grade?

Table 3 presents correlations for each of the LMS activity variables and the students' course grade as measured by the weighted average of graded components. All associations are positive, but only content related hits on Course Materials (%Materials, r = .446, p <.01), and overall activity in terms of Total Hits (%TotalHits, r = .408, p <.01) are moderate in strength [8] and statistically significant consistent with Little-Wiles et al. [15] and Dawson et al. [7]. Correlations between students' course grade and time spent online (% Activity Hours), Assignment related-activities (downloading / uploading files, reading files online), and accesses to the Syllabus are weak and not statistically significant. The lack of a significant correlation between assignment accesses and course grade is not surprising. All students were required to use the Assignment feature to access assignment details (instructions, files) and upload their work for grading. The Materials area on the other hand featured a substantial amount of supplemental non-obligatory information to help students with understanding of course content.

Table 3. Correlations (n = 42)

		%ActivityHours	%Assignments	%Materials	%Syllabus	%TotalHits
Weighted Course Grade	Correlation	.220	.161	.446**	.184	.408**
	Sig.	.161	.307	.003	.242	.007

**p<.01

Research Question 2: Are there different participation profiles that characterize students' LMS activity? If so, do they relate to course performance?

LMS Participation Profiles

Of the three models retained by SPSS Modeler, the 4-cluster solution had the highest Silhouette coefficient of .484 which indicates that there is fair evidence of the reality of the clusters in the data [12]. The smallest cluster contained 6 students (14% of the sample), the largest cluster consisted of 14 students (33%). LMS participation profiles are shown in Table 4 and Figure 1.

Table 4 provides the size of each cluster, mean values for the variables used to form the clusters, and the average course grade for each cluster. To allow for easy comparison between clusters, Table 4 also shows the ranking of each variable mean across the four clusters. Gender and academic major information for each cluster are provided in Table 4 as well. The proportion of students majoring in MIS in Table 4 serves as a proxy for tech-savviness to assess if more tech-savvy students use the LMS more frequently or extensively. The data in Table 4 does not reveal such a pattern. Cluster 3 with the lowest LMS participation has the 2nd highest proportion of MIS students, while Cluster 4, the fully participating cluster has the 2nd lowest proportion of MIS students.

Table 4. Cluster Profiles with Variable Means and Rankings Across Clusters

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Descriptive Label	*Syllabus-Focused, Good Performance*		*Highly Engaged, Very Good Performance*		*Minimally Active, Low Performance*		*Fully Participating, High Performance*	
• **Clustering Variables**	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
- Assignments	3.53	3	4.53	2	2.95	4	5.37	1
- ActivityHours	3.99	3	4.63	2	2.40	4	6.39	1
- Syllabus	6.25	2	2.79	3	1.64	4	7.35	1
- Materials	2.99	3	4.88	2	2.85	4	5.68	1
- TotalHits	3.63	3	4.49	2	2.74	4	5.86	1
• **Overall LMS Participation**	4.08	3	4.26	2	2.52	4	6.13	1
• **Course Grade**	81.93 (BC)	3	84.15 (B)	2	76.58 (C)	4	91.80 (AB)	1
	n	%	*n*	%	*n*	%	*n*	%
Cluster size	9	21.4	14	33.3	13	30.9	6	14.3
MIS Majors	8	88.9	9	64.3	10	76.9	4	66.7
Female Students	2	22.2	2	14.3	2	15.4	2	33.3

Note: 1 = Highest rank, 4 = Lowest rank

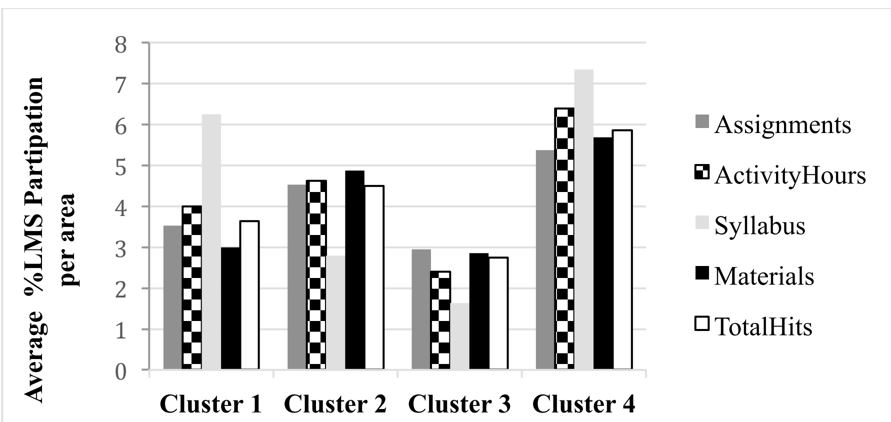

Figure 1. Visual Presentation of Cluster Variable Means

A visual presentation of the variable means used for clustering is provided in Figure 1. As shown in Figure 1 and Table 4, k-means clustering to profile LMS usage of students in a traditional course environment resulted in four distinct groups of students regarding both type and level of engagement with the institutional LMS Blackboard. Descriptive labels in Table 4 reflect the differences in students' LMS participation described below.

Focusing horizontally on each component of LMS activity, students in Cluster 3 tend to be the most active participatns as they rank first on all dimensions of LMS activity. In contrast, students in Cluster 3 participate the least with the lowest mean and rank on every LMS activity variable. Cluster 2 students rank 2nd on all the LMS activity variables with the exception of Syllabus access frequency which has the second-lowest mean of 2.79. Cluster 3 students rank third on all the LMS activity variables with the exception of Syllabus access frequency, which has a relatively high mean of 6.25. The participation patterns in Figure 1 further reveal that Cluster 1 and Cluster 4 are similar in terms of relative ranking of variable means within each cluster, particularly high access frequency to the syllabus. However, all means are lower in Cluster 1 than in Cluster 4. Clusters 2 and 3 have similar patterns in terms of the relative height of each variable mean within the cluster, particularly the low access frequency to the syllabus with Cluster 3 students having the lower means on each activity.

Focusing on variable means within each cluster, both Clusters 4 and 1 distinguish themselves by the high frequency of accesses to the syllabus with means of 7.35 and 6.25 respectively. In both clusters, Syllabus access represents the highest mean. Notable about Cluster 1 is further that accesses to Materials have the lowest mean. Within both Clusters 2 and 3, Syllabus accesses are associated with the lowest mean in each cluster (2.79 and 1.64 respectively). Access to Materials has the highest mean among the LMS variables within Cluster 2 while access to Assignments has the highest mean of all the variable means witin Cluster 3.

LMS Activity Profiles and Course Performance

Letter grade distributions per cluster are presented in Table 5 and are visualized in Figure 2 to further profile the clusters and examine the relationship between LMS usage and student performance.

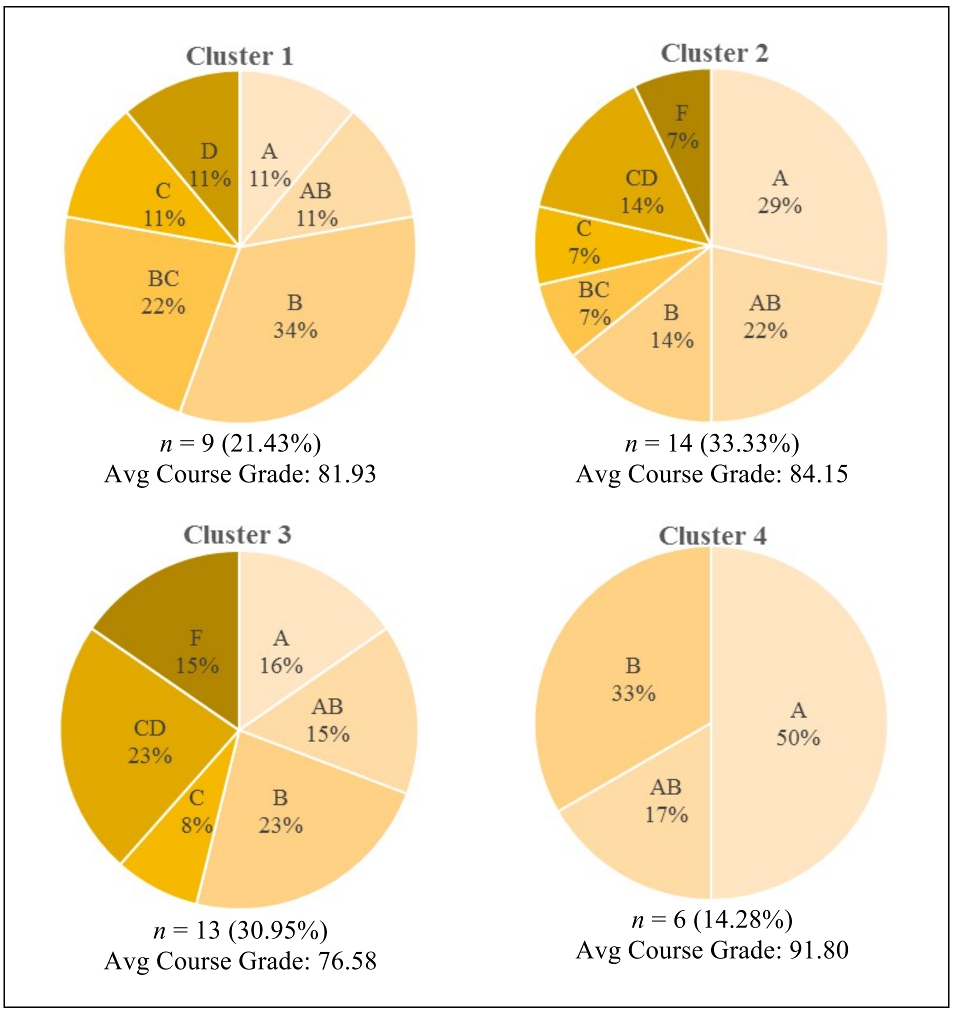

Figure 2. Letter Grade Distributions by Cluster

Table 5. Letter Grade Distributions by Cluster

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Number of Students per Letter Grade	A	1	4	2	3
	AB	1	3	2	1
	B	3	2	3	2
	BC	2	1	0	0
	C	1	1	1	0
	CD	0	2	3	0
	D	1	0	0	0
	F	0	1	2	0

Tables 4 and 5 as well as Figure 2 show a clear alignment between overall high versus low LMS participation in terms of course performance. The highest performance Cluster 4 shows all-around high LMS usage while the lowest-performance Cluster 3 has the lowest means and ranks last on every dimension of LMS usage. Also, each cluster is associated with a different weighted course average and letter grade. Based on the students' LMS usage and course performance, the empirical clusters that emerged from this analysis can be described as follows.

Cluster 4 - Fully Participating, High Performance Cluster. The highest weighted course grade of 91.80 (letter grade AB) is associated with Cluster 4 students who make the highest usage of the LMS in all respects. This fully-participating cluster is the most homogenous with respect to course letter grades consisting of only A, AB, and B students. 67% of this cluster consists of students who earned an A or AB, the remaining 33% earned a B grade. Students in this cluster are very engaged with the course outside of the classroom in terms of both activity hours and total accesses. They also consult the syllabus and assessment materials most often, and engage with non-obligatory discipline content most extensively.

Cluster 2: Highly Engaged, Very Good Performance Cluster**.** Similar to Cluster 4, this cluster with an average grade of 84.15 (letter grade B) has a high proportion of A and AB students (51%), and another 21% of students earning B and BC letter grades. With overall second-highest LMS usage, these students engage considerably with non-obligatory course materials made available by the instructor. At the same time, they are less concerned with consulting the syllabus.

Cluster 1: Syllabus-Focused, Good Performance Cluster. Third-ranked Cluster 1 with an average course grade of 81.93 (letter grade BC) consists primarily of good students (56%) who earned grades of B or BC. While these students appear to pay considerable attention to the syllabus, their engagement with non-obligatory course material posted to Blackboard is low and similar to that of students in Cluster 3.

Cluster 3: Minimally Active, Low Performance Cluster. Finally, the lowest mean course grade across the clusters (76.58, letter grade C) goes hand in hand with the group of students who participate the least on all dimensions of LMS usage. The proportion of students who earned a grade below C is 38% which is the highest proportion of students in this grade range among the 4 clusters. At the same time, almost a third of the students in this cluster (31%) are A and AB students. Within this minimal usage cluster, assignment accesses, which are required, represent the highest area of engagement.

CONCLUSIONS

The focus of this study was on LMS participation and student performance in two face-to-face undergraduate courses that used Blackboard Learn extensively to support traditional classroom instruction. Participation here was conceived as frequency and duration of activity in line with Gillingham [9]. Specific LMS features examined in this study address two of the four areas of the LMS tool design framework described by Dawson et al. [7]: Assessment (Assignments), and Content (Syllabus and Course Materials). The results are consistent with prior research from both online and hybrid course environments that found a positive relationship between LMS participation and student achievement. Correlation analysis showed that frequency of access overall as well as frequency of access to course materials were significantly related with the weighted average students earned in the course. In addition, detailed LMS participation profiles were obtained from using k-means clustering, an unsupervised data mining method, where the number of clusters is *a priori* unknown. Four distinct activity profiles (clusters) emerged. The significant positive correlations between course grade and frequency of access overall and frequency of access to class materials are particularly evident in Clusters 2 and 4 each of which consists of high proportions of A and AB students. The relatively high frequency of access overall and to non-obligatory class materials is indicative of students who participate extensively beyond the face-to-face interaction that occurs in the traditional classroom environment. At the same time, relatively infrequent accesses to class materials in Cluster 1 are associated with a majority proportion of B and BC students (56%). Differences in empirically derived participation profiles are also associated with distinctly different average final grades in each cluster. While the small sample size does not permit generalizations, the results from this study are nonetheless consistent with Dawson et al. [7] who found "strong indications that there are discrete online behavior patterns that are representative of students' overall engagement and final assessment grade".

The present study shows the usefulness of clustering as one approach to investigating LMS participation, and better understanding students and the settings in which they learn. Given the small sample size of the dataset, further research could involve capturing data for these courses over additional semesters. Inclusion of other variables such as demographics (age, GPA) that were beyond the scope of this study, or hit consistency [2] may yield further useful insights. Finally, learning management systems such as Blackboard track large amounts of user activity and interaction data. The analyses conducted here show that the captured tracking variables can be pedagogically meaningful in traditional course environments where much of the interaction takes place face-to-face. LMS-captured data can be viewed as another tool in the arsenal of instructors to get to know the students and measure engagement with the course and its content in a quantifiable manner. As suggested by Macfadyen et al. [16], instructors can benefit from analyzing this data and identify areas that can serve as "early-warning" indicators, in this case access to non-obligatory materials and total accesses to the LMS. Monitoring access to these areas throughout a course can then help flag potential at-risk students that may require additional help or early intervention [16].

REFERENCES

1. Baker, R.S.J.d. (2010) Data Mining for Education. In McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education (3rd edition)*, vol. 7, pp. 112-118. Oxford, UK: Elsevier. Available: http://www.columbia.edu/~rsb2162/Encyclopedia%20Chapter%20Draft%20v10%20-fw.pdf
2. Baugher, D., Varanelli, A., & Weisbord, E. (2003). Student hits in an internet-supported course: How can instructors use them and what do they mean? *Decision Sciences Journal of Innovative Education, 1*(2) , 159–179.
3. Biktimirov, E. N., & Klassen, K. J. (2008). Relationship between use of online support materials and student performance in an introductory finance course. *Journal of Education for Business*, *83*(3), 153-158.
4. Campbell, J. P. (2007). Utilizing Student Data within the Course Management System to Determine Undergraduate Student Academic Success: An Exploratory Study (Doctoral dissertation, Purdue University, 2007). (UMI No. 3287222).
5. Cobo, G., García-Solórzano, D., Eugènia Santamaria, E., Morán, J. A., Melenchón, J., & Monzo, C. (2012). Using agglomerative hierarchical clustering to model learner participation profiles in online discussion forums. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. New York: ACM, 2012, 248-251.
6. Davies, J., & Graff, M. (2005). Performance in e-learning: online participation and student grades. British *Journal of Educational Technology, 36*(4), 657-663.
7. Dawson, S., McWilliam, E., & Tan, J. P.-L. (2008). Teaching smarter: How mining ICT data can inform and improve learning and teaching practice. In *Proceedings ascilite Melbourne 2008,* Available: http://www.ascilite.org.au/conferences/melbourne08/procs/dawson.pdf
8. Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4^{th} ed.). Thousand Oaks, CA: Sage Publications Inc.
9. Gillingham, P. (2009). Ghosts in the Machine: Student Participation and Grade Attainment in a Web-Assisted Social Work Course, *Social Work Education: The International Journal*, *28*(4), 423-435.
10. Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3^{rd} ed.). Waltham, MA: Morgan Kaufmann Publishers.

11. Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *The Journal of Educational Research*, 98, 184-191.
12. Larose, D.T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.
13. Lauría, E. J. M., Baron, J. D., Devireddy, M., Venniraiselvi Sundararaju, V., & Sandeep M. Jayaprakash, S. M. (2012). Mining Academic Data to Improve College Student Retention: An Open Source Perspective. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. New York: ACM, 2012, 139–142.
14. Little-Wiles, J. M., & Naimi, L. L. (2011). A study of traditional undergraduate student engagement in Blackboard Learning Management System. 2011 *Conference Proceedings of the American Society for Engineering Education* (ASEE). Available: https://www.asee.org/public/conferences/1/papers/1264/view
15. Little-Wiles, J., Feldhaus, C., and Fox, P. (2013). Are they engaged? Strategies for determining student engagement online. *Journal of Online Engineering Education*, *4*(2), December 2013. Available: http://www.onlineengineeringeducation.com/v4n2.html
16. Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588-599.
17. Mijatovic, I., Cudanov, M., Jednak, S., & M. Kadijevich, D. M. (2013) How the usage of learning management systems influences student achievement, *Teaching in Higher Education*, *18*(5), 506-517.
18. Morris, L., Finnegan, C. & Wu, S. (2005). Tracking student behavior, persistence, and achievement in online courses. *Internet and Higher Education*, 8, 221-231.
19. Pratt-Phillips, S. (2011). Comparison between Online Activity and Performance in a Distance Education Equine Science Course. *NACTA Journal*, March 2011, 21-25.
20. Rapposelli, J. A. (2014). The correlation between attendance and participation with respect to student achievement in an online learning environment. *Doctoral Dissertations and Projects*. Paper 813. Available: http://digitalcommons.liberty.edu/doctoral/813
21. Tan, P.N., Steinbach, M. and Kumar, V. (2006). *Introduction to Data Mining*. London, Pearson Education
22. Wang, A.Y., & Newlin, M.H. (2002). Predictors of webased performance: the role of self-efficacy and reasons for taking an on-line class. *Computers in Human Behavior Journal*, 18, 151-163.
23. Whitmer, J. (2013). Logging On to Improve Achievement: Evaluating the Relationship between Use of the Learning Management System, Student Characteristics, and Academic Achievement in a Hybrid Large Enrollment Undergraduate Course, Dissertation, Available: https://johnwhitmerdotnet.files.wordpress.com/2013/01/jwhitmer_dissertation_complete_1-21-2013.pdf