Unrevealing the digital thread: Exploring students’ LMS digital behavior and its impact on academic performance in Kuwait higher education.

Ibtisam L. Almutairi
Public Authority for Applied Education and Training
E_almutairi@hotmail.com

Brad McKenna
University of East Anglia
b.mckenna@uea.ac.uk

Adrian Benfell
University of East Anglia UK
a.benfell@uea.ac.uk

Abstract

This study aimed to investigate the influence of students' digital behavior in the Learning Management System (LMS) on their academic performance. Educational Data Mining (EDM) algorithms, specifically clustering analysis, will be used to analyze student log data, specifically within the context of Kuwait University (KU). By utilizing EDM algorithms, various aspects of students' actual digital behavior will be analyzed, including forum posts and views, frequency of logins, files downloaded, attempts and finalization at exams, and quizzes. Then multiple linear regression will be applied to examine the influence of students' digital behavior in the LMS on their academic performance represented by their grades in LMS log data. The findings of this research could help to better understand students' digital behavior through LMS, which can assist in formulating strategies to enhance student engagement and optimize the learning experience. In addition, these findings can inform the design and implementation of LMS at KU, ensuring that it is more closely aligned with the preferences and expectations of students. Since this alignment comes at a cost, it would be wise to invest in it only if it ultimately contributes to enhancing student academic performance which is the question that will be answered in this study.

Key words: Kuwait University (KU), Learning Management System (LMS), Educational Data Mining (EDM), LMS log data, academic performance.

Unrevealing the digital thread: Exploring students' LMS digital behavior and its impact on academic performance in Kuwait higher education.

1. Introduction

In an era marked by the integration of advanced Information and Communication Technologies (ICTs) and the widespread adoption of e-learning systems such as LMS in higher education, it is important to note that the mere availability of these systems does not guarantee improved academic or teaching performance (Al-Fraihat et al., 2020; El-Sayad et al., 2021; Rajabalee & Santally, 2021). This paradox is particularly evident in the context of Kuwait (Ghinea et al., 2013). Despite significant investments in e-learning, the desired benefits have not materialized, and performance has often fallen short of expectations as evident by relatively Low evaluation metrics of KU's performance (QS Quacquarelli Symonds, 2022; RUR Rankings Russian Federation Agency, 2021; Times Higher Education, 2022).

LMS is a server or cloud-based systems that stores and manages information related to users, programs, and content, thereby meeting the needs of all stakeholders within educational institutions or organizations (Veluvali & Surisetti, 2022). Moodle, one open-source LMS, that is widely recognized for its adaptability and extensive capabilities, making it a popular choice among educators and students (Al-Fraihat et al., 2020b). KU, the primary public institution of higher education in Kuwait, has embraced Moodle as the primary system to foster an interactive learning environment with the aim of enhancing educational outcomes (Kuwait University, 2018). The university integrates Moodle as part of a blended e-learning approach rather than as a replacement for traditional instructional methods. While students typically engage in full-time on-campus attendance, the university utilizes e-learning systems, such as Moodle, to maximize their benefits (KU E-Learning Centre, 2018).

2. Literature Review

Students engage in a range of digital activities while utilizing the LMS (Huang et al., 2020). The LMS records the students' clicks, with each clickable function containing a click counter that tracks the students' digital behavior. EDM is a significant area of study focused on analyzing data obtained from the LMS system. EDM algorithms facilitate the examination of various aspects of students' actual digital behavior, such as their time spent on the LMS, frequency of logins, and types of activities performed. This study aims to address existing knowledge gaps concerning the effects of different indicators on student academic performance within blended learning settings (Bessadok et al., 2021), with a specific focus on at KU which would offer practical insights and contribute to this area. In the field of Information Systems (IS)/ e-learning literature, this study aims to fill the gap in understanding methods for visualizing and structuring students' online behavior by analyzing action logs from LMS rather than depending solely on surveys or interviews (Bessadok et al., 2021; Kara & Yildirim, 2022; Leem, 2023). Furthermore, it extends beyond this by examining the impact of students' online behavior on their academic performance, thereby paving the way for potential future research endeavors.

The substantial volume of data produced by LMS logs has led numerous researchers to utilize various analysis methods to address research questions. Yildirim & Gülbahar (2022) employed Moodle engagement analytics and learners' characteristics to forecast superior final performance by utilizing the Decision Tree algorithm, a machine learning technique that constructs a tree-like model of decisions and their potential outcomes to identify the most significant predictors of high final performance. The study findings support that learner behavior is a significant predictor of their final performance (Yildirim & Gülbahar, 2022). The literature has raised concerns about the Decision Tree algorithm due to its limitations in capturing complex interactions and nonlinear relationships and its classification as a black-box model presents difficulties in interpreting the decision-making process and the rules generated by the algorithm (Huang et al., 2020).

Bessadok et al. (2021) employed EDM, specifically K-means clustering, to uncover the digital activities of students recorded within the Blackboard LMS, including uploaded files, viewed courses, completed quizzes, and finalized homework assignments. Subsequently, the authors utilized univariate analysis of variance (ANOVA) to evaluate the impact of students' digital activities on their academic performance. The research findings indicated that the analysis of data derived from an LMS platform is essential for enhancing student achievement in e-learning courses (Bessadok et al., 2021). Riestra-González et al. (2021) used similar analysis methods to predict student performance using LMS log data. The research revealed that students who demonstrated promptness in completing quizzes outperformed their peers, whereas those who procrastinated and accessed course materials later exhibited poorer performance (Riestra-González et al., 2021). Clustering analysis has demonstrated its efficacy in discerning patterns, correlations, and outliers within a complex and multifaceted dataset, such as LMS log data which can be instrumental in facilitating informed decision-making and identifying potential issues or challenges within the LMS system (Ramadan et al., 2020). The following section will discuss the literature take of conceptualizing students' LMS data log activities/actions to certain research variables in order to develop research hypothesis.

2.1.Content Interaction

This study specifically focuses on student-content interaction, which can be defined as the manner in which students interact with instructional materials and planned activities within the learning systems (Muir et al., 2022). The literature refers to students' actions registered in the LMS related to all files uploaded/download on the course, and course views as student content interaction (Bessadok et al., 2021; Muir et al., 2022; Vlachopoulos & Makri, 2019). The course view refers to the process of signing into the user interface in a LMS that displays the tasks assigned to students (Bessadok et al., 2021). Buckley et al. (2021) indicated that the number of views a course receives serves as a reliable indicator of a student's interest and understanding of a subject. Furthermore, Vlachopoulos & Makri (2019) suggests that students who choose to download their course materials more frequently demonstrate greater independence in their academic pursuits. These interactions may also signify students' intention to review the information, which can aid in memorization and comprehension. Vlachopoulos & Makri (2019) and Leem (2023) discovered that the interaction between students and the content in e-learning platforms has a direct and positive influence on their performance. Additionally, Muir et al. (2022) underscores the significance of learner-content interaction in promoting student engagement and improving their learning outcomes. As a result, it can be hypothesized that:
H1: Content interaction has a positive effect on student's academic performance.

2.2.Student Engagement

The LMS plays a role in facilitating student engagement, with a specific focus on the significance of student behavioral engagement in relation to LMS usage (El-Sayad et al., 2021). This study will focus on behavioral engagement as it is the engagement type that is documented in the LMS log data, defined as the level of involvement and actions demonstrated by students

in their learning and academic pursuits, as measured through active participation in discussions, consistent efforts, and contributions to class discussions (Salas-Pilco et al., 2022). In an e-learning context, it is more feasible to monitor and assess students' engagement behavior, such as their participation in online discussions, Forum posts, Forum views (Lu & Cutumisu, 2022; Muir et al., 2022; Salas-Pilco et al., 2022). Student engagement with LMS through structured and facilitated discussions, such as those in forums or discussion boards, can enhance the quality of relationships and cultivate an environment conducive to voluntary engagement which offers opportunities a deeper thinking (Vlachopoulos & Makri, 2019). Huang et al. (2020) utilized data on students' online discussion activities and their findings indicated that students who were more engaged and participated more actively in online discussions tended to achieve higher academic performance compared to those who were less active or did not participate at all (Huang et al., 2020). Additionally, Vlachopoulos & Makri (2019) asserted that students who actively engage in online courses are more likely to attain improved learning outcomes. Therefore, it can be argued that:
H2: Student behavioural engagement has a positive effect on student's academic performance.

2.3. Assessment

Assessment is the process of finalizing or attempting to finalize LMS graded activities, such as assignments, exams, or quizzes, by students (Bessadok et al., 2021). Assessment has become a crucial element of higher education and is integral to academic achievement (Lu & Cutumisu, 2022). Online assessment offers learners the opportunity for self- and feedback, enabling them to adjust and regulate their own learning (Lu & Cutumisu, 2022). Additionally, it can provide deeper insights into comprehending the subject matter (Yildirim & Gülbahar, 2022). Assessment activities demonstrate a high level of accuracy in predicting students' academic performance. Furthermore, comparing the predictive capabilities of different subsets of features, the assessment dataset showing a superior accuracy rate (Bessadok et al., 2021). Yildirim & Gülbahar (2022) research indicated that assessments in the form of pre-tests or midterms activities had a positive influence on final grades. Furthermore, Lu & Cutumisu (2022) and Zainuddin et al. (2019) confirmed that online assessments through LMS were effective in evaluating students' learning performance. Consequently, it can be hypothesized that:
H3: Assessment has a positive effect on student's academic performance.
Hypothesis are presented in the expected model in Figure 1.

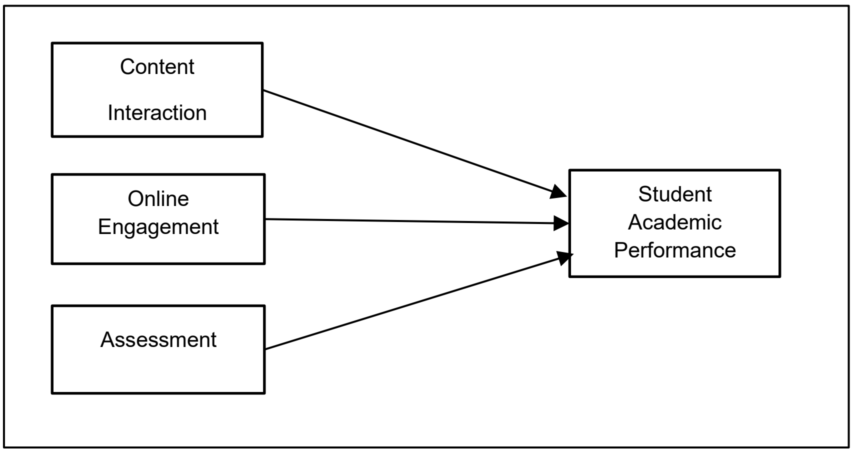

Figure 1 The expected research model.

3. Methodology

3.1. Data Set

The data analysed and processed in this study was obtained from the KU e-learning Centre, specifically the data recorded in the LMS across the various colleges within the university for the academic year 2022-2023. Consisting of 17 colleges, this study focuses specifically on the undergraduate programs offered by its 15 colleges with a population of 35,910 students (Kuwait University, 2018). Each student registered in the system has a record with several attributes, and the variety of these attributes depends on how the system is activated and utilized by the course instructor. To address the research question, a combination of activities and course grades is necessary. Consequently, only student records that meet these criteria are retained for subsequent analysis in this study. Specifically, 8,204 records from the first semester and 7,158 records from the second semester were obtained, resulting in a total of 15,362 student records.

During data preparation phase, the initial log files were processed to clean and prepare them for further analysis. This is essential because Moodle's extracted data sets often contain missing values, noisy data, and irrelevant or redundant information. To address this, the cleaning procedures occurred after the raw log files were initially imported into an Excel worksheet. Identifying data was selectively removed, and the dataset was anonymized by eliminating any personal information.

This study focused on three categories for analysis: content interaction, behavioural engagement, and assessment/evaluation as discussed in the literature review section. The student records kept represent these categories and reflect the digital activities of the students within the LMS, in line with the relevant literature discussed earlier. Detailed set of activities/action types used in the study for analysing students' digital behaviour presented in Table1.

Description	Activity type
Contents interaction	Course view, files downloaded on the course (Bessadok et al., 2021).
Behavioral engagement	Forum posts, Forum views (Lu & Cutumisu, 2022).
Assessment	Exam/ quizzes attempt, finalize exam/ quizzes, Assignment attempt, Finalize assignment (Bessadok et al., 2021; Lu & Cutumisu, 2022; Yildirim & Gülbahar, 2022) .
Student academic performance	Course grade.

Table 1 Student Activities Description

3.2. Analysis Methods

EDM and Learning Analytics are the two main fields of study that focus on analysing the data obtained from an LMS platform (Aldowah et al., 2019). Although there are some differences between the two concepts' purposes and scope, learning analytics is frequently related to EDM (Lemay et al., 2021). While EDM is about utilizing methods for the analysis of learning data to discover previously unknown trends, for example, cluster analysis, learning analytics is a data-driven decision-making strategy that deals with the explanation and contextualization of that data mostly to address predetermined questions for learning improvement (Tomasevic et al., 2020). Due to the exploratory nature of this study, where we examine the kind of digital activities students engage in within the LMS and whether these behaviours influence on academic performance, EDM appears relevant.

In order to analyse the digital behaviour of distinguished students and categorize them into clusters, the study utilized EDM clustering analysis, specifically K-means, a non-hierarchical method that aims to partition a given dataset into a predetermined number of clusters (k) by iteratively assigning data points to the nearest cluster centroid (Ramadan et al., 2020). One challenge associated with this type of analysis is the determination of the number of clusters by the researcher (Tomasevic et al., 2020). Various methods can be employed to determine the number of clusters, one of which involves using Hierarchical Cluster Analysis (Ramadan et al., 2020). In this study, IBM SPSS 29 was used to apply Hierarchical Cluster Analysis using Ward's method to create evenly sized clusters. The dendrogram was visually inspected to identify a reasonable number of clusters for the final clustering solution, resulting in the identification of two distinguished clusters (two unique digital behaviours).

Subsequently, K-means algorithms were applied using IBM SPSS 29, involving the performance of several clusters with different predetermined numbers until the optimal number of clusters was obtained, which also found to be two, confirming the previous analysis. This approach enabled the identification of distinct digital behaviours for each of the two clusters.
In order to examine the significance of the differences in means among the independent variables of the clusters, this research will extend the analysis by employing Multiple Linear Regression through IBM SPSS 29. This approach is especially suitable for forecasting student achievement using LMS log data (Kara & Yildirim, 2022). The Multiple Linear Regression will be utilized to assess the research hypothesis and ascertain whether the student digital behaviour determined by clustering, can predict student academic performance.

4. Conclusion

In conclusion, this study aims to uncover the hidden relationships between students' digital behaviour in the LMS and their academic performance at Kuwait University. By utilizing EDM algorithms and multiple regression, researchers plan to identify key patterns and predictors of success within the KU LMS environment. Such insights have the potential to inform personalized interventions. Also, understanding student preferences and engagement patterns within the LMS can guide the development of adaptive learning platforms and personalized instruction. Furthermore, the findings of this study would facilitate data-driven decision-making in resource allocation, curriculum development, redesigning learning systems, and prioritizing features.

References

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, *37*, 13–49. https://doi.org/10.1016/J.TELE.2019.01.007

Al-Fraihat, D., Joy, M., & Sinclair, J. (2020). Evaluating E-learning systems success: An empirical study. *Computers in Human Behavior*, *120*, 67–68. https://doi.org/10.1016/j.chb.2019.08.004

Bessadok, A., Abouzinadah, E., & Rabie, O. (2021). Exploring students digital activities and performances through their activities logged in learning management system using educational data mining approach. *Interactive Technology and Smart Education*. https://doi.org/10.1108/ITSE-08-2021-0148/FULL/HTML

Buckley, K. , Fairman, K. , Pogge, E., & Raney, E. (2021). Novel use of LMS data to predict online learning success in a pharmacy capstone course. *American Journal of Pharmaceutical Education.* https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&as_ylo=2019&inst=15450697844440806189&q=Novel+Use+of+LMS+Data+to+Predict+Online+Learning+Success+in+A+Pharmacy+Capstone+Course.&btnG=

El-Sayad, G., Hasliza, N., Md Saad, N., & Thurasamy, R. (2021). How higher education students in Egypt perceived online learning engagement and satisfaction during the COVID-19 pandemic. *Journal of Computers in Education*, *8*(4), 527–550. https://doi.org/10.1007/s40692-021-00191-y

Ghinea, G., Alkharang, M. M., & Ghinea, G. (2013). E-learning in Higher Educational Institutions in Kuwait: Experiences and Challenges. *International Journal of Advanced Computer Science and Applications*, *4*(4). https://doi.org/10.14569/IJACSA.2013.040401

Huang, A. Y. Q., Lu, O. H. T., Huang, J. C. H., Yin, C. J., & Yang, S. J. H. (2020). Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs. *Interactive Learning Environments*, *28*(2), 206–230. https://doi.org/10.1080/10494820.2019.1636086

Kara, M., & Yildirim, Z. (2022). Faculty Performance Improvement in Distance Education: Causes of the Performance Deficiencies (Part I). *Performance Improvement Quarterly*, *34*(4), 573–601. https://doi.org/10.1002/PIQ.21367

KU E-Learning centre. (2018). *Kuwait eLearning strategy*. http://ku.edu.kw/DLC/index.htm

Kuwait University. (2018). *About Kuwait University*. https://kuweb.ku.edu.kw/ku/index.htm

Leem, B. (2023). Impact of interactivity on learning outcome in online learning settings: Ordinal logit model. *Journal of Engineering Business Management*, *15*, 1–10. https://doi.org/10.1177/18479790231203107

Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence*, *2*, 100016. https://doi.org/10.1016/J.CAEAI.2021.100016

Lu, C., & Cutumisu, M. (2022). Online engagement and performance on formative assessments mediate the relationship between attendance and course performance. *International Journal of Educational Technology in Higher Education*, *19*(1). https://doi.org/10.1186/S41239-021-00307-5

Muir, T., Wang, I., Trimble, A., & Mainsbridge, C. (2022). Using interactive online pedagogical approaches to promote student engagement. *Education Sciences*, *12*(6), 415. https://www.mdpi.com/2227-7102/12/6/415

QS Quacquarelli Symonds. (2022). *QS World University Rankings*. https://www.qs.com/

Rajabalee, Y. B., & Santally, M. I. (2021). Learner satisfaction, engagement and performances in an online module: Implications for institutional e-learning policy. *Education and Information Technologies*, *26*(3), 2623–2656. https://doi.org/10.1007/S10639-020-10375-1

Ramadan, R. A., Alhaisoni, M. M., & Khedr, A. Y. (2020). Multiobjective clustering algorithm for complex data in learning management systems. *Complex Adaptive Systems Modeling*, *8*(1). https://doi.org/10.1186/S40294-020-00071-9

Riestra-González, M., Del Puerto Paule-Ruíz, M., Ortin, F., & Del Puerto Paule-Ruiz, M. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers & Education, 163,* 104108–104128. https://doi.org/10.1016/j.compedu.2020.104108

RUR Rankings Russian Federation Agency. (2021). *Round University Ranking RUR*. Thomson Reuters. https://roundranking.com/

Salas-Pilco, S. Z., Yang, Y., & Zhang, Z. (2022). Student engagement in online learning in Latin American higher education during the COVID-19 pandemic: A systematic review. *British Journal of Educational Technology*, *53*(3), 593–619. https://doi.org/10.1111/BJET.13190

Times Higher Education. (2022). *Times Higher Education's global portfolio of university rankings*. https://www.timeshighereducation.com/

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, *143*, 103676. https://doi.org/10.1016/J.COMPEDU.2019.103676

Vlachopoulos, D., & Makri, A. (2019). Online communication and interaction in distance higher education: A framework study of good practice. *International Review of Education*, *65*(4), 605–632. https://doi.org/10.1007/S11159-019-09792-3

Yildirim, D., & Gülbahar, Y. (2022). Implementation of Learning Analytics Indicators for Increasing Learners' Final Performance. *Technology, Knowledge and Learning*, *27*(2), 479–504. https://doi.org/10.1007/S10758-021-09583-6

Zainuddin, Z., Shujahat, M., Chu, S. K. W., Haruna, H., & Farida, R. (2019). The effects of gamified flipped instruction on learner performance and need satisfaction: A study in a low-tech setting. *Information and Learning Science*, *120*(11–12), 789–802. https://doi.org/10.1108/ILS-07-2019-0067/FULL/HTML