Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Exploring the Behavior Patterns of Students Accessing Online Learning Material in Online Course: A Case Study at Hung Vuong University

Pham-Duc THO[a*] & Chih-Hung LAI[b]
[a]*Department of Software Engineering, Hung Vuong University, Vietnam*
[b]*Department of Computer Science and Information Engineering, National Dong Hwa University, Taiwan*
*thopham@hvu.edu.vn

Abstract: Understanding students' behavior in online courses may provide teachers with useful information to improve their educational design and provide insights for content and instructional designers to develop personalized learning support. This research uses cluster analysis to explore learners' interaction with online learning materials behavior in an online course at Hung Vuong University, Vietnam and identified three clusters (Less-engaged students, Moderately-engaged students, and Highly-engaged students) which evince different behavior patterns with regards to the time spent interacting with various resources. Based on the findings, several suggestions are also proposed for future research.

Keywords: students behavior, online courses, cluster analysis, behavior patterns

1. Introduction

In Vietnamese context, online learning or blended learning has been introduced for years (Maheshwari, 2021; Dinh & Nguyen, 2020; Van Anh & Nguyen, 2020); however, we still lack of the information on how students performs or interactive in the online course. It is well recognized that the interaction with online learning materials is one of the most commonly performed online learning activities (Li & Tsai, 2017). Teachers and students typically publish and create different kinds of online resources for learning, and such materials give different learning advantages. For example, lecture slides give an outline of teaching contents for students and facilitate students' note-taking (Worthington & Levasseur, 2015), students may review challenging concepts and prepare for examinations through video lectures (Hong, Pi, & Yang, 2018), while peers' assignments and messages posted in discussion forums are essential resources for self-reflection (Sun, Lin, Wu, Zhou, & Luo, 2018; Zheng, Cui, Li, & Huang, 2018). Students may demonstrate different levels of engagement and patterns of behavior when interacting with online learning materials for different purposes and based on different preferences (Cerezo, Sánchez Santillán, Paule Ruiz, & Núñez, 2016; Li & Tsai, 2017); these levels of engagement and patterns of behavior may, in turn, affect their learning performance (Cerezo et al., 2016; Lust, Elen, & Clarebout, 2013a).

Consequently, understanding how students interact with different types of learning materials and how their behavior in interacting with these materials affects their learning performance may provide teachers with useful information to improve their educational design and provide insights for content and instructional designers to develop personalized learning support. However, only using statistical methods is not enough to explore students' interaction with online learning material behavior. Cluster analysis (e.g., the K-means method) can be used to investigate the behavior cluster patterns of a group regarding various indicators (such as the frequency of a particular discussion behavior) (Huei Tse Hou, 2011). Thus, the use of clustering techniques on these behavior sets enables the potential cluster patterns of learners' different behaviors to be explored when interacting with online learning materials (Bakhshinategh, Zaiane, ElAtia, & Ipperciel, 2018; Huei Tse Hou, 2015; Huei Tse Hou & Li, 2014; H. T. Hou & Wu, 2011; Li & Tsai, 2017).

Hence, our study is focused on providing more in-depth perspectives and insightful information derived from students' interaction with online learning material behavior, for instance: their interaction with online learning material behavioral patterns that occurred during their learning process. Our research question is proposed as follow: What are the students' clusters of interacting with online learning material in an online class?

2. Literature Review

Analyzing students' behavior in interacting with online learning materials helped in identifying learners with poor performance (Li & Tsai, 2017; Zhang, Zou, Miao, Zhang, Hwang, & Zhu, 2019), and hence in providing improvement suggestions (Lerche & Kiel, 2018; Zhang, Zhang, Zou, & Huang, 2018; Zhang et al., 2019). Researchers have also pointed out that correlation analysis can help the instructor to determine the relevance between students' learning behavior and performance (Zhang et al., 2019), as well as assist in decision-making and improving teaching and learning processes (Zhang et al., 2019).

In addition to using statistical methods, several studies have used cluster analysis to classify students into distinct groups (Huei Tse Hou, 2015; Li & Tsai, 2017) and to investigate their learning performance (Perera, Kay, Koprinska, Yacef, & Zaïane, 2009). In recent research, Li and Tsai (2017) concluded that different behavior patterns were associated with students' motivation and learning performance. Cluster analysis (e.g., the K-means method) can be used to investigate the behavior cluster patterns of a group regarding various indicators (such as the frequency of an individual discussion behavior) (Huei Tse Hou, 2011). By applying cluster analysis, the potential cluster patterns of learners' various behaviors can be explored (Huei Tse Hou, 2012, 2015; Huei Tse Hou & Li, 2014; Li & Tsai, 2017) (for example, by analyzing the overall learning process of a group of students, questions can be raised: How many potential clusters of learners with similar behavioral traits are being formed? What are the characteristics of each cluster?). In other words, it provides an opportunity to discover meaningful data from learners individually (Perera et al., 2009).

This study applied the most frequently performed interacting learning material activities, as stated in the previous research (Cerezo et al., 2016; Li & Tsai, 2017; Su, Ding, & Lai, 2017) with a Learning Management System (LMS) applied. Therefore, seven activities were identified and selected: Page Hits on Questions, number of Answers Posted, number of Answers Revised, Page Hits on Lecture Slides, number of Comment Posted, number of Discussion Posted, and number of Discussion Edited.

3. Method

3.1 Research Design and Participants

This study aimed to examine the effects of online learning behavior on online learning regarding students' academic performance in a class with the use of an LMS. The participants were 38 university students (33 males and five females) enrolled in a course named INT326 English for Computer Science. The course was compulsory for all the students, and after passing the final examination, they were awarded three credits counting towards their graduation.

3.2 Experimental Procedure

The class took place on a weekly basis for the duration of 15 weeks; however our experiment only took 8 weeks of the whole class duration. Class time was the main point of interaction between teachers and participants. Each lecture took three hours and the course is purely online during the COVID-19 pandemic. During the first week of the experiment (week 1 of the semester), an introductory class was held in order to instruct students on how to interact with an LMS system named HVU LMS and access the course-related resources. Students were familiarized with the environment, compulsory class components, and evaluation processes.

Subsequently, from week 2 to week 8 of the experiment, students were taught 3 hours a week using the proposed online learning system as an environment for submitting assignments. The students were encouraged to use the learning system after class.

3.3 HVU Learning Management System

HVU LMS is an online learning environment, a Moodle-based eLearning platform developed at Hung Vuong University. In this system, students are able to generate questions and discuss with each other by asking, answering questions, and commenting through the provided functions. Instructors are also able to generate questions, share learning resources, and develop the effectiveness of class management. HVU LMS main interface offers multiple functions that can be used to promote online learning can be seen in Figure 1.

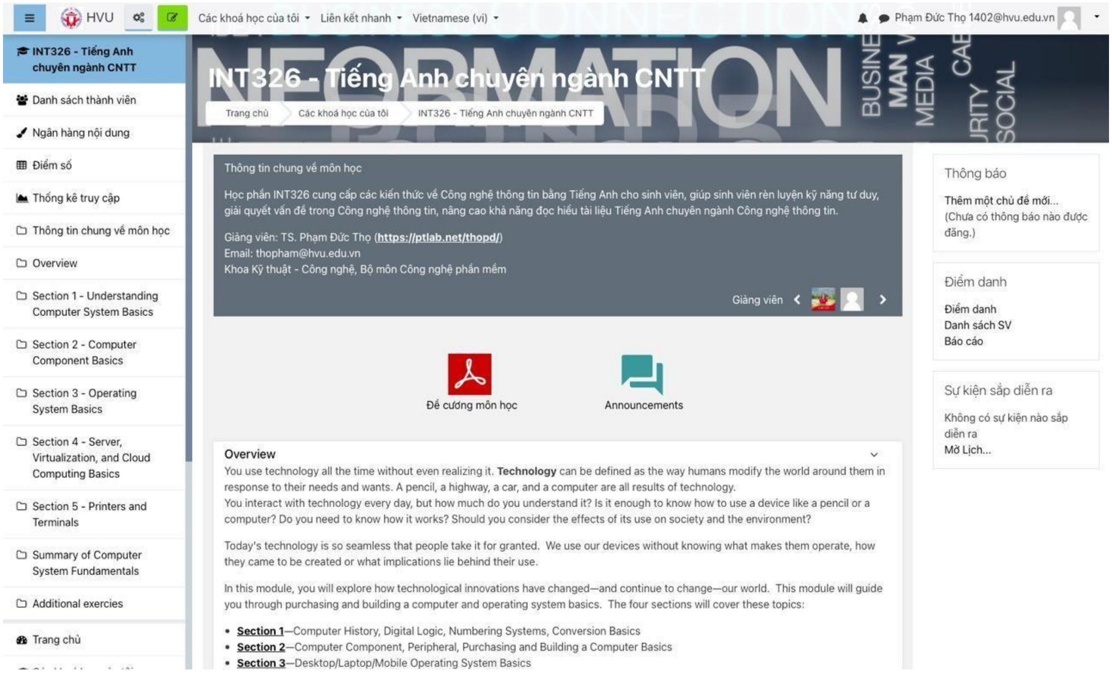

Figure 1. HVU LMS User Interface

3.4 Data Collection and Analysis

In this study, analyzed data were in the forms of log files, which contain the participants' interactions and all information needed on HVU LMS from a database powered by MySQL. The researcher collected data in a total of eight weeks. The number of questionings, comment, revision, and access to learning materials was calculated by simple SQL queries based on unique user IDs. The data were gathered from an HVU LMS database via phpMyAdmin; luckily, missing values were not found in the dataset. Afterward, they were exported into a CSV file for further transformation.

After completing the data cleaning process, the data were then carefully transformed into a sav file for SPSS analysis. Importantly, the student's behavior was extracted from log files individually by using SQL queries based on unique user IDs. To differentiate the participants into groups according to the similarities of their interaction with learning materials behavior (e.g., questioning, commenting, assignment completion, revision, and access to learning materials) that occurred during their computer programming learning progress on the proposed online learning system (i.e., HVU LMS), we extracted a total of seven variables for the analysis as listed in Table 1. A complete enumeration of these variables, along with their basic statistical properties, can be found in Table 2. All of the time-related variables are measured in the total number of occurrences. Despite the small size of our test group, Box Plots of our seven crucial variables still revealed numerous cases that were very distant from the IRQ region, as illustrated in Figure 3. Since these deviations could negatively project onto the clustering process, we decided to transform these variables to a scale of 1-3 in order to reduce the bias in the cluster analysis, following the methodology of Li and Tsai (2017). The 33.33% lowest, intermediate, and highest access times were allocated a value of 1, 2, and 3, respectively, indicating low, moderate, and high access times. In the following, we will refer to the transformed variables as $t^T_{QV}, t^T_A, t^T_R, t^T_L, t^T_C, t^T_{QP}, t^T_{QE}$. Further, we deployed k-means clustering among various subsets of variables as dimensions of the Euclidean space to search for learning behavior patterns. The number of clusters to consider was decided based on the size of the underlying dataset and the dendrogram resulting from its Hierarchical Agglomerative Clustering (HAC). We proceeded in our analysis with clusters that appeared to be consistent, balanced, and mutually distant.

Table 1. *Variables Extracted from HVU LMS*

#	Variable	Variable Description
1	t_{QV}	Page Hits on Questions
2	t_A	Answers Posted
3	t_R	Answers Revised
4	t_L	Page Hits on Lecture Slides
5	t_C	Comment Posted
6	t_{QP}	Discussion Posted
7	t_{QE}	Discussion Edited

Table 2. *Mean and Standard Deviation of Variables Extracted from HVU LMS*

#	Variable	Variable Description	Mean	SD
1	t_{QV}	Page Hits on Questions	619.42	607.41
2	t_A	Answers Posted	48.53	17.44
3	t_R	Answers Revised	34.79	49.37
4	t_L	Page Hits on Lecture Slides	63.76	35.19
5	t_C	Comment Posted	59.32	166.9
6	t_{QP}	Discussion Posted	4.08	2.78
7	t_{QE}	Discussion Edited	6.79	14.58

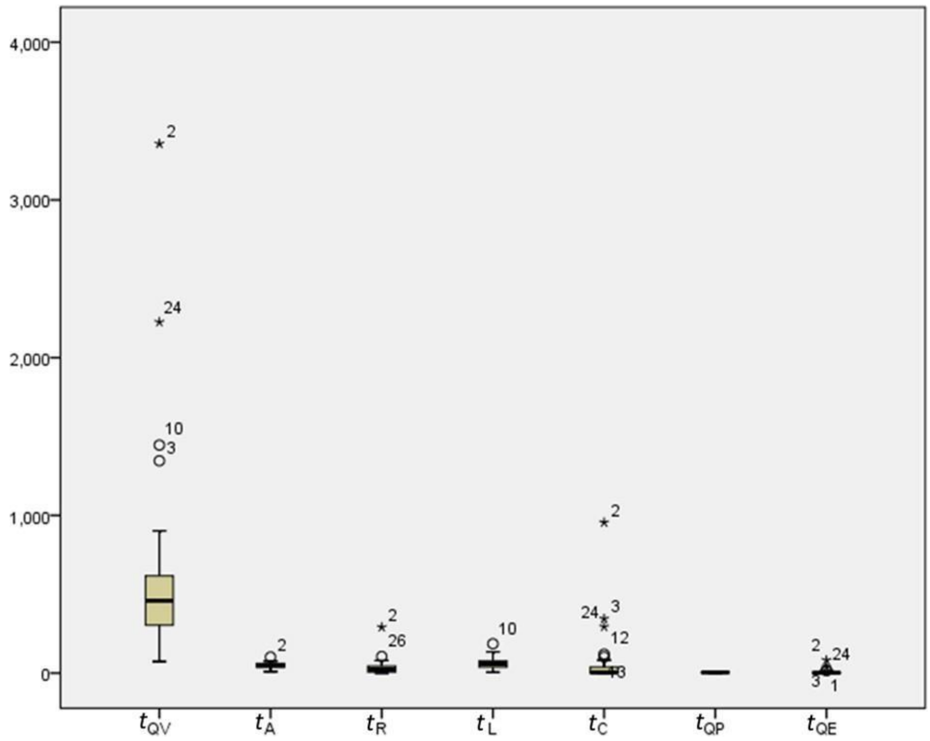

Figure 2: Boxplot of t_{QV}, t_A , t_R , t_L , t_C , t_{QP}, t_{QE}

After identifying the participants' similarities and clustering them into groups, the significant differences, in terms of their learning performance, among the generated clusters must be revealed statistically. Traditionally, a parametric analysis, such as one-way ANOVA, can be used to analyze data if the assumptions are met. The assumptions are as follows:

- ★ Random independent samples
- ★ Interval or ratio level of measurement
- ★ Normal distribution
- ★ No outliers
- ★ Homogeneity of Variance
- ★ A good amount of sample size

However, the data used in this experiment had not met the assumptions mentioned above. In this case, a non-parametric test can be used to analyze the data (Li & Tsai, 2017). Even though non- parametric tests do not have statistical power compared to parametric ones, they are more conservative. Consequently, this study implemented a Kruskal-Wallis test as the primary data analysis method. Furthermore, if a Kruskal-Wallis test demonstrates at least one significant difference among the clusters, a Mann-Whitney test will be conducted as a post hoc test (Li & Tsai, 2017; López, Valenzuela, Nussbaum, & Tsai, 2015). It should be noted that the significance level was set at $p = .05$.

4. Results and Discussions

To classify the students with similar interaction patterns into a homogeneous group, k-means cluster analysis was performed on the seven transformed variables t^T_{QV} , t^T_A, t^T_R, t^T_L, t^T_C, t^T_{QP} , t^T_{QE} . As shown in Table 3, three clusters were identified. These clusters evince differences in students' learning behavior patterns, and therefore we assigned them slightly suggestive names:

(1) Less-engaged students
(2) Moderately-engaged students
(3) Highly-engaged students

As shown in Table 3, from the variance on the average frequency of the seven main behaviors – View Question, Answer, Answer Revision, Learning, Comment, Generate Discussion, and Discussion Edit (t^T_{QV}, t^T_A, t^T_R, t^T_L, t^T_C, t^T_{QP}, t^T_{QE}), as exhibited by the three clusters of students, we learned that students' interaction with learning materials behavior patterns in the online class was distinctively different. The three clusters comprise 16,14, and 8 people, respectively, accounting for 42.11%, 36.84%, and 21.05% of the total students.

Table 3. *Cluster analysis of Interacting Online Learning Material behavior*

Indicators of cluster analysis	Clusters			F
	Less-engaged students (N=16, 42.11%)	Moderately-engaged students (N=14, 36.84%)	Highly-engaged students (N=8, 21.05%)	
t_{QVT}	1.19	2.36	2.88	49.461***
t_{AT}	1.25	2.43	3.00	52.991***
t_{RT}	1.88	1.64	2.75	6.169**
t_{LT}	1.19	2.43	2.63	29.02***
t_{CT}	1.44	2.29	2.75	13.074***
t_{QPT}	1.19	2.21	2.88	39.003***
t_{QET}	1.50	1.79	3.00	15.122***

$**p < 0.01$, $***p < 0.001$

More than 20% of the students are centered in the Highly-engaged students Cluster (N = 8, 21.05%), and the average learning behavior frequency of their behaviors – View Question, Answer, Answer Revision, Learning, Comment, Generate Discussion, and Discussion Edit (t^T_{QV}, t^T_A, t^T_R, t^T_L, t^T_C, t^T_{QP}, t^T_{QE}) – was higher than that of the other two clusters. This suggests that 21.05% of the students learning this course exhibited behaviors with more action than the other two clusters. On the other hand, it is to say that more than 40% of the students learning this course exhibited behaviors with significant inactively than the other two clusters.

After classifying the students into homogeneous groups based on similarities in their course material viewing patterns, we performed the Kruskal–Wallis test in order to compare Less-engaged students, Moderately-engaged students, and Highly-engaged students with regards to the set of collected variables. The test outcome is depicted in Table 4. We observed a statistically significant difference in all the aspects measured.

Our result is aligned with the previous study conducted by Li and Tsai (2017), and provide evidence that Less-engaged students spent significantly less effort on most activities, namely t_{QV}, t_A , t_L , t_C , t_{QP}, when compared to Moderately-engaged students and Highly-engaged students. However, we cannot conclude the difference between Less-engaged students and Moderately-engaged students in the revising activities t_R, t_{QE}. On the other hand, our results identified a Highly-engaged students cluster, which consists of students with significantly more effort measured in all kinds of learning materials when compared to both the Less-engaged students and the Moderately-engaged students. Moreover, although we could not establish any relationship with regards to the average time spent on Learning and Commenting t_L, t_C between the Moderately-engaged students and the Highly- engaged students, our results reveal that students from both the Highly-engaged and the Moderately- engaged clusters spent a significantly longer time on average on Learning and Commenting access than the Less-engaged students.

Table 4: *Analysis of Online Learning Behavior*

Var	Less-engaged students (1)		Moderately-engaged students (2)		Highly-engaged students (3)		Kruskal–Wallis Test	Mann-Whitney U Test
	Mean	SD	Mean	SD	Mean	SD	p	
t_{QV}	278.81	91.09	565.79	124.42	1394.50	972.70	0.000***	2<3 2>1 3>1
t_A	33.31	8.94	53.64	8.21	70.00	14.22	0.000***	2<3 2>1 3>1
t_R	22.37	22.01	23.50	26.15	79.38	87.73	0.011*	2<3 3>1
t_L	37.19	17.98	74.21	16.55	98.63	46.08	0.000***	2>1 3>1
t_C	3.44	7.14	35.00	43.59	213.63	328.74	0.000***	2>1 3>1
t_{QP}	1.62	1.02	4.86	1.66	7.62	2.07	0.000***	2<3 2>1 3>1
t_{QE}	1.56	2.94	2.50	3.11	24.75	24.87	0.000***	2<3 3>1

*p < 0.05, **p < 0.01, ***p < 0.001

5. Conclusion and Future Works

5.1 Conclusion

In this research, we explored and revealed students' interaction patterns with regard to online resources based on students' different identified groups of interaction with online learning material behavior. Based on the information gathered, we attempted to answer our research question by identified three clusters (Less-engaged students, Moderately-engaged students, and Highly-engaged students) which evince different behavior patterns with regards to the time spent interacting with various resources, i.e.
t_{QV}, t_A, t_R, t_L, t_C, t_{QP}, t_{QE}. We detected one cluster of students (Highly-engaged students) that dominated the other two (Less-engaged students, Moderately-engaged students) in all leading variables. This result aligned with a previous study by Li and Tsai (2017), who identified a single cluster on the lower-access end ("low-use-students") and two clusters on the higher end ("slide-intensive-students" and "consistent-use-students"). However, we cannot conclude the difference between the Less-engaged students and Moderately-engaged students in the revising activities t_R, t_{QE}. Moreover, although we could not establish any relationship with regards to the average time spent on Learning and Commenting
t_L, t_C by the Moderately-engaged students and the Highly-engaged students, our results indicate that students from both the Highly-engaged and the Moderately-engaged clusters spent significantly more time on average Learning and Commenting than the Less-engaged students.

5.2 Future works

Based on the findings, this study provides several suggestions for future research:

Future works can deep investigate the content analysis of comments and discussions to students' engagement and students' behavior. It is also interesting to investigate the effect of the automated reply feature of HVU LMS on students' engagement and students' behavior.

For the future development of HVU LMS, we suggest embedding the automatic analysis and instant feedback mechanisms along with early-detection behavior groups into the learning system as a future trend (Huei Tse Hou, Chang, & Sung, 2010). Integrating real-time computing with early detection sequential patterns of learning behavior in HVU LMS may be enhanced by developing mechanisms that provide real-time learning feedback as scaffolding. This approach not only promptly provides teachers with diagnoses of student misconceptions or bottlenecks in learning as important reference information but also offers corresponding real-time guidance regarding the behavior patterns of specific incorrect manipulations. Such an automatic feedback design may optimize the learning process, allowing continuous adjustments to problem-solving strategies and helping teachers identify a variety of misconceptions and incorrect manipulations that the students often display in online courses.

References

Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies, 23*(1), 537-553. doi:10.1007/ s10639-017-9616-z

Cerezo, R., Sánchez Santillán, M., Paule Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education, 96*, 42- 54. doi:10.1016/j.compedu.2016.02.006

Dinh, L. P., & Nguyen, T. T. (2020). Pandemic, social distancing, and social work education: Students' satisfaction with online education in Vietnam. *Social Work Education, 39*(8), 1074-1083.

Heffner, M., & Cohen, S. H. (2005). Evaluating Student Use of Web-Based Course Material. *Journal of Instructional Psychology, 32*(1), 74-81.

Hong, J., Pi, Z., & Yang, J. (2018). Learning declarative and procedural knowledge via video lectures: cognitive load and learning effectiveness. *Innovations in Education and Teaching International, 55*(1), 74-81. doi:10.1080/14703297.2016.1237371

Hou, H. T. (2011). A case study of online instructional collaborative discussion activities for problem-solving using situated scenarios: An examination of content and behavior cluster analysis. *Computers & Education, 56*(3), 712-719. doi:10.1016/j.compedu.2010.10.013

Hou, H. T. (2012). Exploring the behavioral patterns of learners in an educational massively multiple online role- playing game (MMORPG). *Computers & Education, 58*(4), 1225-1233.

Hou, H. T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. *Computers in Human Behavior, 48*, 424-435.

Hou, H. T., Chang, K. E., & Sung, Y. T. (2010). Applying lag sequential analysis to detect visual behavioural patterns of online learning activities. *British Journal of Educational Technology, 41*(2), E25-E27. doi:10.1111/j.1467-8535.2009.00935.x

Hou, H. T., & Li, M. C. (2014). Evaluating multiple aspects of a digital educational problem-solving-based adventure game. *Computers in Human Behavior, 30*, 29-38. doi:10.1016/j.chb.2013.07.052

Hou, H. T., & Wu, S. Y. (2011). Analyzing the social knowledge construction behavioral patterns of an online synchronous collaborative discussion instructional activity using an instant messaging tool: A case study. *Computers & Education, 57*(2), 1459-1468. doi:10.1016/j.compedu.2011.02.012

Lerche, T., & Kiel, E. (2018). Predicting student achievement in learning management systems by log data analysis. *Computers in Human Behavior, 89*, 367-372. doi:10.1016/j.chb.2018.06.015

Li, L. Y., & Tsai, C. C. (2017). Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance. *Computers & Education, 114*, 286-297. doi:10.1016/j.compedu.2017.07.007

López, X., Valenzuela, J., Nussbaum, M., & Tsai, C.-C. (2015). Some recommendations for the reporting of quantitative studies. *Computers & Education, 91*, 106-110. doi:10.1016/j.compedu.2015.09.010

Lust, G., Elen, J., & Clarebout, G. (2013a). Regulation of tool-use within a blended course: Student differences and performance effects. *Computers & Education, 60*(1), 385-395.

Lust, G., Elen, J., & Clarebout, G. (2013b). Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior, 29*(5), 2013-2021. doi:10.1016/j.chb.2013.03.014

Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In Search of user profiles. *Computers & Education, 57*(3), 2135-2144.

Maheshwari, G. (2021). Factors affecting students' intentions to undertake online learning: an empirical study in Vietnam. *Education and Information Technologies, 26*(6), 6629-6649.

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and Sequential Pattern Mining of
Online Collaborative Learning Data. *IEEE Transactions on Knowledge and Data Engineering, 21*(6), 759-772. doi:10.1109/TKDE.2008.138

Su, Y. S., Ding, T. J., & Lai, C. F. (2017). Analysis of Students Engagement and Learning Performance in a Social Community Supported Computer Programming Course. *Eurasia Journal of Mathematics, Science and Technology Education, 13*, 6189-6201.

Sun, Z., Lin, C. H., Wu, M. H., Zhou, J. S., & Luo, L. M. (2018). A tale of two communication tools: Discussion- forum and mobile instant-messaging apps in collaborative learning. *British Journal of Educational Technology, 49*(2), 248-261. doi:10.1111/bjet.12571

Worthington, D. L., & Levasseur, D. G. (2015). To provide or not to provide course PowerPoint slides? The impact of instructor-provided slides upon student attendance and performance. *Computers & Education, 85*, 14-22. doi:10.1016/j.compedu.2015.02.002

Zhang, J. H., Zhang, Y. X., Zou, Q., & Huang, S. (2018). What Learning Analytics Tells Us: Group Behavior Analysis and Individual Learning Diagnosis based on Long-Term and Large-Scale Data. *Educational Technology & Society, 21*(2), 245-258.

Zhang, J. H., Zou, L. C., Miao, J. J., Zhang, Y. X., Hwang, G. J., & Zhu, Y. (2019). An individualized intervention approach to improving university students' learning performance and interactive behaviors in a blended learning environment. *Interactive Learning Environments*, 1-15. doi:10.1080/10494820.2019.1636078

Zheng, L. Q., Cui, P. P., Li, X., & Huang, R. H. (2018). Synchronous discussion between assessors and assesses in web-based peer assessment: impact on writing performance, feedback quality, meta-cognitive awareness and self-efficacy. *Assessment & Evaluation in Higher Education*, *43*(3), 500-514.