Effect of Behavior Patterns of Accessing Learning Materials on Learning Performance in Student-generated Question Activities

Pham Duc Tho, Chih-Hung Lai, Ondrej Navratil

Department of Computer Science and Information Engineering, National Dong Hwa University, Taiwan
810321006@gms.ndhu.edu.tw, laich@gms.ndhu.edu.tw, navratil@gms.ndhu.edu.tw

Abstract

This study explores the effects of student behavior patterns on students' learning performance in a student-generated question experiment. Thirty-three engineering students from a C programming course were recruited as participants. For data collection, student-generated questions, exercises, a final examination, and system logs recorded by an online student-generated questions system were analyzed. The results revealed several significant findings. Firstly, we found that the students accessed the questions generated by the classmate more frequently and for longer than other materials (i.e., exercises and lecture slides). Our results also reveal great diversity among student viewing activities; we therefore partitioned them into three behavior clusters: "Highly-engaged students" who dominate other clusters in the use of all kinds of materials, "Moderately-engaged students" who spent more time on the lecture slides and SGQ, and "Less-engaged students" who seldom used learning materials but accessed exercises more frequently than the "Moderately-engaged students" and also had the highest Test Anxiety. We also observed that motivation and learning performance are strongly associated with behavior patterns. We discuss possible explanations of our findings and propose suggestions for future research.

Keywords: Student-generated questions, Learning behavior pattern, Learning management system, Learning analytics

1 Introduction

Student-generated questions [1-3], also referred to as student-constructed questions [4] or problem posing [5] in the literature, form an essential strategy that has been observed to enhance comprehension of learned content [6], to encourage and monitor awareness [2], and to promote motivation [7-9] and learning performance [1, 9]. In the following, we will refer to this method as Student Generated Questions (SGQ).

While studies have provided evidence that SGQ can promote and facilitate the personal growth of learners [2], few studies have investigated how the students access SGQ with comparison to other materials (i.e., lecture slides, or exercises) during classes [9]. Students may demonstrate various engagement levels and behavior patterns when accessing online learning materials, and do so for different purposes and based on different preferences [10]; these engagement levels and behavior patterns may affect their learning performance in turn [11]. In order to efficiently investigate and define the impact of specific learning strategies, it is essential to choose an appropriate framework. In recent years, following the expansion of online learning environments, a new discipline called Learning Analytics (LA) has emerged [12-13]. LA utilizes our newly-gained ability to capture and process information during the learning process. LA has the potential to deepen our knowledge about learning behavior and improve learning outcomes [12-14]. Therefore, this study aims to track and examine how students access SGQ and other online learning materials, and how their behavior affects their motivation and learning performance in an online environment.

2 Literature Review

2.1 Student Generated Questions

During recent years, an increasing number of researchers have advocated the use of SGQ for learning [15-16]. These researches also indicated the benefits of integrating SQG into education, and recognized numerous aspects of the positive effects on students [2, 17-20].

SGQ has been successfully deployed [1-2, 21-23] and observed to positively affect learning outcomes such as cognitive and affective growth [24], efforts to further enhance learning effects [21, 25-27] and personal growth [1, 9]. Impacts of SGQ have been researched both independently [1, 21-22] and in a combination with various pedagogical designs [6-7, 16, 25, 28]. However, there are only few studies have inspected the students usage of SGQ with comparison to other learning materials (i.e., lecture slides, or

*Corresponding Author: Chih-Hung Lai; E-mail: laich@gms.ndhu.edu.tw
DOI: 10.3966/160792642020012101013

exercises) during classes [9].

2.2 Students' Learning Behaviors

With the growing technology [29-35], some scholars have analyzed students' behavior in higher education, such as reading of materials [10, 13-14, 36], interaction with classmates [13, 37-39], submission of homework and examination attempts in online courses [40]. This helped in identifying learners with poor performance [41], and hence provided improvement suggestions [42]. Another researcher pointed out that correlation analysis can help the instructor to determine the relevance between students' learning behavior and performance [14], as well as assist in decision-making and improving teaching and learning processes [14, 43]. Some studies have deployed cluster analysis in order to partition students into distinct groups [44] and to investigate their learning performance [10, 44]. In their study, Lust et al. [44] managed to isolate a cluster of intensive participants that accessed Web lectures more frequently and intensively in comparison with incoherent-use and no-use participants. In another research, Li and Tsai [10] concluded that different behavior patterns were associated with students' motivation and learning performance.

Despite the importance of investigating students' learning behavior, only few studies have observed students' learning behavior with the use of SGQ approach. Therefore, the current study aimed to reveal any relationship between students' behavior pattern and students' motivation and learning performance during SGQ enabled class.

2.3 Student Motivation and Online Engagement

During an online learning process, students control their learning style as they access Online Learning System. This poses a challenge to the efficiency of an individual's self-regulation. One of the key factors affecting this efficiency is motivation [16, 45]. Motivation can significantly impact study strategies and thus also influence online learning behaviors [46-47] and outcomes. Moreover, various studies have shown evidence that motivation is mutually connected with students' online engagement. Li and Tsai [10] examined relations between viewing behavior, motivation, and learning performance. They conclude that there is an influence between students' behavior patterns and their motivation, in both directions. These results are in alignment with observations by Yi and Hwang [48] and Barba et al. [46]. The relationship between student motivation, engagement, and learning performance was also investigated by Giesbers, et al. [49]. Their findings bear evidence that the intrinsic motivation scores of the students participating in four web-videoconferences were significantly higher than those of students who only participated once, twice, or not at all. Therefore, different motivational factors seem to have different effects on the way participants behave within online classes, and this effect can later project onto their learning performance.

2.4 Research Questions

Significant evidence emerging from the previous research pointed out the advantages of SGQ. Further, a number of studies has been carried out in order to understand and evaluate the efficiency of various student learning behavior profiles; however, likely due to the lack of appropriate software support, only a fraction of them measured non-discrete events. In this article, we introduce an experiment conducted on a custom implemented Online Learning System called Peer-Interaction Programming Learning System (PIPLS). We performed analysis in accordance with that of Li and Tsai [10] in order to establish access patterns with regards to SGQ resources. Further, we examined these clusters against other variables that were measured with the aim of extracting and discussing their influence on the learning outcome. Deploying this approach, we aim to find clues for the following set of questions:

(1) Which materials do students prefer in the learning process in which SGQ are applied?

(2) Which study material access behavior patterns can be recognized in a class with SGQ?

(3) How are these behavior patterns related to learning performance in a class with SGQ?

(4) How are these behavior patterns related to student motivation in a class with SGQ?

3 Methods

3.1 Participants and Research Context

Our test group consists of 33 students (4 females, 29 males), who enrolled in a C-language course Introduction to Computer Programming. All of the participants are undergraduate students with a major in engineering; however, none of them are Computer Science students. Therefore, for most of the attendees, this was the first purely programming class they had ever selected throughout their academic curriculum. Although the course itself was classroom-based, it included compulsory online components. The classroom was equipped with a dedicated PC for every student attending the course. The course was elective (i.e., not compulsory) for all of the students, and after passing the final examination, they were awarded with 3 credits counting towards their graduation.

Since our research requirements surpass standard off-the-shelf software capabilities, major customization was imperative. In order to conduct the experiment and collect the core data, a Peer-Interaction Programming Learning System (PIPLS) was built based on an open-source project developed by Greenspan and Contributors [50]. PIPLS is web-based, with the server side running on PHP and MySQL. As for UI, PIPLS provides standard functionality, such as management of course information, participants, materials, exercises and scores on the teacher side, and access to all of these on the student side. The GUI does not have any ambition to deviate from standards and therefore requires close to no training. In addition, an SGQ module was implemented, allowing students to post and answer questions. Answers could be later evaluated and optionally counted towards the final grade. A major upgrade in PIPLS is the ability to track and record various interaction events. In addition to the conventional "page hits," the system also keeps track of a range of mouse/window events using asynchronous server requests, thus determining the user's focus and estimating the time spent on individual activities, rather than just their frequency of occurrence. Besides research purposes, PIPLS also displays statistical data about user interactions to the teacher, which in return helps them to adjust the study material composition accordingly.

3.2 Procedure

The class took place on a weekly basis for the duration of 11 weeks. Class time was the main point of interaction between teachers and participants. Each lecture took 3 hours. In the first week of the semester, an introduction class was held in order to instruct students in how to interact with PIPLS and access course-related resources. Students were familiarized with the environment, compulsory class components and evaluation processes. For the rest of the course, the lecture content was scheduled as follows:

(1) Chapter topics were introduced, using commented slides.

(2) Sample SGQ were presented to the students, involving chapter content and providing a sample of the result expected.

(3) Exercises and SGQ tasks were assigned to the students.

(4) Students were instructed to complete the assignments either in the remaining time of the lecture (1 hour), or anytime within the timespan of the following 3 days. Students were randomly assigned two SGQ every week by the system. Students were also provided with slides and exercises via PIPLS. Lecture slides were published one week ahead for student's convenience. Each week, teachers published three exercises related to the chapter discussed in the respective lecture. The exercises were similar to an examination question, but the scores did not count towards the final grade.

3.3 Measurement Instrument

The database for our analysis was assembled throughout the semester. Students could access three kinds of online resources via PIPLS, namely (1) lecture slides, (2) exercises, (3) SGQ (ask & answer). Detailed information regarding the quantity of each resource available can be found in Table 1. These statistics represent the total amount of materials published at the end of the semester; however, they were uploaded and created gradually throughout the course timespan. At the end of the semester, students' performance was evaluated in a final examination. This examination was designed by the course professor and two teaching assistants (Master students majoring in Computer Science). The validity of the questions was later evaluated by two expert committees with more than ten years of teaching C Programming at university level, who reviewed the content and scope of the questions as well as their proportional relevance. The final examination contained a total point sum of 100. Any score above 50 points was considered a pass. Table 2 provides a detailed description of the content and evaluation.

Table 1. Materials available in PIPLS and respective events

Resource Type	Resource Quantity	Supporting Activity
Lecture Slides	6 PDF documents / 108 pages	Comments
Exercises	33	Comments, Answers
Proposed SGQ	157	Comments, Answers, Votes

Table 2. Final examination profile and reliability

Assignment Type	Evaluation Method	Quantity	Points per item	Total Points
Coding Question	Automated test case	3	15	45
Open question	Manual correction	2	15	30
Multiple-choice	Correct/incorrect	4	5	20
True/False Question	Correct/incorrect	1	5	5
Total		10		100

The level of students' motivation was measured by the Motivated Strategies for Learning Questionnaire (MSLQ) [51]. The questionnaire includes a total of 31 questions in six subscales, as illustrated in Table 3. The test was translated into Chinese to avoid any misunderstanding. Students answered using a 5-point

Likert scale, ranging from (5) strongly agree to (1) strongly disagree. An individual's score on a particular scale was calculated as an average of answers to the subscale's respective questions. We also evaluated the reliability of each subscale by Cronbach's alpha and the results are displayed in Table 3. The overall reliability of the questionnaire was 0.88.

Table 3. MSLQ profile

Subscale	Count	Reliability	Subscale	Count	Reliability
Intrinsic Goal Orientation	4	0.78	Extrinsic Goal Orientation	4	0.63
Task Value	6	0.91	Control Belief of Learning	4	0.89
Self-efficacy	8	0.94	Test Anxiety	5	0.79
			Overall	31	0.88

3.4 Data Collection and Analysis

Resource type accessed and duration of the access itself played a major role in the description of the student behavior. Based on the students' interaction with PIPLS, we extracted a total of 13 variables for the analysis. Complete enumeration of these variables, along with their basic statistical properties, can be found in Table 4. All of the time-related variables are measured in seconds. Note that only variables 1-3 and 5-7 originate from the raw measurement data. The remaining variables are derived according to the following formulas:

$$\begin{gathered} t_t = \sum_{x \in (l, e, q)} t_x \\ h_t = \sum_{x \in (l, e, q)} h_x \\ \forall x \in \{l, e, q, t\} : a_x = \frac{t_x}{h_x} \\ f_q = \frac{t_q}{t_t} \times 100 \end{gathered} \tag{1}$$

Despite the small size of our test group, Box Plots of our crucial variables, t_t, t_e, t_q still revealed numerous cases very distant from the IRQ region, as illustrated in Figure 1.

Table 4. Variables extracted from PIPLS

#	Variable	Variable Description
1	t_l	Time Spent Accessing Lecture Slides
2	t_e	Time Spent Accessing Exercises
3	t_q	Time Spent Accessing SGQ
4	t_t	Total Time Spent accessing All Resources
5	h_t	Page Hits on Lecture Slides
6	h_e	Page Hits on Exercises
7	h_q	Page Hits on SGQ
8	h_t	Total Page Hits on All Resources
9	a_t	Average Time per Lecture Slide Access
10	a_e	Average Time per Exercise Access
11	a_q	Average Time per SGQ Access
12	a_t	Average Time per Access
13	f_q	Fraction of Time Spent on SGQ (%)

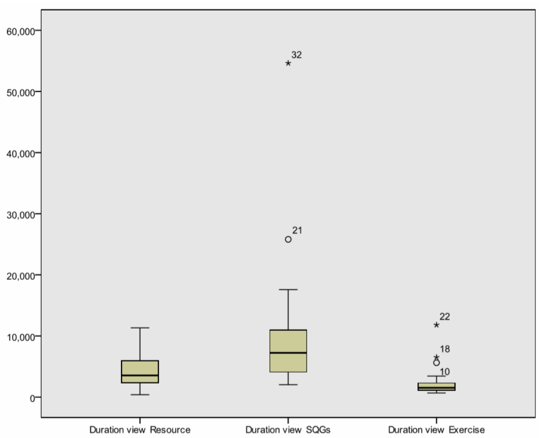

Figure 1. Box Plot of t_l^T, t_e^T, t_q^T

Since these deviations could negatively project onto the clustering process, we decided to transform these variables to a scale of 1-3, following the methodology of Li and Tsai [10]. In the following, we will refer to the transformed variables as t_l^T, t_e^T, t_q^T. Further, in our search for learning behavior patterns, we attempted to group students based on the variables we collected. We deployed k-means clustering among various subsets of variables as dimensions of the Euclidean space. 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. Eventually, after defining the behavior patterns of our test unit, we attempted to perform group comparison methods in order to answer our research questions. There are two significant methods that come into consideration: (1) Analysis of variance (ANOVA), and (2) Kruskal–Wallis one-way analysis of variance (K-W Test). The deciding factor for selection of a proper method is distribution/homogeneity of variance among the variables. Apparently, all variables except for overall performance deviated from normal distribution, as assessed by the Shapiro-Wilk test. Therefore, option (2) the K-W Test appeared to be the right choice. In cases where the K-W Test pointed out a significant hypothesis, we used the post-hoc Mann-Whitney U Test (M-W U Test) for pairwise comparison of the individual clusters (t-test is not suitable for distribution reasons, similar to ANOVA).

4 Result

4.1 Students' Resource Engagement

Without any further analysis, Table 5 gives us some basic insight into students' preferences for study materials. We can see that, based on the mean values, SGQ were the most widely used study materials. We can clearly deduce the importance of time measurements rather than just page hits. This follows from the fact that mean of Time Spent Accessing Lecture Slides is much greater than mean of Time Spent Accessing Exercises, although mean of Page Hits on Lecture Slides is much less than mean of Page Hits on Exercises. Therefore, any conclusion based solely on page hits could be misleading. We also observe that despite a comparatively high mean value $\overline{t_q}$, students only spent an average of 30.25% of their time on SGQ, as stated by $\overline{f_q}$. We attribute this effect to the fact that a small number of students spent an extremely long time on the SGQs, as illustrated in Figure 1, and a relatively high standard deviation of t_q. These extremities along with arguably higher values in the SD column underline the need to separate students into more homogeneous groups, thus forming their access behavior.

Table 5. Mean and standard deviation of variables extracted from PIPLS

#	Variable	Variable Description	Mean	SD
1	t_l	Time Spent Accessing Lecture Slides	4220.61	2756.87
2	t_e	Time Spent Accessing Exercises	2244.88	2143.06
3	t_q	Time Spent Accessing SGQ	9478.03	9610.68
4	t_t	Total Time Spent accessing All Resources	15943.52	11669.54
5	h_l	Page Hits on Lecture Slides	22.67	13.35
6	h_e	Page Hits on Exercises	92.58	68.71
7	h_q	Page Hits on SGQ	106.67	104.73
8	h_t	Total Page Hits on All Resources	221.91	151.74
9	a_t	Average Time per Lecture Slide Access	181.71	21.23
10	a_e	Average Time per Exercise Access	22.96	3.74
11	a_q	Average Time per SGQ Access	89.17	7.69
12	a_t	Average Time per Access	89.17	7.69
13	f_q	Fraction of Time Spent on SGQ (%)	30.25	8.55

4.2 Students' Learning Access Patterns

Deploying the analysis mentioned in the previous chapter, we identified three clusters based on variables t_l^T, t_e^T, t_q^T. These clusters evince differences in students' viewing patterns, and therefore we assign them slightly suggestive names:

(1) Less-engaged students, C_{less}

(2) Moderately-engaged students, C_{mod}

(3) Highly-engaged students, C_{high}

In the following, however, we will provide evidence showing that certain cluster properties do not follow the intuitive assumption. Exact values regarding the cluster's centroids and sizes are displayed in Table 6. It comes as no surprise that the centroid of C_{high} dominates both C_{mod} and C_{less} in all of the dimensions t_l^T, t_e^T, t_q^T. We presume that these are performance-oriented, highly motivated students. An arguably less expected result, however, emerged with the comparison of C_{mod} and C_{less} with regards to Exercise access, t_e^T. Whereas C_{mod} dominates C_{less} in the remaining dimensions, C_{less} carries a significantly higher value of t_e^T than (supposedly) more engaged C_{mod}. Moreover, if we look at a similar cluster analysis performed by Li and Tsai [10], we can see that the least engaged cluster is inferior to the remaining ones in every dimension considered. After classifying the students into homogeneous groups based on similarities in their course material viewing patterns, we performed the K-W Test in order to compare C_{high}, C_{mod} and C_{less} with regards to the set of variables collected. The test outcome is depicted in Table 7. We observed a statistically significant difference in all of the aspects measured except for a_l, a_e, a_q. The results of the K-W Test yielded ($x^2(2, N=33) =$ 2.594, $p = 0.273$), ($x^2(2, N=33) = 0.258$, $p = 0.879$), and ($x^2(2, N = 33) = 2.457$, $p = 0.293$), respectively.

Table 6. Result clusters and their centroids

Variable	Description	Cluster		
		C_{mod}	C_{high}	C_{less}
n	Elements in cluster	10	12	11
t_l^T	Time Spent accessing Lecture Slides (transformed)	2.20	2.67	1.09
t_e^T	Time Spent accessing Exercises (transformed)	1.20	2.83	1.82
t_q^T	Time Spent accessing SGQ (transformed)	2.10	2.67	1.18

Table 7. Analysis of material access

Var	C_{mod}		C_{high}		C_{less}		p	M-W U Test
	Mean	SD	Mean	SD	Mean	SD		
t_l	2018.6	683.61	3291.5	1430.93	905.09	356.27	0.000**	$C_{high} > C_{mod}$ * $C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ **
t_e	1131.9	275.41	3903.67	2903.69	1447.09	336.34	0.000**	$C_{high} > C_{mod}$ ** $C_{high} > C_{less}$ ** $C_{less} > C_{mod}$ *
t_q	7565.4	2892.76	16092.58	13313.29	4000.91	1977.32	0.000**	$C_{high} > C_{mod}$ ** $C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ **
t_l	10715.9	3184	23287.75	13307.48	6353.09	2315.35	0.000**	$C_{high} > C_{mod}$ ** $C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ **
h_l	22.4	7.53	34	13.08	10.55	3.93	0.000**	$C_{high} > C_{mod}$ * $C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ **
h_e	52.3	8.62	152.58	85.11	63.73	14.67	0.000**	$C_{high} > C_{mod}$ ** $C_{high} > C_{less}$ ** $C_{less} > C_{mod}$ *
h_q	85.9	34.43	180.42	142.84	45.09	21.48	0.000**	$C_{high} > C_{mod}$ * $C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ **
h_t	160.6	38.93	367	167.02	119.36	32.33	0.000**	$C_{high} > C_{mod}$ ** $C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ *
a_l	90.15	1.79	95.5	14.74	86.43	8.41	0.273	-
a_e	21.43	2.67	24.49	5.48	22.69	0.24		-
a_q	88.74	4.9	90.48	10.37	88.12	6.76		-
a_t	65.97	4.81	62.75	11.14	52.38	5.71	0.001**	$C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ **
f_q	33.79	4.96	31.46	11.26	25.71	5.39	0.032*	$C_{mod} > C_{less}$ **

*$p < 0.05$. **$p < 0.01$.

In a previous study conducted by Li and Tsai [10], the authors concluded that the less-use students spent significantly less effort accessing any of the learning materials provided when compared to every other cluster. Our result, in contrast, provides evidence that students from C_{less} spent significantly more effort on Exercise-related activities, namely t_e, h_e, when compared to C_{mod}. Therefore, we have the opposite relation between C_{mod} and C_{less}, namely $C_{less} > C_{mod}$. Further, the same authors identified two clusters of highly-engaged students, with a significant difference in preferred learning materials among them. Our research, on the other hand, identified a single cluster C_{high}, which consists of students with significantly more effort measured in all kinds of learning materials when compared to both C_{mod} and C_{less}. Moreover, although we could not establish any relation with regards to the average time spent on individual study material types, our result shows that students from both C_{high} and C_{mod} spent a significantly longer time on an average access than C_{less}. Eventually, we observed a significant result with regards to the fraction of the total time students spent on SGQ, namely f_q. Students in C_{mod} spent proportionally more time on SGQ than those from C_{less}, thus appreciating them as more valuable when compared to C_{less}.

4.3 Students' Learning Performance

We collected two distinct variables that allowed us to measure student learning performance, namely Exercise Score and Exam Score. Findings gathered from analysis using the K-W Test and M-W U Tests are illustrated in Table 8. We can observe a significantly better performance in Exercise Score by C_{high} when compared to the remaining clusters. On the other hand, we could not establish any significant relation between C_{mod} and C_{less} in this aspect. What is more, we can notice that their mean values in Exercise Score are rather close. In the Exam Score part, our findings reveal that students from C_{less} performed significantly worse than those from C_{mod} and C_{high}. This result contrasts with our observation of Exercise Score in Table 8 and Exercise-related variables in Table 7, where C_{less} evinced more effort but not a significantly different result from C_{mod}. In addition, we did not conclude any significant difference between C_{mod} and C_{high} in the Exam Score, although overall engagement of students in C_{high} was observed to be significantly greater than those of C_{mod}, as follows from Table 7.

Table 8. Analysis of performance access

Variable	C_{mod}		C_{high}		C_{less}		K-W Test	M-W U Test
	Mean	SD	Mean	SD	Mean	SD	p	
Exercise Score	45.2	11.08	59.08	9.15	43.36	12.53	0.005**	$C_{high} > C_{mod}$ ** $C_{high} > C_{less}$ **
Exam Score	70.8	11.32	76.17	17	56.73	15.42	0.019*	$C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ *

*$p < 0.05$. **$p < 0.01$.

4.4 Students' Motivation

We have performed six K-W Tests in order to understand the differences in students' motivation. The MSLQ results revealed significant effects of C_{high}, C_{mod} and C_{less} on all six subscales. After the K-W Tests confirmed the differences between the three clusters, Pairwise M-W U Tests revealed statistically significant differences between C_{high} and C_{mod}, C_{high} and C_{less} in all six scales. These results reveal that C_{high} had higher Intrinsic Goal Orientation, Extrinsic Goal Orientation, Task Value, Control of Learning Beliefs, Self-Efficacy for Learning and Performance than C_{mod} and C_{less} Also C_{high} showed lower Test Anxiety than C_{mod} and C_{less}. In addition, the Intrinsic Goal Orientation $(U = 85.0, z - 2.872, p = 0.002)$, Task Value $(U = 97.0, z = -2.021, p = 0.022)$, Control of Learning Beliefs $(U = 87.0. z = -2.768, p = 0.003)$, and Self-Efficacy for Learning and Performance $(U = 75.0, z = -3.491, p = 0.000)$ of C_{mod} and C_{less} showed significant differences. The Test Anxiety level between C_{mod} and C_{less} only demonstrated a marginally significant difference $(U = 88.5, z = -1.638, p = 0.055)$. These results reveal that C_{mod} had higher Intrinsic Goal Orientation, Task Value, Control of Learning Beliefs, Self-Efficacy for Learning and Performance than C_{less} (see Table 9).

Table 9. Motivation questionnaire statistics

Variable	C_{mod}		C_{high}		C_{less}		K-W Test	M-W U Test
	Mean	SD	Mean	SD	Mean	SD	p	
Control of Learning Beliefs	2.30	0.48	2.92	0.29	1.55	0.52	0.000	$C_{high} > C_{mod}$ ** $C_{high} > C_{less}$ **
Extrinsic Goal Orientation	2.70	0.48	3.40	0.51	2.27	0.79	0.001	$C_{high} > C_{mod}$ ** $C_{high} > C_{less}$ **
Intrinsic Goal Orientation	2.00	0.47	2.83	0.39	1.27	0.47	0.000	$C_{high} > C_{mod}$ ** $C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ **
Self-Efficacy for Learning and Performance	2.40	0.52	2.92	0.29	1.27	0.47	0.000	$C_{high} > C_{mod}$ ** $C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ **
Task Value	3.00	0.47	3.50	0.52	2.27	0.90	0.001	$C_{high} > C_{mod}$ * $C_{high} > C_{less}$ ** $C_{mod} > C_{less}$ *
Test Anxiety	3.40	0.70	2.67	0.49	4.00	0.77	0.001	$C_{mod} > C_{high}$ ** $C_{less} > C_{high}$ **

*$p < 0.05$. **$p < 0.01$.

5 Discussion and Conclusion

5.1 Discussion

In this research, we collected data about the behavior of 33 students during 11 weeks of using PIPLS with SGQ activities. Based on the information gathered, we attempt to answer our four research questions as follows.

As for Question 1, we discovered that among the learning materials, the SGQ were one of the most favored. However, an average access to Lecture Slides was measured as being significantly longer than that of both SGQ Exercises. We conjecture that this phenomenon can be attributed to the fact that lecture slides contain wider content and a larger amount of information than the latter two.

Also, students tend to use SGQ and Exercises as a secondary source of information, or in other words as an extension to the slide content. This finding follows the previous studies that concluded similar relations between use of materials related to lectures and other learning materials. In the light of these studies, we assume that the fact that students spent more time on SGQ than lectures, i.e. $\overline{t_q} > \overline{t_l}$, is due to the proportion of the number of SGQs and the number of lecture slides provided to the students. Also, we are convinced that participation in the SGQ activities stimulate students' competitiveness and therefore increases their engagement in any future assignments of a similar nature.

Regarding Question 2, our research identified three clusters (C_{less}, C_{mod}, C_{high}) which evince different behavior patterns with regards to the time spent accessing various resources, i.e. t_l, t_e, t_q. We detected one cluster of students (C_{high}) that dominated the other two (C_{mod}, C_{less}) in all three leading variables. Among the other two groups, we could not conclude any universal dominance in their access behavior. This is implied by the significance of $C_{mod} > C_{less}$ in t_l, t_q, and $C_{less} > C_{mod}$ in t_e. We will provide an explanation of this behavioral abnormality in our answer to Question 4. This result contrasts with a previous study by Li and Tsai [10], 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"). We attempt to explain this difference by the size of the samples tested in both studies, and the individual composition of learning materials provided. We would also like to mention that this difference motivates future research in this area, especially with regards to the count of the experiment participants.

In response to Question 3, we successfully observed some significant differences among clusters. In particular, C_{high} showed significantly greater performance in Exercise Score; however, we did not identify any significant relation between C_{mod} and C_{less} This may be a little surprising since, as mentioned in our answer to Question 2, C_{less} showed significantly higher engagement in Exercise-related activities than C_{mod}.

C_{less}, on the other hand, showed worse Exam Score performance when compared to both C_{mod} and C_{high}.

This result appears to be in line with Li and Tsai [10] who concluded that "the students who invested more time and effort in viewing the online learning materials had better learning performance," and also with Lust et al. [44]. However, even though C_{high} showed significantly higher effort in all learning activities than C_{mod}, Final Score did not identify any statistically significant relation between them. Based on these facts, we conclude that different learning patterns may yield variable time-to-performance ratios, and therefore conjecture that proper composition of the learning materials used is as important as the total time spent accessing them. This assumption also points towards the importance of self-regulation when SGQ are in effect.

As for Question 4, our result revealed a significant difference between the clusters. C_{high} dominates the remaining clusters in all of the positive scales and is dominated by the remaining clusters in the negative scale (Test Anxiety). Our findings extend Li and Tsai [10] results as they only found a relationship between clusters in Intrinsic Goal Orientation, Task Value, Self-Efficacy for Learning and Performance. With regards to Test Anxiety, we measured only a marginal significance in the relation between C_{mod} and C_{less}, in favor of C_{less}. We suggest that this result is the reason why students of C_{less} spent significantly longer viewing the Exercise materials than students of C_{mod}. However, our data only bear a marginal significance; therefore, we propose this conjecture for future research.

5.2 Conclusion

SGQ is regarded as one of the essential cognitive strategies that encourage and monitor awareness and enhance self-regulatory capabilities. Besides generating questions, viewing SGQ and other online learning materials is the most frequently performed online learning activity during SGQ treatments. Hence, the need to understand how students view SGQ and other different learning materials and how that behavior influences their learning outcome raises a critical concern. Our experiment with 33 students of an elective programming course revealed several significant findings. We found that the students viewed SGQ for longer and more frequently than other materials (i.e., lecture slides and exercises). More important, our results revealed that the viewing behaviors of students during the SGQ activities showed great variety and were divided into three behavior clusters: the "Highly-engaged students" cluster which dominates the other clusters in the use of all kinds of materials, "Moderately-engaged students" who spent more time on lecture slides and SQG, and "Less-engaged students" who rarely used any learning material but more frequently used exercises than the "Moderately-engaged students." We also observed that viewing behavior is connected with learning achievement during SGQ activities. More specifically, the results showed significantly worse performance of "Less-engaged students" compared to both "Highly-engaged students" and "Moderately-engaged students" during the final examination. Finally, students' viewing behaviors were strongly associated with their motivation. "Less-engaged students" may have more concerns and worry about failing examinations, and therefore spent significantly longer viewing the Exercise materials than the "Moderately-engaged students." Further, our research has established the importance of the measurement of time-related variables alongside simple page-hits, as presented in the result analysis. This evidence also calls for verification of previous research studies that did not have time measuring tools at their disposal. We believe that the newer version of PIPLS, which was successfully deployed in an experiment of a moderate size, can serve future research in larger scale classes.

Acknowledgements

This study is supported by the Ministry of Science and Technology of the Republic of China under contract number: MOST 105-2511-S-259-004-MY3.

References

[1] C. H. Lai, P. D. Tho, J. S. Liang, Design and Evaluation of Question-Generated Programming Learning System, *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, Hamamatsu, Japan, 2017, pp. 573-578, DOI: 10.1109/IIAI-AAI.2017.107.

[2] F. Y. Yu, Scaffolding Student-generated Questions: Design and Development of a Customizable Online Learning System, *Computers in Human Behavior*, Vol. 25, No. 5, pp. 1129-1138, September, 2009, DOI: 10.1016/j.chb.2009.05.002.

[3] F.-Y. Yu, Y.-H. Liu, T.-W. Chan, The Efficacy of a Web-based Domain Independent Question-posing and Peer Assessment Learning System, *Proceedings International Conference on Computers in Education*, Auckland, New Zealand, 2002, pp. 641-642 Vol. 1, DOI: 10.1109/CIE.2002.1186028.

[4] E. Balajthy, Using Student-constructed Questions to Encourage Active Reading, *Journal of Reading*, Vol. 27, No. 5, pp. 408-411, February, 1984.

[5] A. T. Barlow, J. M. Cates, The Impact of Problem Posing on Elementary Teachers' Beliefs About Mathematics and Mathematics Teaching, *School Science and Mathematics*, Vol. 106, No. 2, pp. 64-73, February, 2006, DOI: 10.1111/j.1949-8594.2006.tb18136.x.

[6] F. Y. Yu, K. J. Pan, The Effects of Student Question-Generation with Online Prompts on Learning, *Educational Technology & Society*, Vol. 17, No. 3, pp. 267-279, July, 2014.

[7] F. Y. Yu, Y. J. Chen, Effects of Student-generated Questions as the Source of Online Drill-and-practice Activities on Learning, *British Journal of Educational Technology*, Vol. 45, No. 2, pp. 316-329, March, 2014, DOI: 10.1111/bjet.12036.

[8] F. Y. Yu, C. P. Wu, C. C. Hung, Are There Any Joint Effects of Online Student Question Generation and Cooperative Learning?, *The Asia-Pacific Education Researcher*, Vol. 23, No. 3, pp. 367-378, September, 2014, DOI: 10.1007/s40299-013-0112-y.

[9] A.-F. Lai, The Undergraduate Student's Perception and Navigation Behaviors toward On-line Student-generating Questions Activities of Data Structures Curriculum in Engineering Education, *5th International Conference on Information and Education Technology*, Tokyo, Japan, 2017, pp. 39-43.

[10] L.-Y. Li, C.-C. Tsai, Accessing Online Learning Material: Quantitative Behavior Patterns and Their Effects on Motivation and Learning Performance, *Computers & Education*, Vol. 114, pp. 286-297, November, 2017, DOI: https://doi.org/10.1016/ j.compedu.2017.07.007.

[11] G. Lust, J. Elen, G. Clarebout, Regulation of Tool-use within a Blended Course: Student Differences and Performance Effects, *Computers & Education*, Vol. 60, No. 1, pp. 385-395, January, 2013, DOI: https://doi.org/10.1016/j.compedu.2012.09.001.

[12] O. H. T. Lu, J. C. H. Huang, A. Y. Q. Huang, S. J. H. Yang, Applying Learning Analytics for Improving Students Engagement and Learning Outcomes in an MOOCs Enabled Collaborative Programming Course, *Interactive Learning Environments*, Vol. 25, No. 2, pp. 220-234, April, 2017, DOI: 10.1080/10494820.2016.1278391.

[13] J. H. Zhang, Y. X. Zhang, Q. Zou, S. Huang, What Learning Analytics Tells Us: Group Behavior Analysis and Individual Learning Diagnosis Based on Long-Term and Large-Scale Data, *Educational Technology & Society*, Vol. 21, No. 2, pp. 245-258, April, 2018.

[14] J. C. Y. Sun, C. T. Lin, C. Chou, Applying Learning Analytics to Explore the Influence of Online Learners' Motivation on Their Online Learning Behavioral Patterns, *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, Kumamoto, Japan, 2016, pp. 377-380, DOI: 10.1109/IIAI-AAI.2016.186.

[15] C.-W. Kuo, F.-Y. Yu, Difficulties of Multiple-choice Question-Generation Encountered by Elementary Students in Social Studies: A Case Study, *International Journal of Information and Education Technology*, Vol. 8, No. 4, pp. 259-266, April, 2018.

[16] F. Y. Yu, W. S. Wu, H. C. Huang, Promoting Middle School Students' Learning Motivation and Academic Emotions via Student-Created Feedback for Online Student-Created Multiple-Choice Questions, *The Asia-Pacific Education Researcher*, Vol. 27, No. 5, pp. 395-408, October, 2018, DOI: 10.1007/s40299-018-0398-x.

[17] F. Y. Yu, Y. H. Liu, T. W. Chan, A Web-based Learning System for Question-Posing and Peer Assessment, *Innovations in Education and Teaching International*, Vol. 42, No. 4, pp. 337-348, November, 2005, DOI: 10.1080/14703290500062557.

[18] C. Chin, D. E. Brown, B. C. Bruce, Student-generated Questions: A Meaningful Aspect of Learning in Science, *International Journal of Science Education*, Vol. 24, No. 5, pp. 521-549, May, 2002, DOI: 10.1080/09500690110095249.

[19] F.-Y. Yu, Y. Liu, The Comparative Effects of Student Question-posing and Question-answering Strategies on Promoting College Students' Academic Achievement, Cognitive and Metacognitive Strategies use, *Journal of Education and Psychology*, Vol. 31, No. 3, pp. 25-52, September, 2008.

[20] M. Teplitski, T. Irani, C. J. Krediet, M. Di Cesare, M. Marvasi, Student-Generated Pre-Exam Questions is an Effective Tool for Participatory Learning: A Case Study from Ecology of Waterborne Pathogens Course, *Journal of Food Science Education*, Vol. 17, No. 3, pp. 76-84, July, 2018, DOI: 10.1111/1541-4329.12129.

[21] C.-C. Hsu, T.-I. Wang, Applying Game Mechanics and Student-generated Questions to an Online Puzzle-based Game Learning System to Promote Algorithmic Thinking Skills, *Computers & Education*, Vol. 121, pp. 73-88, June, 2018, DOI: https://doi.org/10.1016/j.compedu.2018.02.002.

[22] C. H. Lai, P. D. Tho, Development of a Programming Learning System Based on a Question Generated strategy, *24th International Conference on Computers in Education*, Mumbai, India, 2016, pp. 445-450.

[23] F. Y. Yu, C. L. Su, A Student-constructed Test Learning System: The Design, Development and Evaluation of Its Pedagogical Potential, *Australasian Journal of Educational Technology*, Vol. 31, No. 6, pp. 685-698, December, 2015.

[24] H. C. Yeh, P. Y. Lai, Implementing Online Question Generation to Foster Reading Comprehension, *Australasian Journal of Educational Technology*, Vol. 28, No. 7, pp. 1152-1175, August, 2012.

[25] F.-Y. Yu, The Learning Potential of Online Student-constructed Tests with Citing Peer-generated Questions, *Interactive Learning Environments*, Vol. 27, No. 2, pp. 226-241, 2019, DOI: 10.1080/10494820.2018.1458040.

[26] F.-Y. Yu, C.-P. Wu, The Effects of an Online Student-Constructed Test Strategy on Knowledge Construction, *Computers & Education*, Vol. 94, pp. 89-101, March, 2016, DOI: https://doi.org/10.1016/j.compedu.2015.11.005.

[27] F.-Y. Yu, Y.-L. Chang, H.-L. Wu, The Effects of an Online Student Question-Generation Strategy on Elementary School student English Learning, *Research and Practice in Technology Enhanced Learning*, Vol. 10, No. 1, pp. 1-16, December, 2015, DOI: 10.1186/s41039-015-0023-z.

[28] F. Y. Yu, Multiple Peer-assessment Modes to Augment Online Student Question-Generation Processes, *Computers & Education*, Vol. 56, No. 2, pp. 484-494, February, 2011, DOI: 10.1016/j.compedu.2010.08.025.

[29] S. Zhang, X. Cai, W. Zhou, Y. Wang, Green 5g Enabling Technologies: An Overview, *IET Communications*, Vol. 13, No. 2, pp. 135-143, January, 2019.

[30] X. Ba, Y. Wang, J. Fan, W. Xiang, User-priority-based Resource Allocation for Device-to-device Communications in 5G Underlaying Cellular Networks, *IET Communications*, Vol. 13, No. 8, pp. 1016-1024, May, 2019.

[31] T. Mohamed, M. Ali, I. Hani, A Challenge-response Mechanism for Securing Online Social Networks Against Social Bots, *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 32, No. 1, pp. 1-13, August, 2019.

[32] S. Sabri, A. Aloui, A New Approach for the Recognition of Human Activities, *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 32, No. 4, pp. 211-223, October, 2019.

[33] J. Qin, Y. Jiang, Recommender Resources Based on Acquiring User'S Requirement and Exploring User'S Preference with Word2Vec Model in Web Service, *International Journal of Internet Protocol Technology*, Vol. 12, No. 3, pp. 144-152, July, 2019.

[34] M. Nirmala, M. R. Babu, Fuzzy-based Fake Information Detection Algorithm to Define the User Trust on the Content of Social Networks, *IET Networks*, Vol. 8, No. 6, pp. 367-371, November, 2019.

[35] A. Kumari, S. Tanwar, S. Tyagi, N. Kumar, Verification and Validation Techniques for Streaming Big Data Analytics in Internet of Things Environment, *IET Networks*, Vol. 8, No. 3, pp. 155-163, May, 2019.

[36] P.-Y. Cheng, Y.-N. Su, Y.-C. Chien, T.-T. Wu, Y.-M. Huang, An Investigation of Visual Fatigue in Elementary School Students Resulting from Reading e-books, *Journal of Internet Technology*, Vol. 19, No. 5, pp. 1285-1292, September, 2018.

[37] K. Brennan, Audience in the Service of Learning: How Kids Negotiate Attention in an Online Community of Interactive Media Designers, *Learning, Media and Technology*, Vol. 41, No. 2, pp. 193-212, April, 2016, DOI: 10.1080/17439884.2014.939194.

[38] J. Wang, Network Distance Education Platform Control System Based on Big Data, *International Journal of Internet Protocol Technology*, Vol. 12, No. 3, pp. 173-180, July, 2019.

[39] C.-H. Lai, C.-M. Chu, C.-H. Chen, An Examination of Online Discussion Quality of STEM-based Instruction, *Journal of Internet Technology*, Vol. 19, No. 6, pp. 1943-1950, November, 2018.

[40] E. Fouh, D. A. Breakiron, S. Hamouda, M. F. Farghally, C. A. Shaffer, Exploring Students Learning Behavior with an Interactive Etextbook in Computer Science Courses, *Computers in Human Behavior*, Vol. 41, pp. 478-485, December, 2014, DOI: https://doi.org/10.1016/j.chb.2014.09.061.

[41] M. Bienkowski, M. Feng, B. Means, *Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief*, https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf.

[42] S. Freitas, D. Gibson, C. Du Plessis, P. Halloran, E. Williams, M. Ambrose, I. Dunwell, S. Arnab, Foundations of Dynamic Learning Analytics: Using University Student Data to Increase Retention, *British Journal of Educational Technology*, Vol. 46, No. 6, pp. 1175-1188, November, 2015, DOI: 10.1111/bjet. 12212.

[43] F. Martin, J. C. Whitmer, Applying Learning Analytics to Investigate Timed Release in Online Learning, *Technology, Knowledge and Learning*, Vol. 21, No. 1, pp. 59-74, April, 2016, DOI: 10.1007/s10758-015-9261-9.

[44] G. Lust, M. Vandewaetere, E. Ceulemans, J. Elen, G. Clarebout, Tool-use in a Blended Undergraduate Course: In Search of User Profiles, *Computers & Education*, Vol. 57, No. 3, pp. 2135-2144, November, 2011, DOI: https://doi.org/10.1016/j.compedu.2011.05.010.

[45] K. M. Y. Law, S. Geng, T. M. Li, Student Enrollment, Motivation and Learning Performance in a Blended Learning Environment: The Mediating Effects of Social, Teaching, and Cognitive Presence, *Computers & Education*, Vol. 136, pp. 1-12, July, 2019, DOI: 10.1016/j.compedu.2019.02.021.

[46] P. G. Barba, G. E. Kennedy, M. D. Ainley, The Role of Students' Motivation and Participation in Predicting Performance in a MOOC, *Journal of Computer Assisted Learning*, Vol. 32, No. 3, pp. 218-231, June, 2016, DOI: 10.1111/jcal.12130.

[47] H.-T. Hung, Clickers in the Flipped Classroom: Bring Your Own Device (BYOD) to Promote Student Learning, *Interactive Learning Environments*, Vol. 25, No. 8, pp. 983-995, November, 2017, DOI: 10.1080/10494820.2016.1240090.

[48] M. Y. Yi, Y. Hwang, Predicting the Use of Web-based Information Systems: Self-Efficacy, Enjoyment, Learning Goal Orientation, and the Technology Acceptance Model, *International Journal of Human-Computer Studies*, Vol. 59, No. 4, pp. 431-449, October, 2003, DOI: https://doi.org/10.1016/S1071-5819(03)00114-9.

[49] B. Giesbers, B. Rienties, D. Tempelaar, W. Gijselaers, Investigating the Relations between Motivation, Tool Use, Participation, and Performance in an e-learning Course Using Web-videoconferencing, *Computers in Human Behavior*, Vol. 29, No. 1, pp. 285-292, January, 2013, DOI: https://doi.org/10.1016/j.chb.2012.09.005.

[50] *Question2Answer*, http://www. question2answer.org/.

[51] P. R. Pintrich, *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*, NCRIPTAL-91-B-004, January, 1991.

Biographies

Pham Duc Tho is a Ph.D. student in the Department of Computer Science and Information Engineering of National Dong Hwa University, Taiwan. His current research focuses on digital learning, learning technology, and learning analytics.

Chih-Hung Lai is an associate professor in the Department of Computer Science and Information Engineering of National Dong Hwa University, Taiwan. He has won Distinguished Person Award on Computer Education from Ministry of Education. Dr. Lai's current research focuses on digital learning, learning technology, and learning analytics.

Ondrej Navratil is a Ph.D. Candidate at NDHU, Taiwan. He graduated Master in Mathemathical Methonds in Computer Science at FIT BUT, Czechia. Expertise in Graph Theory, Complexity and Formal Languages. Experienced teacher/lecturer and full stack developer.