

Does a Distributed Practice Strategy for Multiple Choice Questions Help Novices Learn Programming?

<https://doi.org/10.3991/ijet.v15i18.10567>

Lishan Zhang
Central China Normal University, Wuhan, China

Baoping Li (✉), Qiuji Zhang
Beijing Normal University, Beijing, China
libp@bnu.edu.cn

I-Han Hsiao
Arizona State University, Arizona, USA

Abstract—Learning how to program is becoming essential in many disciplines. However, programming cannot be easily learned, especially by non-engineering students. Therefore, it is important to teach non-engineering students to learn with efficient strategies. To discover an efficient learning strategy, we had 64 students practice programming with a simple learning management system and tracked all of their practice behaviors on multiple choice questions. The learning management system assigned one multiple choice question per day, but let students themselves decide their own practice frequencies. Students could also make unsynchronized communications by commenting on the questions. By analyzing their behavior patterns and other performance indicators, this paper compared the effect of two different practice strategies for multiple choice questions: distributed practice and massed practice. Our analysis found that students who adopted distributed practice significantly outperformed those who adopted massed practice on final exams ($p=0.031$). We further explored the possible reasons that led to this significant difference. Students who adopted distributed practice strategy tended to make higher percentage of first submission correctness, be more cautious while correcting errors, and be more constructive in posting question-related comments.

Keywords—Distributed practice, massed practice, programming language learning, multiple choice question, data analysis.

1 Introduction

Because programming is an essential skill for data analysis in domains such as economics, chemistry, biology and social science, mastering a programming language has become required for many college students [1]. Learning how to program can also be treated as a training process for computational and logical believing [2, 3]. However, programming cannot be easily learned by novices, which often leads to high classroom

dropout rates [4, 5]. Students often complain that programming is non-intuitive and find it hard to become accustomed to. To prevent students from becoming nervous about learning code before the start of instruction, in this study we explored how multiple-choice question (MCQ), which is the most common exercise type, can be used to help novices learn programming. MCQ has been used to help students learn programming in different ways. For example, Yang et al. [6] found that providing appropriate explanations for each alternative in MCQ can enhance students' learning. In this study, we hypothesized that different practice strategies using the same set of multiple-choice questions might also affect students' learning. Therefore, we assigned all of the practice questions through a customized learning management system, so that we could track students' practice behaviors and discover efficient and inefficient practice strategies, if any existed. Specifically, we observed whether students could learn more when their practice sessions were relatively distributed.

Practice is considered distributed when there is time between two consecutive exercises [7]. Distributed practice has proven to be an effective method in improving students learning outcomes in tasks such as word, text, and face memorization [8-10]. Students' memory can stay for a longer time through distributed practice than massed practice. This is also called the "spacing effect". However, when the idea of distributed practice was applied to domains that contain more procedural knowledge than declarative knowledge, it produced contradictory results. Cepeda et al. [7] found that distributed practice was not effective in improving student efficiency in understanding complex mathematics concepts. Budé et al. [11] claimed that students had a better understanding of statistical concepts in an introductory level class when learning was distributed. The lag between each learning seems to help students digest what they have learned. These existing studies suggest that distributed practice helps students to memorize things for a longer time and understand basic procedural knowledge but hardly help students develop a deep understanding of procedural knowledge. Programming is considered as procedural knowledge, because students need to learn how to describe computational process. But a programming novice also needs to remember and understand a lot of syntax, which just like word and text memorization. MCQ is the simplest practice type that is suitable for helping students recall programming syntax. We then hypothesized that distributed practice for MCQ could help novices learn how to program.

To explore the effect of distributed practice on multiple choice questions in the programming language learning, we had students practice with a system called "QuizIT" [12]. This is a system designed for students to conveniently conduct distributed practice. The preliminary experiment results using the system showed that students who frequently practiced with the system outperformed those who did not. However, it was not clear whether the outperformance was due to a higher volume of practice, better practice strategies, or both. High frequency of log-in activities is usually a strong predictor of students' grades, even without considering practice distribution [13]. In this study, we fixed the total amount of distinct multiple-choice questions that a student could practice to explore the effect of different practice strategies on programming language learning.

This study aims to answer the following research questions:

1. Do different multiple-choice question practice strategies affect students' programming learning outcomes?
2. How might a multiple-choice question practice strategy impact students' programming learning?

The remainder of the paper is organized as follows: We first review the related work, then we describe our study method. In the third section, we report our data analysis and relate the analysis to our research questions. In the last section, we discuss the results and conclude with remarks.

2 Related Work

2.1 Distributed practice

In the cognitive psychology literatures, it is widely accepted that learned knowledge is retained longer when practice is distributed rather than massed [14], [15]. This is referred to as the distributed practice effect [16], lag effect or the effect of spacing [7]. Distributed practice often co-appears with massed practice. The latter is defined as studying subject matter uninterruptedly or with only short breaks [11]. Sobel, Cepeda and Kapler [17] believed that the spacing effect is because of a memory advantage that occurs when people learn materials on several occasions. Shimoni [18] claimed that distributed practice not only consolidated memory but also refined students' understanding of the knowledge by allowing time for them to forget over the interval between successive presentations. Gerbier and Toppino [20] claimed that the spacing effect can be described by the deficient processing hypothesis: spaced repetition can result in more efficient encoding and better memory in the brain than immediate repetition. Although distributed practice is a better learning strategy than massed practice in most research studies, massed practice seems to be at least as effective as distributed practice in learning complex skills [16]. This likely explains why distributed practice has been applied most in domains that contain a large quantity of declarative knowledge, such as English. In this work, we used distributed practice to improve non-engineering students' learning of C programming, which was considered a basic but also difficult class for students who were studying programming for the first time.

Previous research has shown that a specific distribution of time can influence learning outcomes [20]. The optimal distribution is varied for different domains. For example, the optimal spacing between practices in memorizing task is approximately one month [14] [21]. In the domain of language learning and reading comprehension, the optimal spacing tends to be reduced to one week or even less [22][23]. In memorization tasks, when the spacing between practices was too long, a student's performance on recall tests gradually declined with an increase in the length of the gaps [24]. There should be some optimal ratio existed between the amount of exercises during each exercise and the space between two exercises. The optimal ratio may change upon domains.

2.2 Programming language learning

Programming is becoming a foundation class rather than a class only for engineering students. However, programming cannot be easily learned by novices [25] [26]. This is likely because many students memorize only the programming language's surface features, such as the syntax of variables and loops, but overlook the logic behind the code. There have been many studies facilitating students to learn how to program in a computer-based environment. Many previous works aimed to help students learn to program through automated assessments of students' codes and by providing feedback regarding coding errors. For instance, Web-CAT [27] and ASSYST [28] are assessment tools that use pattern-matching techniques to compare students' answers to the correct answers. SQL tutor [29] prompts students when their mistakes break a set of constraints that should not be broken by SQL language. QuizJET [30] is one example of facilitating automatic programming evaluation that uses parameterized exercises. CloudCoder [31] captured the knowledge components in the programming problems to discern with which topics the students might be struggling. In contrast, several programs made novice learners focus on programming logic instead of syntax by constructing a fun user interface. For example, Scratch [32] replaced tedious code with a set of colorful blocks and enabled novices to ignore programming syntax. Jakoš and Verber [33] designed an educational game to teach students with no prior knowledge of programming. The students who enjoyed playing also succeeded on final exams. Instead of making interactive and joyful learning environments, students' programming learning can also be leveraged through simple multiple choice questions with a complex mechanism [6].

More recently, researchers have started to use students' work pattern data collected from a learning management system to predict whether a student was improving programming skills [34]. In this study, we used students' work patterns to infer their practice strategy and explored whether and how different practice strategies affected learning outcomes. In particular, we identified the students who adopted the distributed practice strategy and those who adopted the massed practice strategy. We then compared the effect on learning between the two practice strategies.

3 Method

Because students' perceived ease of use can significantly affect students' satisfaction[34], we used a simple learning management system named QuizIT to send out our MCQ practice, and analyzed students' practice strategies from the log file of the system. This section first introduces the system, then our MCQ for practice. At last, we describe how we design the study to collect the data and our analysis strategy.

3.1 QuizIT system

QuizIT is a system that provides students a multiple-choice question to practice every day. The user interface is simple and straightforward. As soon as a student logs

into the system, the question of the day will display if the student has not completed it. A sample question is shown in Figure 1. Students might not log into the system every day; therefore, they do not always finish their questions on time. In this case, they can make up their unanswered questions by accessing the calendar view and navigating to the previous dates, as shown in Figure 2. Students can always retry or review any previously answered questions. All the multiple-choice questions in the system are posted by the programming language instructor in advance. The system displays the questions in the order desired by the instructor. A student can comment on each question, and the comment can be seen by him/her immediately as well as by all the other students. However, a student cannot view the comments made by others until he/she makes his/her own comment.

Every interaction is recorded by the system, and each interaction is given a timestamp. The recorded interactions for the follow-up analysis include: correct attempt, incorrect attempt, question retry (redo a previously answered question), and commenting on a question.

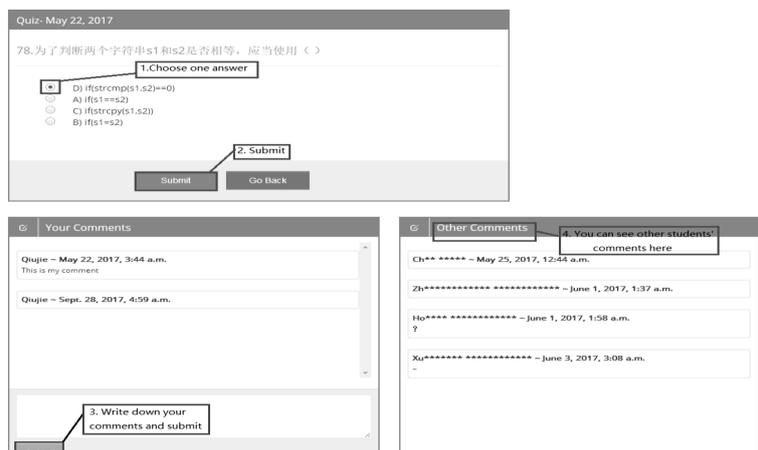


Fig. 1. A sample multiple choice question with comments. The order of action is represented by the number.

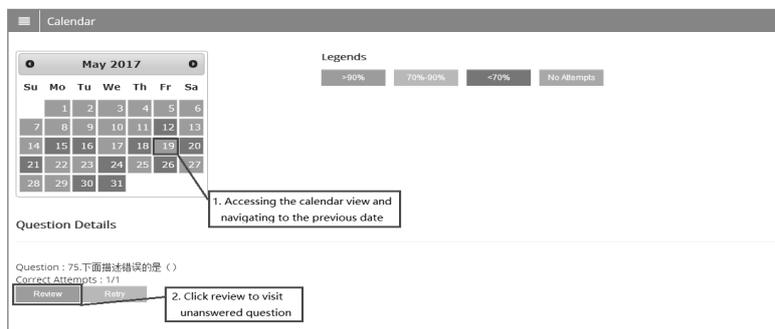


Fig. 2. The calendar view.

3.2 MCQ for practice

Our study sample was from a C programming class that had 64 first-year college students. The study lasted 45 days. Because QuizIT needed to provide one multiple choice question to the students per day, there were 45 multiple choice questions in the system. These questions covered all the topics taught between the midterm exam and the final exam. The order of the questions was synchronized with the class progress.

A total of 7 of the 45 questions asked students to select the output of a small piece of code, as shown in the example below. The example question helped students practice how pointer and loop in C programming is applied, which was one of the most complex of the 45 questions. However, students who were familiar with the concepts should have only needed a few minutes to answer the question.

```
What is the output of the program below?
void fun (int *b, int n, int *s)
{
int i; *s=0;
for (i=1;i<=n;i++) *s=*s+(b+i);
}
main ()
{
int x=1, a[ ]={2,3,4,5,6};
fun (a,3,&x);
printf("\n%d",x);
}
A. 7 B.12 C.18 D.1
The correct choice is B
```

The remaining 38 questions aimed to aid students in recalling the basic concepts taught in the class, as shown in the below example. Students were expected to spend approximately 1 minute to answer this type of question.

The type of returned value of a function "fun" is defined by:

```
The return statement
The function that calls the function "fun"
The context of the function over running time
The type nominated by the definition of the function
The correct choice is D
```

In summary, the practice questions were not designed to be difficult for students to answer but to provide students with the opportunity to review what they learned in class.

3.3 Data collection

All the student participants were majoring in psychology, and this was their first programming class. Students attended the class once a week. The class lasted 210 minutes per day, including a 10-minute break every 45 minutes. The QuizIT system was not the only method students could use to learn and practice. Therefore, to eliminate the effect from students' using different learning methods that could not be recorded by the system, we began to collect the data after students completed their midterm exams and stopped collection one week before the final exam. We assumed that if students used unique practice methods, this would lead to different learning outcomes and be reflected in their performance on the midterm exams. Therefore, the midterm and final exams were used as a pretest and posttest, respectively, to measure student learning efficiency.

Our primary aim was to learn which practice strategies students employed when using the system and the effects of those different strategies. We hypothesized that practice distribution should be one of the primary effects. Because of the learning management system used, we assigned one multiple choice question per day per student. Therefore, there were 45 different multiple-choice questions available for practice. In addition, students could practice the same question multiple times to reinforce their learning. Students did not need to answer the assigned question every day as long as they finished all the questions by the end the study.

We designed the indicators in Table 1 to describe how the students used the 45 multiple choice questions to practice. We wanted to determine how much effort each student put into practicing. To this end, we used two indicators: total amount of MCQ practice and total amount of MCQ practice time. Then, since we suspected that practice distribution would affect learning, we calculated the average number of days between two consecutive practice sessions. If two students practiced approximately the same amount and the first student practiced with fewer days between two consecutive practice sessions than did the second, then the former student would have practiced for less time per session. Therefore, we believed that the former student relatively distributed his/her practice. Because students could repeat a question when they answered it incorrectly, we recorded the time between two consecutive attempts and used the median value to represent the time a student usually took to change his/her answers, which is the resubmission time. The learning management system also allowed students to interact with each other by posting comments to the questions. Previous works suggested that social interaction could enhance students' learning performance [36, 37]. Therefore, we designed two indicators to describe students' interaction behaviors: the number of comments and the ratio of question-related comments.

In addition to students' practice strategies, we also designed indicators to show students' performance using the system. A correct answer on the first submission of a question was usually a positive indicator of a student's competence. Moreover, we calculated students' overall percentage of correct answers, which was students' performance on the questions that they had previously answered.

Table 1. The indicators of students' activities.

Category	Indicator name	Description
How do students practice with the system?	Days between practice (practice frequency)	Average number of days between two consecutive practices
	Resubmission time	Median time used for resubmission of an answer after an incorrect response
	Total amount of MCQ practice time	The total amount of time a student spent practicing using the system
	Total amount of MCQ practice	The total number of exercises a student completed on average, including the first submission and repeat submissions
	Ratio of question-related comments	The percentage of comments that were related to a question over the total number of comments
	Number of comments	The number of comments posted by a student
How well do students perform in the system?	Correctness on the first check	The percentage correct responses on the first attempt
	Overall correctness	The percentage of a student's correct responses over his/her total attempts

4 Data Analysis

4.1 Descriptive results

A total of 64 college students participated in the 45-day study. As required by the instructor, every student had to complete each question at least once. On average, students practiced 86.08 (SD=38.588) questions, including both the attempt and subsequent attempts. The percentage of first check correctness was 0.59 (SD=0.112). On average, students used the system to practice every 8.49 (SD=4.158) days, which was their practice frequency. Students spent 5807.01 (SD=2789.20) seconds using the practice system and took 5.28 (SD=4.575) seconds to resubmit an answer after an incorrect attempt. Students tended not to comment. They made 11.97 (SD=11.449) comments on average over the study, and 3.05 (SD=4.920) of the comments were related to the question commented on. In most cases, students posted meaningless comments to get the access to see others' comments. On average, students scored 79.53 (SD=5.822) on the midterm exam, and 78.08 (SD=14.042) on the final exam.

4.2 Students scored significantly higher when their practice using MCQ was distributed

By definition, distributed practice means that students should practice using many short sessions over a long period. In contrast, massed practice should consist of fewer, but longer, practice sessions. Additionally, the total practice time for each strategy should be equal. Assume that two students both practice 60 questions. One student completes two questions per session and practices 30 times while the other student completes 30 questions per session and practices two times. The former student's practice is more distributed than is the latter students.

Given that the total number of different MCQ is fixed, practice frequency can be potentially used to classify students into distributed and massed practice groups. However, because students can redo the practice questions that they have done before, it is possible that the students with a higher practice frequency practiced as much as the students with a lower practice frequency each time. In this case, practice frequency would show how hard a student studied. If so, the two types of students would have a different number of practices. However, according to our data analysis, the number of questions students practiced did not correlate with practice frequency ($r=0.045$, $p=0.726$). This allowed us to distinguish students' practice strategies from their practice frequencies. We used the median value of the days between practices to divide students into two groups. A student whose practice frequency was below 7.5 days was in the distributed practice group. A student whose practice frequency was at least 7.5 days was in the massed practice group. Because the class was held every 7 days, all the students in the distributed practice group practiced using the system at least once between two consecutive classes. There were 22 students in the distributed practice group and 42 students in the massed practice group. Students in the distributed practice group completed the system exercises every 5.00 (SD=1.349) days and completed 83.73 (SD=23.749) MCQ practices. In contrast, students in the massed practice group completed the system exercises every 10.32 (SD=3.958) days and completed 87.31 (SD=44.662) MCQ practices. The two groups of students performed similarly on their midterm exam. The distributed practice group students scored 80.96 (SD=5.533) on their midterm exam, while the massed practice group students scored 78.79 (SD=5.895). An ANOVA showed that the difference was not significant ($F=2.036$, $p=0.159$). The results confirmed that the two groups of students had similar initial competences and practiced a similar amount of questions overall but differed in their practice frequency. This classification guaranteed that any student in the distributed practice group had his/her practice more "distributed" than did the students in the massed practice group. Therefore, our following analysis shows whether a student learns more when his/her practice is relatively distributed.

Once the students were classified into distributed and massed practice groups, we compared their learning outcomes, which were their grades on their final exams. The students in the distributed practice group scored higher (Mean=84.14, SD=10.72) than the students in the massed practice group (Mean=74.91, SD=14.63). We conducted an ANCOVA using students' grades for the midterm exam as the covariance. The difference was significant ($F=4.862$, $p=0.031$). The comparison of the adjusted means is illustrated in Figure 3. The next section includes a further analysis to explore how distributed practice leveraged learning.

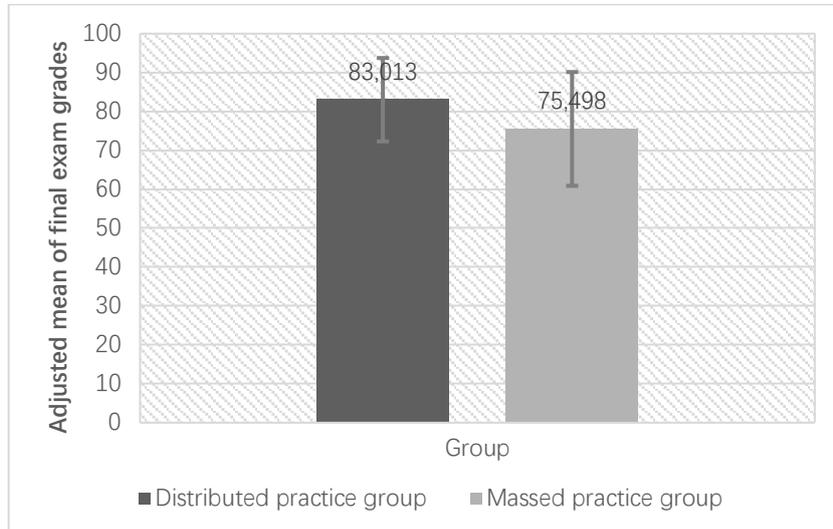


Fig. 3. Adjusted mean of final exam grades (exclude the influence of students’ midterm performance) of the two groups of students. The vertical bar indicated the original standard deviation.

4.3 How did distributed practice leverage learning?

Because the students were divided into two groups, distributed practice vs. massed practice, we had the opportunity to explore the effect of the two practice patterns. Table 2 lists the performance of the two groups of students on all the indicators.

Table 2. Distributed practice group vs. massed practice group

	Distributed practice N=22	Massed practice N=42
Midterm exam grades	80.96 (SD=5.533)	78.79 (SD=5.895)
Final exam grades	84.14 (SD=10.72)	74.91 (SD=14.63)
Total amount of MCQ practice	83.73 (SD=23.749)	87.31 (SD=44.662)
Total amount of MCQ practicing time	6291 second (SD=3296.8)	5553 second (SD=2489.1)
Medium time for answer resubmission	6.45 sec (SD=5.361)	4.67 sec (SD=4.039)
Correctness on the first check	0.643 (SD=0.0628)	0.566 (SD=0.1227)
Proportion of correct attempts	0.643 (SD=0.0765)	0.627 (SD=0.0915)
Number of comments	13.59 (SD=11.999)	11.12 (SD=11.204)
Proportion of topic-related comments	0.346 (SD=0.3476)	0.194 (SD=0.2934)

Table 2 shows that students who adopted a distributed practice strategy had higher rates of first submission correctness. Correctness on the first response could be affected by a student’s initial competence, which was measured using the midterm exam. So we conducted an ANCOVA to test the significance of the difference ($p=0.022$, $F=5.493$) with the midterm score as the covariance. The results suggest that when a student had few practice questions to complete per session, the student might have been more

cautious when answering the question than would another student who had more questions to complete per session [38] [14]. Using the timestamp that recorded when students entered and exited each multiple-choice question, we were able to calculate the total amount of practice time for each student. Students in the distributed practice group practiced for a little longer than students in the massed practice group. However, the difference was not significant ($F=1.010$, $p=0.319$) according to an ANOVA. Students in the distributed practice group spent longer to correct an incorrect response, but this difference was also not significant ($F=2.249$, $p=0.139$).

The students in the distributed practice group made a similar number of comments as the students in the massed practice group ($p=0.416$, $F=0.669$) but made a higher percentage of topic-related comments, which showed a marginal significance ($p=0.068$, $F=3.443$). The tests were conducted using an ANOVA. Because students had to leave several comments to view others' comments, leaving comments not related to the topic was likely a sign of a student being "active", according to Chi's ICAP framework [39]. In contrast, leaving topic-related comments was a sign of being a "constructive" or "interactive" student who had a higher level of engagement in learning than an "active" student, which should lead to higher learning gains.

The results suggest that students tended to make better use of a question when they did not have to complete many practice questions at a time. It seems that, because students in the distributed practice group were more cautious when answering the practice questions, they scored significantly higher than those in the other group while completing a similar amount of practice. To further support this argument, we calculated a new measure, "learning efficiency," to quantify how much improvement a student makes for one question. To calculate learning efficiency, we used the Z score of students' final exam grades to account for the grade distribution and used the minimal value of the Z score to avoid negative learning efficiency. Because many previous studies have shown that a student's learning should be linearly correlated with the logarithmic value of the number of practices [40, 41], $\ln(\text{total practice})$ was used as the denominator. The learning efficiency was calculated using the following formula:

$$\begin{aligned} & \text{learning efficiency}(\text{student}_i) \\ &= \frac{ZScore_Final(\text{student}_i) - \min_{j=1..n} ZScore_Final(\text{student}_j)}{\ln(\text{total amount of practice}(\text{student}_i))} \end{aligned}$$

where $ZScore_Final(\text{student}_i)$ is the grade of final exam of the student_i

Students in the distributed practice group learned more efficiently (Mean=0.71, SD=0.170) than those in the massed practice group (Mean=0.56, SD=0.250), but the difference was only marginally significant based on an ANCOVA using midterm exam grades as the covariate ($p=0.051$, $F=3.946$).

4.4 Discussion of the data analysis

The results produced several interesting findings. In terms of the learning factors, we found that more practice did not lead to better learning outcomes. This seems

inconsistent with the well-known Learning Factors Analysis (LFA) [40] and the Performance Factor Analysis (PFA) [41], which quantifies a learning outcome using the number of correct attempts and incorrect attempts and claims that the learning outcome should increase with the number of practices. However, in contrast to the traditional setting of LFA and PFA, which provided more than enough practice questions, students in our experiment had a fixed number ($N=45$) of practice questions, but they could practice the same question as many times as they wanted. This resulted in a difference in the number of attempts by each student. This experimental setting provides us with an opportunity to observe how to better use a fixed amount of practice questions for learning. We divided students into two groups according to their practice frequency. Our analysis showed that students in the distributed practice group spent slightly more time completing a slightly lower number of practice questions and achieved a higher percentage of first check correctness. It implied that these students were likely more engaged when answering the practice questions. The students in the distributed practice group spent more time correcting their answers and were more constructive when posting comments. This was likely aided them in learning the required materials. Our finding is consistent with other previous studies which showed that having less to learn during each study session might contribute to an improved learning outcome [14] [42]. Unfortunately, many of the differences between student outcomes in our study were not significant. Therefore, we cannot make firm conclusions but only suggest possible reasons for the differences.

We randomly selected 6 students for face-to-face interviews to understand why some adopted a distributed practice strategy while others adopted a massed practice strategy. The 6 students were divided into 4 different groups: frequent system usage with a high final exam score, infrequent system usage with a high final exam score, frequent system usage with a low final exam score, and infrequent system usage with a low final exam score. All the students said that time was the most important factor that prevented them from completing the exercises daily. They were busy with many other classes. Completing MCQ practice was almost the only method they used to reinforce their learning. If they did have time, all but one student preferred to practice every other day. That student felt troubled by having to log into the system and answer questions. The explanation for why students did not want to practice every day was the cost of mental context switching [43]. They often needed to prepare themselves for answering the questions, especially for the questions containing programs. They said that it usually took 3 to 5 minutes to answer one question. Therefore, it was not worth answering just one question after having to spend a minute or more to prepare. In contrast, because answering one or two questions should take less than 10 minutes, it should be possible to encourage students to do so. How to encourage students in a similar setting is one of our next projects. Another explanation for why several students preferred massed practice was that they felt massed practice could lead to a higher number of correct answers. However, students in the distributed practice group earned a higher rate of correct answers on their first submission. Therefore, the beliefs of the students contradicted the truth. It seems that a Self-Regulation Learning tutor might help students use a better practice strategy, and therefore, lead to improved learning outcomes [44].

This study fixed the number of new questions per day at one without regard to question difficulty. On days when the question was too easy, students may have felt that answering was not worth logging into the system and completing the question. An adaptive selection method might be applied to provide students with an appropriate number of questions at an appropriate level of difficulty [45]. This type of adaptation might encourage students to adopt a distributed practice strategy.

We expected students to make more comments. However, they made only 11.97 comments per student, including topic-unrelated comments. This was likely because the comment function was not well introduced to and known of by the students. Therefore, in the future, we will encourage students to share their thoughts even when they have answered correctly. Thus, we can increase the power when determining the correlation between students' commenting behaviors and their learning outcomes.

5 Conclusion and Limitations

By setting the number of distinct multiple-choice questions and having students decide how to space their practice sessions, we found that students practiced with different frequencies. Dividing students into distributed and massed practice strategy groups based on their practice frequencies, we found that the two groups of students showed no significant difference for practice time and initial competences but showed a significant difference in learning outcomes. Therefore, different practice strategies on multiple choice questions did have an impact on students' learning outcome. Further analysis showed that students tended to answer more carefully when they had fewer questions to practice per session, which likely aided them in better internalizing the learning materials. As a result, the students in the distributed practice group achieved a significantly higher percentage of first check correctness and a higher ratio of topic-related comments that had a marginally significant difference than that of the students in the massed practice group.

However, the reader should be careful when drawing conclusions. Our study was not completed in a strictly controlled setting. Completing multiple choice questions was not the only method the students could use to practice. Therefore, there could be other factors, such as textbook reading habits and if they did their homework, that might have affected the students' learning outcomes. We assumed that students' learning strategies using our system were relatively stable throughout the semester. Therefore, the effect of the other study factors should be reflected in students' midterm exam scores, which were similar for the two groups of students. Their midterm exam scores were used as a covariate in the analysis to exclude the effect of other factors. Even so, readers must be aware of the above assumption used in this study.

Because our study showed that students tend to learn more when their practice sessions using multiple choice questions were relatively distributed, our next step is to require students to practice at prescribed frequencies and observe whether differences in practice frequency effect learning outcomes.

6 Acknowledgement

This work was supported by National Natural Science Foundation of China: [grant number 61807004].

7 References

- [1] S. Reardon and B. Tangney, "Smartphones, Studio-Based Learning, and Scaffolding: Helping Novices Learn to Program," *Acm Transactions on Computing Education*, vol. 14, pp. 1-15, 2014. <https://doi.org/10.1145/2677089>
- [2] S. Atmatzidou and S. Demetriadis, "Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences," *Robotics & Autonomous Systems*, vol. 75, pp. 661-670, 2016. <https://doi.org/10.1016/j.robot.2015.10.008>
- [3] M. J. Tsai, C. Y. Wang and P. F. Hsu, "Developing the Computer Programming Self-Efficacy Scale for Computer Literacy Education," *Journal of Educational Computing Research*, p. 073563311774674, 2018. <https://doi.org/10.1177/0735633117746747>
- [4] C. Watson and F. W. B. Li, "Failure rates in introductory programming revisited," 2014, pp. 39-44.
- [5] A. Yadin, "Reducing the dropout rate in an introductory programming course," *Acm Inroads*, vol. 2, pp. 71-76, 2011. <https://doi.org/10.1145/2038876.2038894>
- [6] T. C. Yang, G. J. Hwang, S. J. H. Yang, and G. H. Hwang, "A Two-Tier Test-based Approach to Improving Students' Computer-Programming Skills in a Web-Based Learning Environment," *Journal of Educational Technology & Society*, vol. 18, pp. 198-210, 2015.
- [7] N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer, "Distributed practice in verbal recall tasks: A review and quantitative synthesis.," *Psychological Bulletin*, vol. 132, pp. 354-80, 2006. <https://doi.org/10.1037/0033-2909.132.3.354>
- [8] R. Russo, A. J. Parkin, S. R. Taylor, and J. Wilks, "Revising current two-process accounts of spacing effects in memory.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 24, p. 161, 1998-01-01 1998. <https://doi.org/10.1037/0278-7393.24.1.161>
- [9] A. W. Melton, "Repetition and retrieval from memory," *Science*, vol. 158, p. 532-532, 1967-01-01 1967. <https://doi.org/10.1126/science.158.3800.532-b>
- [10] K. A. Rawson and W. Kintsch, "Rereading effects depend on time of test.," *Journal of Educational Psychology*, vol. 97, p. 70, 2005-01-01 2005. <https://doi.org/10.1037/0022-0663.97.1.70>
- [11] L. Budé, T. Imbos, M. W. V. D. Wiel, and M. P. Berger, "The effect of distributed practice on students' conceptual understanding of statistics," *Higher Education*, vol. 62, pp. 69-79, 2011. <https://doi.org/10.1007/s10734-010-9366-y>
- [12] M. Alzaid, D. Trivedi and I. H. Hsiao. The effects of bite-size distributed practices for programming novices. In 2017 IEEE Frontiers in Education Conference (FIE) (pp. 1-9). IEEE. <https://doi.org/10.1109/fie.2017.8190593>
- [13] E. A. Gamie, S. A. El-Seoud, M. Salama, and W. Hussein, "Multi-Dimensional Analysis to Predict Students' Grades in Higher Education," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 14, pp. 4--15, 2019. <https://doi.org/10.3991/ijet.v14i02.9905>

- [14] F. N. Dempster, "The spacing effect: A case study in the failure to apply the results of psychological research.," *American Psychologist*, vol. 43, pp. 627-634, 1988. <https://doi.org/10.1037/0003-066x.43.8.627>
- [15] L. Litman and L. Davachi, "Distributed learning enhances relational memory consolidation," *Learning & Memory*, vol. 15, pp. 711-716, 2008. <https://doi.org/10.1101/1m.1132008>
- [16] Y. Suzuki and R. Dekeyser, "Effects of distributed practice on the proceduralization of morphology," *Language Teaching Research*, 2015.
- [17] H. S. Sobel, N. J. Cepeda and I. V. Kapler, "Spacing effects in real-world classroom vocabulary learning," *Applied Cognitive Psychology*, vol. 25, pp. 763-767, 2011. <https://doi.org/10.1002/acp.1747>
- [18] R. B. G. W. Shimoni, "Addressing the Needs of Diverse Distributed Students.," *International Review of Research in Open & Distance Learning*, vol. 14, pp. 134-157, 2013.
- [19] E. Gerbier and T. C. Toppino, "The effect of distributed practice: Neuroscience, cognition, and education," *Trends in Neuroscience and Education*, vol. 4, pp. 49-59, 2015-01-01 2015. <https://doi.org/10.1016/j.tine.2015.01.001>
- [20] M. Garzia, G. R. Mangione, L. Longo, and M. C. Pettenati, "Spaced learning and innovative teaching: school time, pedagogy of attention and learning awareness," *Research on Education & Media*, vol. 8, 2016. <https://doi.org/10.1515/rem-2016-0004>
- [21] D. Rohrer, "Student Instruction Should Be Distributed Over Long Time Periods," *Educational Psychology Review*, vol. 27, pp. 635-643, 2015. <https://doi.org/10.1007/s10648-015-9332-4>
- [22] A. M. Glenberg, "Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms," *Journal of Verbal Learning & Verbal Behavior*, vol. 15, pp. 1-16, 1976. [https://doi.org/10.1016/s0022-5371\(76\)90002-5](https://doi.org/10.1016/s0022-5371(76)90002-5)
- [23] D. A. Balota, J. M. Duchek and R. Paullin, "Age-related differences in the impact of spacing, lag, and retention interval.," *Psychology & Aging*, vol. 4, p. 3, 1989. <https://doi.org/10.1037/0882-7974.4.1.3>
- [24] N. J. Cepeda, N. Coburn, D. Rohrer, J. T. Wixted, M. C. Mozer, and H. Pashler, "Optimizing distributed practice: Theoretical analysis and practical implications.," *Experimental Psychology*, vol. 56, pp. 236-246, 2009. <https://doi.org/10.1027/1618-3169.56.4.236>
- [25] A. Robins, J. Rountree and N. Rountree, "Learning and teaching programming: A review and discussion," *Computer Science Education*, vol. 13, pp. 137-172, 2003. <https://doi.org/10.1076/csed.13.2.137.14200>
- [26] J. Bennedsen and M. E. Caspersen, "Failure rates in introductory programming," *Acm Sigcse Bulletin*, vol. 39, pp. 32-36, 2007. <https://doi.org/10.1145/1272848.1272879>
- [27] S. H. Edwards and M. A. Perez-Quinones, "Web-CAT: automatically grading programming assignments: ACM, 2008. <https://doi.org/10.1145/1597849.1384371>
- [28] D. Jackson and M. Usher, "Grading student programs using ASSYST," *Acm Sigcse Bulletin*, vol. 29, pp. 335-339, 1997. <https://doi.org/10.1145/268085.268210>
- [29] A. Mitrovic, "An intelligent SQL tutor on the web," *International Journal of Artificial Intelligence in Education*, vol. 13, pp. 173-197, 2003-01-01 2003.
- [30] I. Hsiao, S. Sosnovsky and P. Brusilovsky, "Guiding students to the right questions: adaptive navigation support in an E-Learning system for Java programming," *Journal of Computer Assisted Learning*, vol. 26, pp. 270-283, 2010. <https://doi.org/10.1111/j.1365-2729.2010.00365.x>

- [31] A. Papancea, J. Spacco and D. Hovemeyer, "An open platform for managing short programming exercises," in International ACM Conference on International Computing Education Research, 2013, pp. 47-52. <https://doi.org/10.1145/2493394.2493401>
- [32] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, and B. Silverman, "Scratch: programming for all," *Communications of the ACM*, vol. 52, pp. 60-67, 2009-01-01 2009. <https://doi.org/10.1145/1592761.1592779>
- [33] F. Jakoš and D. Verber, "Learning Basic Programming Skills with Educational Games: A Case of Primary Schools in Slovenia," *Journal of Educational Computing Research*, vol. 55, 2016. <https://doi.org/10.1177/0735633116680219>
- [34] J. Spacco, P. Denny, B. Richards, D. Babcock, D. Hovemeyer, J. Moscola, and R. Duvall, "Analyzing Student Work Patterns Using Programming Exercise Data," pp. 18-23, 2015. <https://doi.org/10.1145/2676723.2677297>
- [35] J. Ohliati and B. Abbas, "Measuring Students Satisfaction in Using Learning Management System," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, pp. 180--189, 2019. <https://doi.org/10.3991/ijet.v14i04.9427>
- [36] M. T. H. Chi, S. Kang and D. L. Yaghmourian, "Why Students Learn More From Dialogue- Than Monologue-Videos: Analyses of Peer Interactions," *Journal of the Learning Sciences*, vol. 26, 2016. <https://doi.org/10.1080/10508406.2016.1204546>
- [37] M. Alamri, "Undergraduate Students' Perceptions toward Social Media Usage and Academic Performance: A Study from Saudi Arabia," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, pp. 61--79, 2019. <https://doi.org/10.3991/ijet.v14i03.9340>
- [38] J. A. Glover and A. J. Corkill, "Influence of paraphrased repetitions on the spacing effect.," *Journal of Educational Psychology*, vol. 79, pp. 198-199, 1987. <https://doi.org/10.1037/0022-0663.79.2.198>
- [39] M. T. H. Chi and R. Wylie, "The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes," *Educational Psychologist*, vol. 49, pp. 219-243, 2014. <https://doi.org/10.1080/00461520.2014.965823>
- [40] H. Cen, K. Koedinger and B. Junker, *Learning Factors Analysis – A General Method for Cognitive Model Evaluation and Improvement*: Springer Berlin Heidelberg, 2006.
- [41] P. I. Pavlik, H. Cen and K. R. Koedinger, "Performance Factors Analysis --A New Alternative to Knowledge Tracing," in *Conference on Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation To Affective Modelling*, 2009, pp. 531-538.
- [42] R. F. Hopkins, K. B. Lyle, J. L. Hieb, and P. A. S. Ralston, "Spaced Retrieval Practice Increases College Students' Short- and Long-Term Retention of Mathematics Knowledge," *Educational Psychology Review*, vol. 28, pp. 853-873, 2016. <https://doi.org/10.1007/s10648-015-9349-8>
- [43] P. D. Kieffaber and W. P. Hetrick, "Event-related potential correlates of task switching and switch costs," *Psychophysiology*, vol. 42, pp. 56--71, 2005. <https://doi.org/10.1111/j.1469-8986.2005.00262.x>
- [44] V. Aleven, I. Roll, B. M. McLaren, and K. R. Koedinger, "Help helps, but only so much: Research on help seeking with intelligent tutoring systems," *International Journal of Artificial Intelligence in Education*, vol. 26, pp. 205-223, 2016-01-01 2016. <https://doi.org/10.1007/s40593-015-0089-1>
- [45] L. Zhang and K. VanLehn, "Adaptively selecting biology questions generated from a semantic network," *Interactive Learning Environments*, vol. 25, pp. 828-846, 2017-01-01 2017. <https://doi.org/10.1080/10494820.2016.1190939>

8 Authors

Lishan Zhang is an associate professor at Central China Normal University. He received a Ph.D. in Computer Science from Arizona State University. He has published over 10 peer-reviewed academic papers. His research interests include intelligent tutoring systems, student modelling for personalized learning and educational data mining.

Baoping Li is a lecture of Faculty of Education, Beijing Normal University. Her research interests include smart learning environment design and assessment, teacher and students' perception on learning environments, integrating ICTs into teaching and learning activities, and technology enhanced science education.

Qiujie, Zhang is a master student in Faculty of Education, Beijing Normal University, China. Her research orientation is science education theory and practice

I-Han Hsiao is an assistant professor in the School of Computing, Informatics, and Decision Systems Engineering at Arizona State University. She received a Ph.D. in Information Science from University of Pittsburgh (2012). She completed her postdoctoral study in visual learning analytics in Teachers College, Columbia University. She has published over 50 peer-reviewed publications. Dr. Hsiao's research focuses on visual analytics, open student modeling and adaptive technology for personalized learning.

Article submitted 2019-03-28. Resubmitted 2020-01-16. Final acceptance 2020-01-18. Final version published as submitted by the authors.