Exploring the Impact of Voluntary Practice and Procrastination in an Introductory Programming Course

Jiayi Zhang
ejzhang7718@gmail.com
University of Pennsylvania
Philadelphia, PA, USA

Taylor Cunningham
taylorcu@sas.upenn.edu
University of Pennsylvania
Philadelphia, PA, USA

Rashmi Iyer
rashmii@seas.upenn.edu
University of Pennsylvania
Philadelphia, PA, USA

Ryan Baker
ryanshaunbaker@gmail.com
University of Pennsylvania
Philadelphia, PA, USA

Eric Fouh
efouh@cis.upenn.edu
University of Pennsylvania
Philadelphia, PA, USA

ABSTRACT
The effort to learn and the regulation of learning are key to successful learning. Voluntary practice has been shown to improve learning and is associated with having generally good self-regulated learning. At the same time, procrastination often slows the learning process and is associated with less than ideal regulation of learning. In this paper, we present the results of a study exploring the impact of voluntary practice and procrastination on the learning outcomes of novice programmers. We used data from an introductory programming course (CS1) at a large university and found that most students engaged in voluntary practice. However, students with higher prior performance and non-procrastinators were more likely to participate in the voluntary practice. We also found that participating in the voluntary practice did not have a significant impact on course performance. Furthermore, the study showed a weak negative correlation between procrastination and time spent on the homework and a weak negative correlation between procrastination and distributed practice. Finally, we found that non-procrastinators performed significantly better than procrastinators on the majority of homeworks.

CCS CONCEPTS
• Social and professional topics → CS1; • Applied computing → Computer-assisted instruction.

KEYWORDS
CS1, procrastination, voluntary practice, time-on-task, distributed practice

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGCSE 2022, March 3–5, 2022, Providence, RI, USA. ACM, New York, NY, USA, 6 pages.
https://doi.org/10.1145/3478431.3499350

1 INTRODUCTION
Successful learning in introductory programming courses involves using self-regulation strategies and making an effort to learn [2]. Voluntary practice is a strategy that demonstrates an effort to learn, while procrastination can signify less than optimal regulation of learning. To design effective interventions that promote better regulation of learning, computing educators should first understand the learning behaviors of novice programmers. The purpose of this study is to understand the impact of voluntary practice and procrastination on learning outcomes. We looked at the factors influencing participation in voluntary practice; in that regard, we investigated the relationship between prior performance and voluntary practice. We also explored how the difficulty of the assignments affects procrastination and how procrastination affects time-on-task and distributed practice. The study also examined the impact of procrastination and voluntary practice on learning outcomes.

Using data from a CS1 course, we found that higher prior performance and starting assignments early increased the likelihood of engaging in voluntary practice. However, when controlling for prior performance and procrastination, voluntary practice did not seem to impact students’ later performance. Our investigation also found that students procrastinated more on easier assignments and that procrastination was associated with lower/poorer time-on-task, distributed practice, and learning outcomes.

2 RESEARCH QUESTIONS
In this study, we are interested in better understanding the impact of voluntary practice and procrastination on the learning outcomes of novice programmers. In that regard, we will answer the following research questions (RQs):

RQ1: How do prior performance and procrastination influence a student’s likelihood to engage in voluntary practice?

RQ2: What is the impact of voluntary practice on learning outcomes?

RQ3: What is the impact of procrastination on time-on-task and distributed practice (on homeworks)

RQ4: What is the impact of procrastination on learning outcomes?
3 RELATED WORK

Previous studies have explored the impact of voluntary practice in CS1 courses. In [3], the authors used CodeWorkout an online drill-and-practice system in a CS1 course. Students were required to perform some but not all of the problems. The authors found that, voluntary practice contributed to improved performance on exams. In another study [4], an online tool was provided to students in a CS1 course as a supplemental practice resource. The tool contained exercises similar to the written exams only. The study revealed no significant difference in time-on-task between high and low performing students. However for the final exam, students who engaged in voluntary practice contributed to improved performance on exams. When looking at factors that influence students participation in voluntary practice, the study described in [11], found that more students with self-reported prior programming experience did not participate in voluntary practice. The same study also found no statistically significant difference between the performance of the midterm exam, of students who engaged in voluntary practice versus the ones who did not. However for the final exam, students who engaged in voluntary practice performed better (statistically significant at level $p = 0.012$) than the ones who did not.

When looking at the impact of procrastination in introductory programming courses, a previous study [9] showed that high performing and low performing students are harmed differently. High performing students are still able to complete the assignments when they procrastinate whereas low performing students are not able to get the help they need and might not complete the assignments. A recent study [8] looked at procrastination in a CS1 course and found that many students start work on assignments early. However, about 40% of those who started early worked on the assignment until the due day. They also found that, although student who started early performed better on average that those who did not, there was no difference in grade between them and students who started a day before the due date. Only students who started the assignment on the due date received a lower grade. Because of the potential negative impact of procrastination, several interventions were designed to curb its prevalence [7, 10].

4 PARTICIPANTS AND METHODS

Participants were students ($n = 301$) enrolled in the CS1 (in Java) offering at a large highly selective university in the USA. Students in the course did not have previous computing experience and were predominantly in their first semester, with their major undeclared. Students had to complete 9 programming homework assignments, and 2 timed exams (open notes, open internet). Students had an unlimited number of submissions for each homework and received immediate feedback with each submission. The first exam (midterm) was administered towards the middle of the term after homework 4, and the second (non-cumulative final) one at the end of the term after all of the homework assignments were completed. The midterm focused only on the programming topics covered in the first five homeworks. The second exam focused more acutely on the topics covered in homeworks 5 - 8. Students also had to complete 11 online quizzes with unlimited attempts and immediate feedback.

Students completed programming assignments in Codio$^1$. Codio is an online platform IDE functionalities, it also housed the lecture notes (textbook) for 5 of the topics covered in the course (recursion, references, abstract data types, arraylist, and linked lists). Static (PDF) slides were provided for the other topics covered in the course. The lecture notes are part of the OpenDSA project [6] an interactive electronic textbook and contain several learning activities. The textbook was provided to students on week 4 and was made available for the rest of the semester. The students were not required to complete the textbook’s problems for a grade but were encouraged to use it. We defined prior performance as the average grade of the students on the first four homeworks, before the lecture notes were released. We used the lecture notes to investigate voluntary practice among the participants.

We collected interaction data in Codio, and we took a snapshot of each student’s code after every period of inactivity (in Codio) lasting more than 10 minutes. The timestamp of the snapshot gives us a good indication of when the student took a break from working on the homework assignment. We used grades on homework, exams, and online quizzes, as a measure of student performance in the course. A procrastination measure was created to represent how many days a student waited to start working on an assignment after its release date we will refer to it as DSBD (days started before due). Procrastinators are students with DSBD less than the median DSBD for each homework (see Table 5 for more details). Of the 301 students who registered in the CS1 class during the fall 2020 semester, 282 students completed exam 1, exam 2, and at least 6 homework assignments out of the 9 assignments in total. Two students were further excluded due to missing temporal information on homework. 280 students were included in the following analyses.

4.1 Homework assignments

Homework assignments were assigned to students for 7 to 14 days. Since the difficulty of homeworks can change the behavior of a student, we ranked the homeworks by difficulty level. In order to be able to use standard Item Response Theory (IRT) models, we dichotomized homework grades. For each homework assignment, we assigned a score of 1 to a student if their grade was greater or equal to 93 (the letter A grade cutoff) and 0 otherwise. We used the R statistical software$^2$ (LTM package) to perform the IRT analysis.

We used the 2 parameter (2 PL) IRT model. It allows us to identify the difficulty, the discriminating factor of each homework (HW), and the probability for a student with an average ability to get a score of 1 on the assignment. The discrimination represents the degree to which the homework is able to recognize students of different abilities. The difficulty and discrimination coefficient are presented in Table 1.

The negative values of all but one difficulty parameter indicate that the problems were not extremely difficult. This is understandable given that the students had unlimited attempts and immediate feedback on the homeworks. HW8 was the most difficult homework and covered linked lists. We grouped the other homeworks as follows: HW1, HW6, HW2, HW7, and HW3 were moderately difficult (hardest to easiest), HW4, HW5 were easy, and HW0 was very easy.

\footnote{\url{https://www.codio.com/}}\footnote{\url{https://www.r-project.org/}}
For each homework, the table lists the difficulty, the discriminating factor, and the probability for a student with an average ability to get a score of 1 on the assignment.

When considering the discrimination parameter, the model showed that HW3, HW5, HW6, and HW4 had high discriminatory power, and HW2, HW1, HW8, HW7, and HW0 had moderate discrimination power. The low discrimination power of HW0 is explained by the fact that is an easy assignment (a variant of the "Hello world" program).

5 RQ1: HOW DO PRIOR PERFORMANCE AND PROCRASTINATION INFLUENCE A STUDENT’S LIKELIHOOD TO ENGAGE IN VOLUNTARY PRACTICE?

5.1 Analysis
To understand how prior performance and procrastination influence a student’s likelihood to participate in voluntary practice, we looked at students’ use of the course textbook. Therefore, in this analysis, we operationalized whether students participated in the voluntary practice as whether they have spent more than zero minutes on any modules of the textbook. Since homework grades are left-skewed, the prior performance scores were transformed to ranks.

5.2 Results
The majority (n = 248) of the students participated in voluntary practice (i.e., used the textbook). On average, students who participated in the voluntary practice ranked higher in previous performance and had started homework a day sooner than students who did not participate in the voluntary practice (Table 2). Logistic regression was used to estimate the impact of prior performance ranking and procrastination on determining which students participated in voluntary practice. Results from two simple logistic regression and a multiple logistic regression models are reported in Table 3. As shown in Table 3, students who had higher prior performance ranking (Model 1) or had started homework sooner (Model 2) were more likely to participate in the voluntary practice. The significance level of each factor attenuated in Model 3 when the two factors were examined together. The odds of participation increases slightly per one better rank in prior performance, and the odds decreases by 3% per each day procrastinated. Results found in this analysis identified two factors that could potentially explain why some students participated in the voluntary practice while others did not. Specifically, students who obtained better grades and

students who procrastinated less were more likely to participate in the voluntary practice. The two factors may reflect a phenomenon where students who were already doing well in class were more likely to voluntarily engage in additional practice.

6 RQ2: WHAT IS THE IMPACT OF VOLUNTARY PRACTICE ON LEARNING OUTCOMES?

6.1 Analysis
We examined the impact of voluntary practice on learning outcomes with the 280 students analyzed in RQ1. The number of chapters read and the percent correct on answering the questions in the textbook, were identified as additional measures of the effort involved in the voluntary practice. A composite score representing a student’s performance after the release of the textbook was calculated. The score contains students’ performance on the two exams and homework assignments 5 to 9 (excluding the first four assignments). The composite scores were then transformed into ranks due to skewness.

6.2 Results
Three multiple regression models were tested to understand the impact of the voluntary practice on learning outcomes. With prior performance and procrastination controlled, none of the variables measuring the use of the textbook or the effort of using the textbook statistically significantly predicted the composite score ranks (Table 4). Therefore, participating in the voluntary practice did not have a significant impact on the composite score rank. However further investigation is needed to understand how well the content in the voluntary practice aligns to the content that is being assessed in the homework assignments and exams, which constitute the
7 RQ3: WHAT IS THE IMPACT OF PROCRASTINATION ON TIME-ON-TASK AND DISTRIBUTED PRACTICE

7.1 Analysis

For each homework, we ran a linear regression analysis to evaluate the correlation for each of the following: DSBD vs homework difficulty, DSBD vs Time Spent, DSBD vs Number of Snapshots. We tested the significance of each correlation, and then used the Benjamini Hochberg procedure [1] to help control for false positives. Sample sizes across the homework assignments ranged from 269 to 280 based on the number of students who started and submitted each specific assignment.

7.2 Results

7.2.1 Procrastination and homework difficulty. We found a statistically significant (p = 0.018) correlation (r = 0.14) between the start date (procrastination) and the difficulty. HW0 was not included in the correlation because it is by far the easiest homework and students had 14 days to complete it. This indicates that students are able to correctly assess the difficulty of the homeworks and are able to adjust their starting date accordingly. They started harder homework earlier than the easier ones.

7.2.2 Procrastination and Time-on-task. We found a statistically significant positive correlation (at p = 0.05) between the DSBD and time-on-task, for all the homework assignments except HW4 (which had a marginally significant positive correlation). The significance of the results remained the same once the Benjamini Hochberg procedure was run. For all homeworks except HW4, this indicates that starting the homework assignments earlier is correlated with a greater time-on-task.

7.2.3 Procrastination and Distributed practice. We found a statistically significant positive correlation (at p = 0.05) between the DSBD and the number of snapshots for all of the nine homework assignments. For each assignment, this indicates that starting the homework earlier is correlated with a greater number of breaks while working on the assignment.

Based on our analysis, we can conclude that procrastination (starting the homework late) is correlated with less distributed practice. Starting the homework assignments earlier (no procrastination) is correlated with an increase in distributed practice.

8 RQ4: WHAT IS THE IMPACT OF PROCRASTINATION ON LEARNING OUTCOMES?

8.1 Analysis

Due to the relevance of certain homework assignments (and their covered topics) to each of the exams, we computed the following correlations: procrastination (DSBD) and homework performance, procrastination over Homeworks 0 through 4 and exam 1 performance, procrastination over Homeworks 5 through 8 and exam 2 performance, procrastination over all Homeworks and exam 2 performance, and procrastination over all Homeworks and class performance.

Since there is no way easy way to identify precisely when a student began preparing for an individual exam, we looked at the interactions with relevant homeworks as an indication to how much a student procrastinated on learning relevant topics. Sample sizes differed slightly, as not all students completed all assignments, and ranged from 275 to 280 across the homeworks and 265 to 275 across the correlations to exams performances.

8.2 Results

8.2.1 Procrastination and homework performance. We found a weak correlation between DSBD and homework grade. The correlation was statistically significant at p = 0.01 for all homework
assignments except Homework 0. However, non-procrastinators performed statistically better than procrastinators on all the homeworks except HW0, HW5, and HW8. The results are available in Table 5. It is not particularly surprising that there was no statistically significant difference for HW0 and HW8 given that they are the easiest and the hardest homeworks. Further work is required to better understand why HW5 grades were not different.

8.2.2 Procrastination and exams performance. We found a weak but statistically significant (at $p = 0.05$) correlation between average DSBD over the homeworks completed before the first exam (HW0 through HW4). This indicates that starting earlier on (early topics) homeworks correlated with better performance on the first exam. There was almost no correlation between starting later homeworks (5-8) early and exam 2 performance. Similarly, there was very little correlation between starting all the homeworks early and exam 2 performance.

<table>
<thead>
<tr>
<th></th>
<th>HW0</th>
<th>HW1</th>
<th>HW2</th>
<th>HW3</th>
<th>HW4</th>
<th>HW5</th>
<th>HW6</th>
<th>HW7</th>
<th>HW8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median, DSBD</td>
<td>8.37</td>
<td>2.99</td>
<td>3.13</td>
<td>3.14</td>
<td>2.86</td>
<td>2.85</td>
<td>3.90</td>
<td>2.15</td>
<td>6.05</td>
</tr>
<tr>
<td>Avg. Grade Procrastinators</td>
<td>99.21</td>
<td>87.14</td>
<td>89.17</td>
<td>86.77</td>
<td>94.51</td>
<td>95.02</td>
<td>81.14</td>
<td>79.60</td>
<td>73.35</td>
</tr>
<tr>
<td>Avg. Grade Non-Procrastinators</td>
<td>98.33</td>
<td>91.55</td>
<td>92.45</td>
<td>92.45</td>
<td>96.14</td>
<td>96.63</td>
<td>87.65</td>
<td>86.65</td>
<td>74.53</td>
</tr>
<tr>
<td>Mann-Whitney Test</td>
<td>$p = 0.43$</td>
<td>$p &lt; 0.01$</td>
<td>$p &lt; 0.05$</td>
<td>$p &lt; 0.01$</td>
<td>$p &lt; 0.05$</td>
<td>$p = 0.36$</td>
<td>$p &lt; 0.05$</td>
<td>$p &lt; 0.05$</td>
<td>$p = 0.10$</td>
</tr>
</tbody>
</table>

Table 5: Procrastination vs. non procrastinators homework grades

Further work is necessary to understand why this is the case. However, in our situation, we hypothesized that since only half of the topics covered in the class were available for voluntary practice, the learning benefits of voluntary practice were not large enough.

Students in our study are able to adjust their start date based on the homework difficulty. This result is quite significant since procrastination has been described as a “form of self-regulatory failure” [13]. Our results suggest that more effort should be devoted in studying the “incubation period” and its impact in CS education. The incubation period is defined as “a period of time in which a problem is set aside prior to further attempts to solve” [12]. A meta review found that divergent thinking tasks benefited the most from the incubation effect.

Our study found that procrastinators spent less time-on-task and had lower distributed practice. This result is not surprising given that procrastination reduces the amount of time students can work on an assignment. Moreover, procrastinators performed worse than non-procrastinators, and we found a weak negative correlation between procrastination and learning outcomes. Given that our study revealed that students were able to adjust their start date based on the homework difficulty, this result is in line with a previous study [8]; and indicates that many students still do not know how well to adjust their start date.

10 THREATS TO VALIDITY

We considered the students in our CS1 sample as a group. It is possible that our results might have differed based on why students were taking the course (though analyzing this properly would have required a larger sample). In our course, about 26% of students were enrolled in an engineering program, and only 11% of them had Computer Science as their declared major. The majority of students (60%) did not have a declared major. Following the course, less than half of the undeclared students declared an engineering major (including computer science). As discussed in [5], students who do not have programming or computer science as their primary major may have lower prioritization of course work as those whose primary major is programming. Another item to consider when generalizing our findings is the threshold that we used when assessing assignment difficulty. We use a cutoff grade of 93% (A letter grade) given the skewness of the grades. However, it is possible that choosing a different level could yield different results. The same applies to the inactivity period that we used when taking snapshot of students’ code. We used 10 minutes, however, it is not clear if the results will hold if we used a longer inactivity period. In general, the lack of qualitative data and the fact that only half of the topics covered in the class were available for voluntary practice were limitations to our current study, but are relevant areas for future work.
11 CONCLUSION AND FUTURE WORK

In this paper, we collected fine-grained programming data to study the impact of voluntary practice on procrastination in a CS1 course. We found that not procrastinating and having good grades on homework increased the likelihood of engaging in voluntary practice. However, we found that when controlling for prior performance and procrastination, voluntary practice did not have an impact on final grade (rank). We found that procrastination was negatively correlated with the difficulty of the homework, indicating that students could accurately assess the difficulty of the homework. However, many students did not adjust their start date well given that non-procrastinators outperformed procrastinators. Our study confirmed that procrastination negatively affected time-on-task and distributed practice.

The next steps in this study is to provide the students with interactive optional activities (textbook) that cover the entire curriculum. We will also collect qualitative and demographic data through surveys and interviews to account for students’ interest in computing and to get a better insight into the impact of the “incubation effect” in introductory programming courses.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. DUE-1946150.

REFERENCES