The Effects of Procrastination Interventions on Programming Project Success

Joshua Martin, Stephen H. Edwards, and Clifford A. Shaffer
Virginia Tech
2202 Kraft Drive
Blacksburg, VA 24060
+1 540-231-5723
jdm522@vt.edu, edwards@cs.vt.edu, shaffer@vt.edu

ABSTRACT

In computer science, procrastination and related problems with managing programming projects are viewed as primary causes of student attrition. Unfortunately, the most successful techniques for reducing procrastination (such as courses in study skills) are resource-intensive and do not scale to large classrooms. In this paper, we describe three course interventions that are designed to be scalable for large classrooms and require few resources to implement. *Reflective writing assignments* require students to consciously consider how their time management choices impact their classroom performance. *Schedule sheets* force students to actively plan out the time required to solve a programming project. *Email alerts* inform students of their progress relative to their peers as they work on an assignment, and suggest ways to improve behavior if their progress is found to be unsatisfactory. We implemented these interventions in a junior-level data structures course and analyzed data from 330 students over two semesters. Separate analyses of reflective writing responses, schedule sheet contents, and e-mail alert contents are discussed, along with student opinions about the value and effectiveness of each treatment. We found a statistically significant relationship between the time when work is completed and its quality, with late work being of lower quality. We found that one of the three interventions had a statistically significant effect on reducing late work: e-mail alerts sent to students to make them more aware of how they were doing with respect to expectations were associated with both a reduction in assignments completed late, and an increase in assignments completed at least one day early. This result was found despite the fact that students reported subjectively that e-mail alerts were of marginal utility.

General Terms

Education

ICER'15, August 09–13, 2015, Omaha, NE, USA

ACM 978-1-4503-3630-7/15/08 ...$15.00.
DOI: http://dx.doi.org/10.1145/2787622.2787730.

Keywords

programming, procrastination, self-regulation, self-efficacy, scheduling, schedule sheets, e-mail alerts, reflective writing

1. INTRODUCTION

It is a common problem in programming-intensive courses that students fail to complete their programming projects in a timely manner. A common hypothesis among CS educators is that such students may have the necessary skills, but lack the time management skills or the commitment needed to complete individual project assignments. Procrastination and poor choices are often used as excuses. 70-95% of undergraduates procrastinate on coursework to some degree, while 20-30% exhibit chronic or severe procrastination [10]. By procrastination, we mean "to voluntarily delay an intended course of action despite expecting to be worse off for the delay" [10]. "Negative procrastinators" are those who procrastinate to the extent that they actually do experience negative consequences from their delays.

In STEM disciplines that involve project-based learning activities, students may be at greater risk when they procrastinate. When a student has two or more weeks to complete a project or paper—one that presumably will take more than a single afternoon to finish—it is certainly easier to procrastinate, because the deadline is farther in the future. However, it is also more dangerous, since putting off the work both reduces the available time, should the project take more effort than the student expects, and also reduces the opportunities available to seek assistance or ask questions, should unexpected difficulties arise. In our own courses, we see that typically a quarter to a third of students are unable to satisfactorily complete any given multi-week programming project.

In a previous study of 1,101 CS students over a period of five years [4], we looked at students who sometimes performed well on work and sometimes performed poorly, and used a within-subjects comparison to look at the differences in when they started their work. We found a statistically significant correlation between when students start working on a project and the quality of their work: when a student starts earlier, he or she is significantly more likely to earn an A or B on work than if the work is started later.

Many techniques have been proposed to combat procrastination. The most successful techniques seem to be supplementary courses or workshops on time management strategies [12]. While this has been shown to be effective, they are costly in terms of time and manpower. An ideal mechanism to reduce procrastination must be feasible at a larger scale and applicable to courses with hundreds of students.

In this paper, we examine three classroom interventions to reduce procrastination. These interventions are designed to require little additional manpower or class time, and so can be used in large courses. These interventions included active reflection writing tasks, schedule sheets, and situational awareness alerts that describe student performance relative to expectations. We examined how these treatments affected the times at which students started submitting work to an automated grading system, as well as when students finished their assignments. While two of the interventions did not provide evidence of significant impact, the e-mail alert intervention did show a significant increase in the number of assignments submitted early and a significant decrease in the number of assignments submitted late, in comparison to the control condition. Further, by examining relationships between on-time vs. late work and quality of student work, this study re-confirms that late assignments score lower and typically contain more bugs, as measured by instructor reference tests. In addition, work completed early, ahead of the deadline, scores significantly higher.

This work extends a preliminary examination of this same experiment presented at ITiCSE 2015 [3], where only the main treatment effect was examined. The contributions of this paper include a more thorough examination of the impact of the treatments after filtering out students who dropped or failed to complete the course, along with an examination of the times when students start submitting work for assessment, rather than only their finish times. Also, separate analyses of the data collected on each individual treatment are provided, along with the results of student opinion surveys indicating student perceptions of the value, effectiveness, and time required for each of the interventions. While the Tuckman procrastination scale [11] was used to measure procrastination tendency among subjects in this study, no significant relationship between procrastination scale scores and assignment submission times was found. Finally, this paper compares the quality of student work with its time of completion to assess whether late work was of measurably lower quality in this study.

2. BACKGROUND

While procrastination is a pervasive problem throughout education, there is still a lack of understanding about the phenomenon. Perhaps the best summary of procrastination research so far comes from Steel, who published a meta-analysis of procrastination research in 2007 [10]. Steel defined procrastination as a "prevalent and pernicious form of self-regulatory failure." Some research indicates that procrastination may be an individual personality trait, and several instruments have been developed to measure this tendency [10, 6, 11]. We used Tuckman's procrastination scale instrument [11] to measure the procrastination tendency of students in our study, described in Section 5.

Several potential causes of procrastination have been proposed. However, procrastination is primarily a failure of self-regulation. In a study of 456 undergraduates, Klassen et al. [5] found that a person's view of their own ability to self-regulate was a strong predictor of procrastination. Tuckman theorized that an inability to overcome procrastination tendencies might be related to the gradual transfer of responsibility from teachers and parents to individual students that occurs throughout the school years. Because this transfer of responsibility reaches its peak during the college years, he theorized that researchers should examine techniques that can assist students in the regulation of their own learning [11]. Such techniques should include providing information to students so that they are aware of the appropriate progress needed to successfully complete a task.

Steel [10] has proposed temporal motivation theory (TMT) for modeling procrastination. TMT incorporates four factors to account for the desirability of a task: expectancy of success (E), value of the task to the individual (V), the delay before one is rewarded for the task (D), and the individual's sensitivity towards that delay (Γ). Utility is defined as $(E \times V)/(\Gamma \times D)$. This theory influences the interventions we designed.

3. INTERVENTIONS

The focus of our study is three interventions that we made, with the goal of reducing procrastination or otherwise improving the performance of students on projects in a junior-level Data Structures and Algorithms course. This course is taken typically one year after a traditional CS2 course, and is typically about the fourth programming course that a student encounters. Each of the interventions studied is relatively easy to administer (assuming availability of the infrastructure that we built to support some of these), and can scale to large courses. In total, there were four course sections involved, over two years. One section was the control (no explicit interventions targeted toward procrastination were administered, and this course was taught in a manner similar to prior years), and each of the interventions was administered to one section of the course. We first describe the three interventions, and then we discuss our experiences and analyze the results. See [3] for details on the specific instruments used in each intervention.

The key activity studied was the semester programming projects. Students in all sections had a similar experience in that they were required to implement four projects during the semester. Each project had a life cycle of approximately one month, from the time when the initial specification for the project was made available (and discussion of the project was initiated in class) until the assignment due date. In total, the projects accounted for 45% of the semester grade. However, the projects are even more important than this figure would indicate, as scores on projects explain about 95% of the variance on total semester score. Thus, good performance on the projects is crucial to a successful grade. The projects are generally considered by the students to be quite challenging, involving interactions between typically two to four major classes (in the object-oriented sense), requirements for student-generated unit tests, meaningful design choices by the students (that affect project scoring), and complex programming skills such as advanced recursion and dynamic memory allocation. Success on the projects requires project management skills such as time management along with skills in software design, testing, and debugging. A typical project might require 30–80 hours, resulting in approximately 500-2,000 lines of code, including software tests but excluding comments.

3.1 Active Reflection

The first intervention we examined is *reflective writing assignments*. These assignments were inspired by active learning techniques, specifically the technique called the "minute paper" [9][2]. The goal of these writing assignments was to engage students in reflection about their own time management behavior and how it affects their individual performance on the programming projects. The initial concept of writing a single response was expanded to four responses after consulting with the course instructors. In our study, these responses were completed using an on-line form near the beginning of each project, asking the student to reflect on the impact of their time management choices on their previous project experience. The activity was designed so that students could complete it in under 15–20 minutes. Students in the targeted section were required to do the reflections, with each one being worth approximately 1% of the semester grade.

3.2 Schedule Sheets

The second intervention examined is the use of *schedule sheets*. In prior years, instructors for thise course have used "painless" schedule sheets [8]. Student survey data indicate a mixed response to the schedule sheets, with some students finding them useful while others find them unhelpful. The goal behind the schedules is to encourage students to break a large project assignment into smaller, more manageable pieces. The sheets are also designed to have students consider their progress on an assignment periodically. This intervention aims to reduce procrastination by helping students form, express, manage, and track smaller-scale deadlines.

To effectively manage the schedule sheets for students, we designed and implemented an electronic system to handle the submission and grading of these sheets. Students entered or changed their work breakdown structure as a series of tasks. Often these tasks consisted of specific classes, program behavior, or modules a particular assignment required. Each task had subfields for the estimated design time, coding time, and testing time, as well as a personal target deadline for when the student anticipated completing the entire task.

Students were required to fill out or edit a schedule sheet three times during a project. Collectively, the three schedule sheets for a given project constituted about 1% of the semester grade (for a total of 4% for schedule sheets over the four projects). For each project, the first sheet was an initial schedule that was due within a week of receiving the assignment. The second sheet was an intermediate schedule due one week before the assignment was due, allowing students to update their progress and modify their own task deadlines as necessary. A final schedule was due after the project was completed, with students reporting the actual amount of time spent on each project task.

To ensure student schedules were reasonable, the system provided automatic feedback as schedule information was entered by the students. A "check my work" button allowed students to get immediate feedback at any point while editing their schedule. The work check mechanism included multiple diagnostics used to verify that a schedule was appropriate for the particular project based on the number of anticipated components, the time estimates made by the student, and the personal task deadlines set by the student. Additionally, the system allowed an instructor to manually review the submitted schedules for any additional discrepancies, to provide their own feedback comments, or to adjust scores where necessary.

3.3 E-mail Alerts

The third intervention examined is the use of automated e-mail *situational awareness alerts*. We developed a mechanism to send periodic e-mail alerts to students throughout the time allotted to work on a project. These alerts were designed to raise awareness of a student's level of effort compared to his or her peers, and compared to expectations.

Instructors will often inform students that working early and often on a particular assignment will yield a higher score, but this information often may be ignored. This intervention is designed to take a different approach by providing individualized information that is more relevant to a student. In particular, the system is integrated into Web-CAT, an automated grading tool used at our institution. Because students were required to submit their work to Web-CAT for evaluation, we could provide feedback on a student as they worked towards a project solution, and include data extracted from their current work to produce more informative messages.

The e-mails sent by this intervention began roughly one week before the project was due. Second and third e-mails were sent at 4 days and 2 days before the assignment due date, respectively. The content of each alert was customized to reflect the work submitted so far to Web-CAT. The student's work was classified along 4 dimensions: the amount of code written (relative to an approximate target size for the given assignment), the proportion of instructor written reference tests passed (an approximation of functional correctness), the degree of testing performed (if the assignment requires students to write their own software tests), and the number of static analysis checks failed (measuring adherence to coding style guidelines, if required by the assignment). Based on the scores for each of the dimensions, each student's work would be given an internal grade: Good (indicating advanced progress compared to the ideal rate of progression), Neutral (indicating average progress compared to the ideal rate), Bad (indicating poor progress compared to the ideal rate), and Undefined (indicating no work has yet been submitted).

Based on these internal ratings, a customized e-mail was constructed and sent to students. The subject line of the e-mail was phrased as "CS 3114: Your progress on Project 2". However, if the student's work indicated insufficient progress in one or more dimensions, the subject line would instead be "CS 3114: You may be at risk on Project 2", or even "CS 3114: You are at risk on Project 2". The body of the e-mail message contained a separate paragraph corresponding to each of the 4 dimensions on which the student's work was rated. Messages were phrased in an attempt to recognize progress and reinforce good practices without being judgmental or negative.

4. EXPERIENCES WITH THE INTERVENTIONS

The three interventions were employed in four sections of CS 3114: Data Structures and Algorithms, a junior-level data structures course at our university. Two of the sections were taught in Fall 2013, with the other two in Fall 2014, all by the same instructor (an author on this paper and Co-PI on the project). Three of the sections each received one of the three interventions, and one section acted as a control. The course involved four separate programming assignments, with students being given approximately one month for each assignment. Each section received the same intervention across all four assignments.

The following subsections describe our experiences employing these interventions. In addition, we gave an opinion survey at the end of each semester asking students for their reactions to the intervention they experienced, in terms of how useful they felt it was, whether they felt it took too much time, and whether they believe it affected the way they managed their time on projects. The assignment consisted of writing one-paragraph responses to four prompts: describing the key elements of the plan they used to manage their time on their most recent project, describing how that affected the quality of their work, describing the plan they intend to use on the new project, and describing their development strategy and how it impacted their results.

4.1 Active Reflection

4.1.1 Response Themes

Many students emphasized starting early as a way to avoid stress and turn in quality work. In fact, on the first reflective writing assignment, 85% of respondents indicated they would use a strategy of submitting earlier than previously. However, analyzing the responses from the next sheet, we found the number of students who reported starting earlier on the project reduced to 68%. This trend did not improve with Project 3, were students who reported working early dropped to around 46%. This indicates students underestimated their own ability to start a project early (as self defined), and while the reflective writing assignments forced students to consider their time management choices, it did little to improve their self-reported performance.

Besides starting early, many students emphasized a lack of team work as a key element to finishing on-time. One student wrote: “After deciding to work with a friend, we broke up sections of the project to focus on. By partially splitting up the project, we were able to maximize our development over time.” Another student lamented a lack of planning, writing “I began working on my last project way too late. The key elements was that my partner and I waited for a mutually agreeable time to work on the project. We waited too long for a perfect time and ended up starting very late.”

Reasons for starting late varied. Some students complained about a lack of knowledge, while others blamed the amount of additional coursework they had. A few students had difficulties working with teammates, which led to a later start than they assumed was ideal.

Overall, the responses reveal that most students know the correct way to avoid procrastination, but fail to follow through—reconfirming a failure of self-regulation. A variety of reasons could be to blame for this, including some that are not the fault of the student. The reflective writing assignments made students consider how to schedule their time, but did not seem to change their actual behavior.

4.1.2 Survey Results

At the end of the semester, students were given a brief on-line opinion survey on their experiences with the intervention consisting of nine Likert-style questions that were answered on a five-point scale. Questions covered whether the intervention helped the student manage projects better, made them think critically about time management, made them more aware of how much they procrastinate, or caused them to change the way they managed their time. Questions also covered whether students felt the intervention was a waste of time, or took too much time to complete.

Overall, student responses to the reflective writing treatment were more positive than for the other interventions. 50% of respondents indicated that the intervention helped them manage their projects better, giving a response of 4 or 5. Only 23% agreed that the reflective writing assignments were a waste of time, although 64% agreed that they took too much time to complete (the highest of any intervention). Finally, when students were asked if the intervention caused them to start at least one assignment sooner, 68% of students either agreed or strongly agreed, the highest of any intervention. This seems to indicate that students valued the experience of reflective writing.

4.2 Schedule Sheet Data

4.2.1 Schedule Data

While individual projects differed in complexity, the average across all students for the student-reported estimates of total project time on the first schedule sheet ranged from 33–54 hours (33 hours for Project 1, s.d. 15.0, 53.6 hours for Project 2, s.d. 17.7, 41 hours for Project 3, s.d. 16.6, and 39.5 hours for Project 4, s.d. 10.8). These initial estimates were usually underestimates, however, with the final reported time spent by students averaging 43–79 hours (66.4 hours for Project 1, s.d. 34.9, 79 hours for Project 2, s.d. 33.0, 43.9 hours for Project 3, s.d. 15.5, and 47.7 hours for Project 4, s.d. 21.3). Individual students underestimated the effort required 72% of the time. Figure 1 summarizes the distribution of initial time estimates made by students, their revised estimates on their intermediate schedule sheets, and the final self-reported time spent (all values are student-reported on their own schedule sheets).

On the intermediate time sheets due one week before the project was due, students reported having spent an average of about one third of the total time they would report by the end of the project. One would expect that estimates of time remaining at this point would be more accurate than

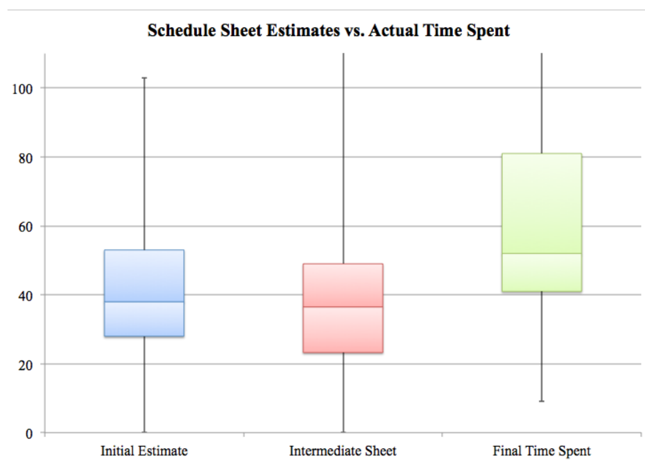

Figure 1: Schedule sheet estimates on the initial schedule and intermediate schedule, compared to total time reported across all sheets.

the initial estimates. However, the opposite appeared to be the case, with most students underestimating the time remaining–in fact, on average, students estimated it would take less total time by the intermediate time sheet than their original (under-)estimate before work began. Figure 1 shows the distribution of estimates across all three sheets, illustrating this underestimation trend.

We examined the accuracy of the final schedule sheet and its potential impact on the time students submitted their final work to Web-CAT using a two-way analysis of variance. We found a significant relationship between these factors (F = 17.79, p <0.001). Students who underestimated the amount of work required for a project submitted their final solutions later, while students who overestimated the amount of work required submitted their final solutions earlier.

We hoped that schedule accuracy would improve as the semester went on, indicating that students were improving their own time management skills. We did find that the assignment itself had a significant relationship with schedule accuracy (F = 10.41, p <0.001), where students had the largest average underestimates for Project 1, followed by Project 2, with the smallest underestimates for the two later projects. Because the assignments differed in level of effort and complexity, one would expect differences in the ability of students to estimate the level of effort required, however.

4.2.2 Survey Results

From the survey of student opinions, student responses to the schedule sheet intervention were less positive. Only 21% of respondents agreed or strongly agreed that schedule sheets helped them manage their projects better, with 56% disagreeing or strongly disagreeing. Additionally, 62% of respondents agreed or strongly agreed that the intervention was a waste of time (the highest of any intervention), although only 29% agreed that they took too much time to complete. Finally, only 44% of respondents agreed or strongly agreed that schedule sheets caused them to start their next assignment earlier.

4.3 E-mail Alert Data

4.3.1 Email Analysis

For the students who received e-mail alerts, students usually received three alerts, seven days before the assignment was due, then again four days before the due date, and finally one last alert two days before the due date. For two of the assignments, we added a fourth alert ten days ahead of the due date as well.

For each e-mail alert, the student's work so far, as characterized by their most recent submission to Web-CAT, was analyzed, and rated on four different aspects. If students had not made any submission to Web-CAT yet at the time of the alert, they received a message specifically addressing that fact, and including encouraging language about the benefit of starting early.

When alerts were sent ten days ahead of the due date, only 2–7% of students had made a submission to the electronic grading system. Thus, the bulk of the students received a "form letter" alert because there was no data available to use for a more personalized message. For alerts one week ahead of the due date, 6–18% of students had made a submission. Four days ahead, 17–33% of students had made at least one submission, and two days ahead, the number had grown to 43–54%.

Because of the low percentages of students who had submitted work, in almost all cases, the majority–sometimes the vast majority–of students simply received an alert indicating they had not yet submitted any work, and reminding them that starting earlier is associated with better success on project assignments.

We examined more closely the alert status of students 4 days ahead of the due date, roughly in the middle of the series of alerts. A two-way analysis of variance indicated a significant difference (F = 12.9, p <0.001) between the project grades earned by students who had made at least one submission to Web-CAT at that point (mean of 80.7%, s.d. 24.6%, not including any extra credit incentives for early completion) and those who had not yet made any submission (mean of 69.2%, s.d. 29.3%).

4.3.2 Survey Results

In the survey of student opinions, student responses to the e-mail alert intervention were somewhat negative. Only 14% of respondents agreed that the e-mail alerts helped them manage their own projects better, with no students at all strongly agreeing. Additionally, 55% of respondents agreed or strongly agreed the e-mail alerts were a waste of time, and 37% agreed or strongly agreed they took too much time. Finally, only 24% of respondents agreed or strongly agreed that the e-mail alerts caused them to start their next assignment earlier, the lowest of any of the interventions.

5. EVALUATION

The primary purpose of this study was to determine if any of the interventions positively affected the timeliness of student work. In particular, we hypothesized that treatment groups would be less likely to turn work in late, and correspondingly more likely to turn work in on time, or even early. Because prior research indicates that work completed late often earns lower scores in multiple dimensions [4], secondarily we also wanted to confirm this link was also present for this study.

Our study involved a total of 330 students in four sections over a two-semester period who agreed to allow access to their assignment data, all enrolled in our university's junior-level data structures course. Of these students, 82 (24.8%) either dropped the course, withdrew from the course, or did not complete all programming assignments. We excluded these students from our analysis, so the data reported in this section are based on students who completed all programming assignments and received a grade for the course. Among the remaining students, the size of each treatment group were similar (control N = 60, reflective writing N = 64, schedule sheets N = 59, and e-mail alerts N = 65).

All four groups completed four programming projects of varying difficulty, with Projects 1 and 3 being somewhat easier or smaller, and Projects 2 and 4 being more involved. All projects focused on implementing complete programs built on data structures implemented as part of that assignment. Sections offered at the same time—the control and reflective writing sections in Fall 2013, and the e-mail alert and schedule sheet sections in Fall 2014—used identical assignments. However, between the two semesters the assignments were changed, although both semesters used assignments that were intended to be comparable in terms of level of effort. The same instructor taught all four sections.

In all programming projects for all sections, students were offered a 10% extra credit bonus for completing their assignment at least one day ahead of the due date. Students did not receive any grade penalty for turning in work late, however. Instead, a "time bank" of individual 1-day extensions was permitted, similar to the model described in [1]. Each student had a small, fixed number of these no-penalty grace days they could use, but late submissions were no longer accepted once a student had expended all of their late days. Students in Fall 2013 were allotted five late days, but this number was reduced to three in Fall 2014.

5.1 Procrastination Tendency

We used Tuckman's procrastination scale [11] to measure the individual procrastination tendency for students in all groups at the start of the course. This instrument has been independently validated as a procrastination measure. The instrument consists of 16 questions answered on a 4-point Likert scale regarding one's perceptions about putting off required tasks versus starting them when necessary. Aggregating answers across all questions produces scores ranging from 16-64, with higher scores indicating a greater tendency to procrastinate. After giving the instrument questions to students using an electronic survey, aggregate scores were compiled and then normalized to a 0–1 scale.

We compared scores across groups to determine whether course sections were of similar procrastination tendencies. A one-way analysis of variance did not indicate any significant differences between the groups ($F = 1.6$, $p = 0.20$), with average scores of 60% for the control group, 58% for the reflective writing group, 64% for the schedule sheet group, and 59% for the e-mail alert group. Also, we did not find any significant relationship between scores on the procrastination scale and the times when students finished work ($F = 0.77$, $p = 0.38$).

5.2 Treatment Impact on Submission Time

Because we wished to determine if any of the interventions positively affected the timeliness of student work, we classified each final submission for each assignment as being either early (finished at least one day before the deadline to earn extra credit), on time (finished on the due date), or late (finished after the due date). Figure 2 shows the relative proportion of student work falling into each category for each treatment group.

Of the groups, the e-mail treatment group had the lowest number of late submissions and also the highest number of early submissions, making that group's on-time performance significantly different than the control group ($\chi^2 = 10.05$, $p = 0.0015$). The other two treatments did not differ from the control group in a statistically significant way (reflective writing: $\chi^2 = 0.03$, $p = 0.87$; schedule sheets: $\chi^2 = 0.52$, $p = 0.47$).

In addition, we also looked at the finish times as a continuous variable, in terms of the difference between the deadline and the time when students completed their projects. The results of a two-way analysis of variance ($F = 3.81$, $p < 0.01$) followed by Tukey's HSD, the students in the e-mail alert group turned in their assignments significantly earlier than those in both the control and reflective writing groups (but not significantly earlier than the schedule sheets group), while other groups were not significantly different from each other. Figure 3 shows the mean finish times for each group. The pooled standard deviation across all groups was 2.2 days, so the difference between the finish times of the e-mail alert group and the control group represents an effect size of 0.25 (schedule sheets: 0.18 (not significant), and reflective writing: 0.03).

We also wanted to check whether students *started* any earlier. Because we only had data from students once they began submitting their work electronically to Web-CAT, we do not have direct access to their start times. However, previous work suggests that student behaviors for starting work [7] and for beginning to submit to an electronic grading system follow similar patterns [4]. As a result, we examined the time of each student's first submission to Web-CAT.

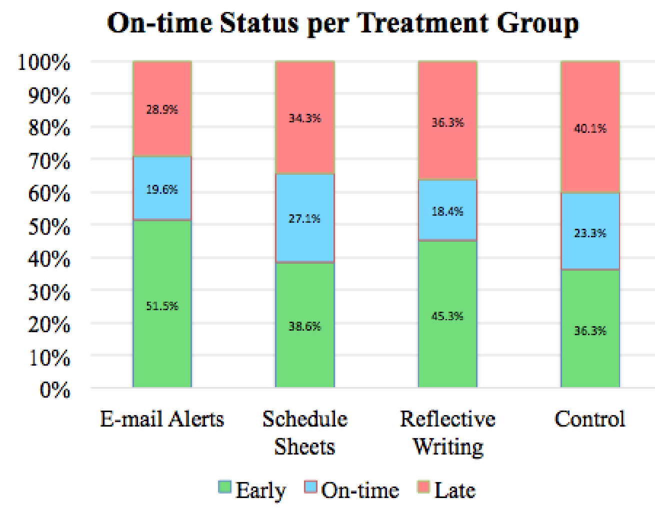

Figure 2: The percentage of early, on time, and late assignments turned in per treatment group.

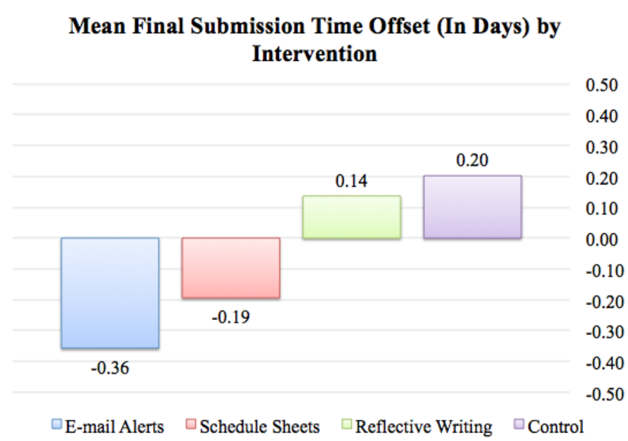

Figure 3: The average project submission time relative to the due date per treatment group. A negative value indicates the mean submission time was before the deadline, while a positive value indicates the mean submission time was late.

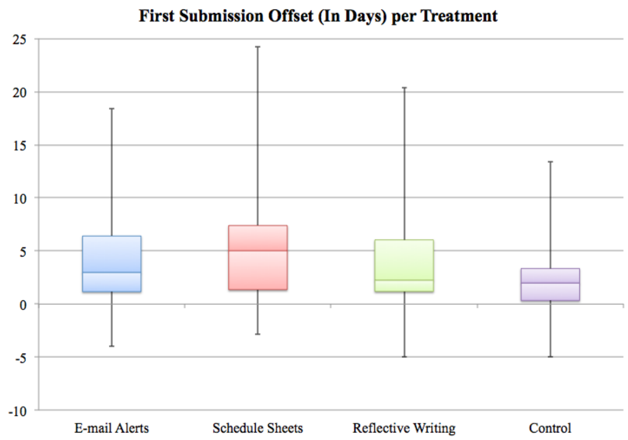

Figure 4: The first submission timestamp offset per treatment group. Note that a larger value indicates an earlier mean first submission time, while a negative value indicates a late first submission.

Figure 4 shows the relationship of first submission times among the groups. An analysis of variance (F = 24.7, p <0.001) followed by Tukey's HSD indicated that the schedule sheet group has significantly earlier first submission times on average (mean 5.3 days, s.d. 4.3) than all other groups. Both the e-mail alerts group (mean 4.0 days, s.d. 3.9) and the reflective writing group (mean 3.4 days, s.d. 3.7) were also significantly earlier than the control (mean 2.4 days, s.d. 3.1). Similarly, students in the schedule sheet group had a larger spread of time between their first submission and final submission (mean 5.1 days, s.d. 3.8) compared to the other groups (e-mail alerts: mean 3.6 days, s.d. 3.4; reflective writing: mean 3.6 days, s.d. 4.1; control: mean 2.6 days, s.d. 3.1). Again the schedule sheets group was significantly different than the other three (F = 19.9, p <0.001), with both e-mail alerts and reflective writing being significantly different than the control, but not each other.

One possible explanation for the earlier start times by the schedule sheets group is that students in that group were required to make an initial submission to Web-CAT before or with their intermediate schedule sheet, which was due one week before the deadline. This may have artificially altered the time of the first submission. One possibility is that students in the schedule sheet group simply turned in their work to meet the requirement, without it being "authentic" in the sense of having reached the point where they believed checking their work with the automated grader was needed. To explore this possibility, we also looked at the sizes of student programs on their first submissions, ignoring comments and blank lines, relative to their final product. Students in the schedule sheet group were the only group to submit significantly different amounts of code in their first submissions compared to the control group (F = 3.83, p <0.01). Students in the schedule sheet group made initial submissions containing an average of 78.6% of their final amount of code (s.d. 28.1%), which was less than the control group (89.2%, s.d. 35.6%). Neither the e-mail alerts group (mean 85.6%, s.d. 30.6%) nor the reflective writing group (mean 84.3%, s.d. 35.0%) were statistically different from any other groups. Based on this, it is plausible that the earlier times of first submission for the schedule sheets group do not consistently represent "starting earlier" in this study, since this group's first submissions were smaller, and were not significantly associated with finishing earlier than other groups.

Finally, we also analyzed the use of time bank days by students—the individual 1-day extensions allowed, as described at the start of Section 5. Analyzing the number of bank days used across all four projects (normalized to account for the different number of available bank days across semesters), there was no significant relationship between treatment and the proportion of bank days used (F = 1.4, p = 0.25, mean proportion of days used, control: 55%, reflective writing: 42%, schedule sheets: 57%, e-mail alerts: 52%). This seems to indicate that no treatment had an impact on a student's tendency to use these days. There also was no significant difference in proportion of bank days used between semesters. That might indicate that the number of actual bank days available is was not important in this study.

5.3 Impact of Lateness on Quality

Previous work indicates that a student's programs tend to score lower and behave less correctly when that student finishes work late, after the due date, compared to when that same student finishes earlier [4]. To analyze the impact of lateness on project quality, we examined two key indicators. First, we considered the grade (total score) received by students on each project, without including any extra credit bonus received for early completion. In addition to the automated score produced by Web-CAT, this score also includes the manual grading performed by course teaching assistants. Second, we also considered the percentage of reference tests passed by each final submission, which is a measure of the functional correctness achieved, independent of other aspects of the grade.

Figure 5 shows the relationship between early, on time, or late submission and project grade. Completing projects early was significantly associated with earning higher scores (t = 4.28, p <0.001), although there was no significant difference between on time and late submissions (t = 1.21, p = 0.23). When considering functional correctness, however, the differences are even more apparent. Figure 6 shows

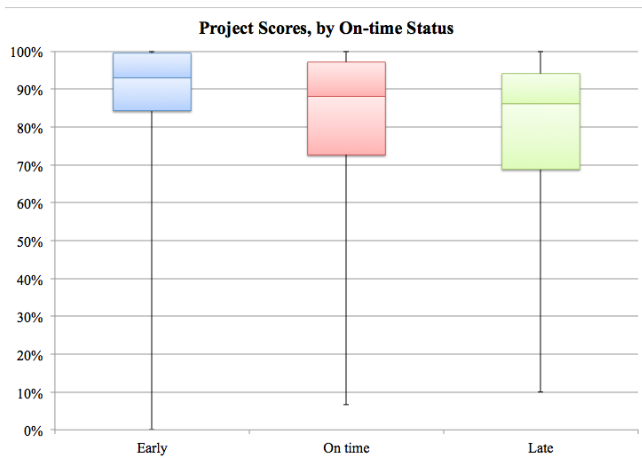

Figure 5: Project score distributions for early, on time, and late submissions.

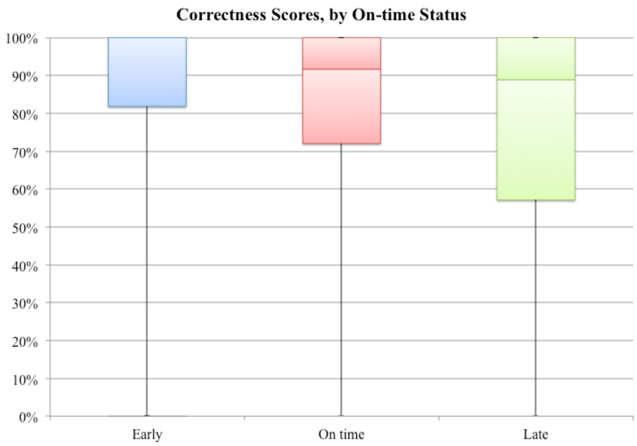

Figure 6: Correctness percentages for early, on time, and late submissions, in terms of instructor reference tests passed.

the relationship between the correctness scores for programs completed at different times, in terms of the percentage of instructor-provided reference tests passed by the program. Here, on-time work was significantly more correct (t = 2.43, p = 0.015) than late work, and early work was significantly more correct than on-time work (t = 3.87, p <0.001).

While these results support the notion that late work is of lower quality, one possibility is that "stronger students" simply finish their work earlier, and that weaker students, who are less able to complete an assignment, finish later. To account for this, we examined our data set to identify students who turned in at least one of the four projects late, and at least one of the four projects on time or early. We then performed a within-subjects comparison of both project scores and of reference test pass rates, comparing each student's late work against that same student's non-late work. In this within-subjects comparison, the on-time status of the student's final submission was a significant factor in his or her overall project grade (F = 7.5, p <0.001), with earlier work scoring higher. Similarly, the on-time status of the student's final submission was also a significant factor in the percentage of reference tests passed by his or her program–that is, its functional correctness (F = 13.7, p <0.001). As in [4], this within-subjects comparison suggests that finishing on time or early is related to project success independent of any individual-specific traits.

6. CONCLUSION AND FUTURE WORK

This paper summarizes experiences with three new classroom interventions designed to reduce student procrastination. While two of the interventions did not provide significant evidence of impact, the e-mail alert intervention showed some promise in reducing the frequency of late submissions and increasing the frequency of early submissions. At the same time, however, students subjectively reported that they did not feel the e-mail alert intervention affected how they managed their time, and generally reported that they believed these alerts were a waste of time. Nevertheless, from the point of view of our interventions, our results indicate the e-mail alert intervention was somewhat effective in the most obvious measures of earlier start times and earlier finish times, and may potentially be more effective if we improve them and base them on more accurate data, perhaps by directly collecting student data from their IDE instead of only from their grading system submissions.

At the same time, this study confirmed earlier findings that late submissions have lower quality, receiving lower scores on functional correctness, while early submissions earn higher scores overall. Further, a within-subjects comparison indicates that individual students see higher scores on the work they complete early, compared to the work they complete late. Clearly, finishing earlier is associated with higher scores.

Based on these results, we plan to continue exploring the e-mail alert intervention described here, with the aim of providing more accurate and directed feedback that students may find more useful. It is possible that by collecting activity data directly from the student's development environment as they work on assignments, it will be possible to gain fine-grained insights into the entire period of development rather than simply the snapshots when the student turns in work for checking. This additional information could lead to more valuable information in e-mail alerts, aimed at increasing awareness of progress against the rest of the group and against expectations.

Ultimately, all of these interventions will only be effective if they can be used in a large scale classroom. While we have demonstrated such scalability in this paper, our implementations definitely are still prototypes. Ideally, we will eventually create a system that can accurately monitor student progress based on years of previous student effort on similar assignments, and provide feedback to instructors based on at-risk students as needed.

7. ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation under grant DUE-1245334. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

8. REFERENCES

[1] J. Aycock and J. Uhl. Choice in the classroom. *SIGCSE Bull.*, 37(4):84–88, Dec. 2005.

[2] C. C. Bonwell and J. A. Eison. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports*. ERIC, 1991.

[3] S. H. Edwards, J. Martin, and C. A. Shaffer. Examining classroom interventions to reduce procrastination. In *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education*, ITiCSE '15, pages 254–259, New York, NY, USA, 2015. ACM.

[4] S. H. Edwards, J. Snyder, M. A. Pérez-Quiñones, A. Allevato, D. Kim, and B. Tretola. Comparing effective and ineffective behaviors of student programmers. In *Proceedings of the fifth International Workshop on Computing Education Research*, pages 3–14. ACM, 2009.

[5] R. M. Klassen, L. L. Krawchuk, and S. Rajani. Academic procrastination of undergraduates: Low self-efficacy to self-regulate predicts higher levels of

procrastination. *Contemporary Educational Psychology*, 33(4):915–931, 2008.

[6] C. H. Lay. At last, my research article on procrastination. *Journal of Research in Personality*, 20(4):474–495, 1986.

[7] C. Norris, F. Barry, J. B. Fenwick Jr., K. Reid, and J. Rountree. Clockit: collecting quantitative data on how beginning software developers really work. In *Proceedings of the 13th Annual Conference on Innovation and Technology in Computer Science Education*, ITiCSE '08, pages 37–41, New York, NY, USA, 2008. ACM.

[8] J. Spolsky. Painless software schedules. `www.joelonsoftware.com/articles/fog0000000245.html`, 2000.

[9] D. R. Stead. A review of the one-minute paper. *Active Learning in Higher Education*, 6(2):118–131, 2005.

[10] P. Steel. The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological bulletin*, 133(1):65, 2007.

[11] B. W. Tuckman. The development and concurrent validity of the procrastination scale. *Educational and Psychological Measurement*, 51(2):473–480, 1991.

[12] B. W. Tuckman. Relations of academic procrastination, rationalizations, and performance in a web course with deadlines 1. *Psychological Reports*, 96(3c):1015–1021, 2005.