More Carrot or Less Stick: Organically Improving Student Time Management With Practice Tasks and Gamified Assignments

Mac Malone
tydeu@cs.unc.edu
University of North Carolina
Chapel Hill, NC, United States

Fabian Monrose
fabian@ece.gatech.edu
Georgia Institute of Technology
Atlanta, GA, United States

ABSTRACT

Students often struggle with time management. They delay work on assignments for too long and/or allocate too little time for the tasks given to them. This negatively impacts their performance, increases stress, and even leads some to switch majors. As such, there is a wealth of previous research on improving student time management through direct intervention. In particular, there is a heavy focus on having students start assignments earlier and spend more time-on-task – as these metrics have been shown to positively correlate with student performance. In this paper, however, we theorize that poor student time management (at least in CS) is often due to confounding factors – such as academic stress – and not a missing skill set. We demonstrate that changes in assignment design and style can cause students to organically manage their time better. Specifically, we compare two alternative designs – a low risk preparatory assignment and a highly engaging gamified assignment – against a conventional programming assignment. While the conventional assignment follows common trends, students do better on the alternative designs and exhibit novel behavior on the usual metrics of earliness of work and time-on-task. Of note, on the preparatory assignment, time-on-task is *negatively* (albeit weakly) correlated with performance – the opposite of what is standard in the literature. Finally, we provide takeaways and recommendations for other instructors to use in their own approaches and research.

CCS CONCEPTS

• **Social and professional topics → Computing education**; • **Applied computing → Education**; • **Security and privacy**;

KEYWORDS

Time management, Earliness, Time-on-task, Prerequisite testing, Gamification, Engagement, Stress, Learning analytics

ACM Reference Format:
Mac Malone and Fabian Monrose. 2023. More Carrot or Less Stick: Organically Improving Student Time Management With Practice Tasks and Gamified Assignments. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2023), July 8–12, 2023, Turku, Finland*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3587102.3588825

ITiCSE 2023, July 8–12, 2023, Turku, Finland

ACM ISBN 979-8-4007-0138-2/23/07…$15.00
https://doi.org/10.1145/3587102.3588825

1 INTRODUCTION

As instructors, we love seeing students learn and grow and are disheartened when circumstances conspire to impede their success. As researchers, we are interested in analyzing both the highest highs and the lowest lows to see how we can further develop our craft to produce more of the good and less of the bad. In this spirit, we believed that gamification could help better motivate and engage students in the educational process and, to this end, developed our own in-house gamified learning platform to teach students. We designed it initially to motivate students to learn cybersecurity techniques by hacking a real game, but it subsequently grew into a fully-featured learning platform. Of note, the latest iteration included a secure automated grading system for standard coding exercises. This system regularly logs student activity, providing us an opportunity to learn more about student behavioral patterns.

Students distribute work across an assignment period in different ways. Previous research suggests how they do so can have a major impact on their performance. If the impact is too negative, it can even lead to students dropping out of computer science altogether [1, 22]. There are many aspects to this connection – student ability, time management strategies, environmental concerns, cognitive factors, *etc.* – and each can produce different results upon analysis [8, 9, 15, 27]. In this paper, we focus on the impact of student time management, since our grader logs provide us a detailed picture of student activity and can thus provide valuable insight.

For time management, there are two major factors that research suggests have a significant effect on student performance: how early a student works on an assignment (earliness) and how long they spend on it (time-on-task) [13, 14]. However, in our intervention, we discovered that student behavior can vary greatly across different styles of assignments. While earliness and time-on-task may predict student performance on a conventional assignment, a preparatory or gamified assignment can produce very different results. In fact, under certain circumstances, the positive correlation between time-on-task and performance can even be inverted! All in all, our results suggest student time management can be organically improved by either increasing engagement with gamification (more carrot) or lowering risk with practice tasks (less stick).

Our Contributions. We analyze the popular metrics of earliness and time-on-task in our own intervention and examine their impact on student performance. In doing so, we discover a number of caveats with the conventional wisdom on the topic. We demonstrate that a low risk preparatory assignment with practice tasks graded as a bonus and a highly engaging gamified assignment graded normally can organically improve student time management and performance. Furthermore, on the preparatory assignment, time-on-task was *negatively* (albeit insignificantly) correlated with performance, challenging the strong positive correlation generally found in the literature [14]. In contrast, a conventional assignment is observed to follow common trends, highlighting the significance of assignment design. We suggest one key difference between assignments is the amount of academic stress produced and finish with recommendations other instructors can apply in their own courses.

2 RELATED WORK

There is a plethora of research on student time management both in computer science and outside it. This includes detailed reports on observed behaviors and their effects [8, 15, 23, 25] along with interventions designed to improve areas of concern [2, 6, 11, 12, 19]. In our system, we pay particular attention to the metrics of earliness and time-on-task, so works with a special focus on those are the most germane to our interests.

Work by Leinonen *et al.* [13] best addresses the metric of earliness. It summarizes the relevant research and highlights the general consensus that students who work earlier perform better. The authors then test this theory in a course of their own. While they observe a positive correlation between earliness and grade, they note that even among those who started a day before the deadline, the most common grade was 5 (the highest). Thus, they argue that a simple correlation is likely an inadequate explanation. We build upon this work by demonstrating more anomalous results and providing a possible explanation.

For time-on-task, the current state of the art is also by Leinonen *et al.* [14]. In it, they compare coarse- and fine-grained metrics of time-on-task and their relationship with student performance. They measure coarse-grained time-on-task using the time elapsed from first keystroke to first submission and then remove all breaks between keystrokes of 10 minutes or more to obtain their fine-grained time-on-task metric. Fine-grained time-on-task is shown to be universally better than coarse-grained at predicting performance, and they use it to predict final course grades with 93% accuracy (via a random forest classifier). Like them, we also account for breaks in our time-on-task metric, though ours are coarser due to a difference in activity metric (we did not track activity to the level of keystrokes). In contrast, we find time-on-task to be a frail metric that does not always correspond well to performance, which highlights potential caveats with the state-of-the-art understanding.

3 METHODOLOGY

The context for our analysis is a Fall 2021 iteration of the Introduction to Computer Security course at UNC Chapel Hill (59 students – 51 males, 8 females). During that semester, there were three distinct styles of assignments: preparatory assignments, highly gamified assignments, and conventional assignments. Students were assigned a grade of S, A, B, C, or F on each assessment, where an S signifies completion of bonus tasks above and beyond those needed for an A. All student work took place within our overarching gamified learning platform, Riposte [18]. We received IRB approval and student consent to use the collected data for research purposes [16].

We use this data to test previous work on the topic of time management and analyze whether our three different styles of assignments change anything. Additionally, we noted that students who start too late run out of time, spending less time on the task and achieving a lower grade than they otherwise would. We theorized this could lead to a correlation between time-on-task and performance simply as a by-product of a correlation between earliness and performance. That is, students who start late will necessarily spend less time and therefore perform worse (*i.e.,* decreased earliness, time-on-task, and grade) whereas those who started early can spend more and perform better (*i.e.,* increased earliness, time-on-task, and grade). This produced three research questions:

(RQ1) Do students who work on assignments earlier and/or spend more time-on-task perform better?
(RQ2) Does assignment design effect student behavior?
(RQ3) Are earliness and time-on-task interrelated?

Assignment Differences

For our analysis, we chose one preparatory, one gamified, and one conventional assignment out of all those in the course to analyze. We pick them because they best represented their category.

Preparatory assignments serve as low-risk assessments of necessary prerequisites. Missing prerequisites can cause a student to perform badly on an assignment even if they completed the learning objective. For example, a student may have trouble communicating over a WebSocket and this may inhibit them from completing an online password guessing assignment even if they learned how to generate good password candidates. Preparatory assignments help mitigate this by highlighting what areas students need to review to prepare for subsequent normal assignments. In structure, our preparatory assignments resemble more the practice task model of Denny et al. [5] rather than the simplified assessments of Edwards *et al.*'s syntax exercises [7] or Parsons Problems [21]. Grades on preparatory assignments do not impact the student's overall course grade. Instead, receiving an A (or S) will guarantee the student a minium of a C- on the subsequent normally graded assignment.

Gamified assignments are completely unlike preparatory and conventional assignments – they are centered around completing challenges through cheat-based gameplay rather than finishing normal programming exercises. For example, in the one we analyze, students hack a 2D action web game by modifying its code using browser developer tools. These hacks augment their game character's ability, allowing them to overcome previously impossible obstacles within the game to earn trophies and improve their standing on the learning platform's leaderboard. Students receive a conventional grade based on the number of trophies they earn. Students also use their hacks in a post-assignment tournament to see who developed the most game-breaking cheats.

Finally, our conventional assignments are standard programming assignments with test-suite-based automatic grading. However, they are also semi-structured and exploratory, reflecting our field of cybersecurity. For example, the one we analyze here covers offline password guessing. Students perform trial-and-error production of password candidates while also optimizing their code so that it does not use too much memory or take too long to complete. Thus, it is worth keeping in mind that some of the differences we observe between the preparatory and conventional assignment may be due to the difference in complexity between the two.

Assessing Earliness

There are multiple ways to measure the earliness of a student's work. One way is to simply measure the start of student's activity on the assignment (e.g., in our case, first log entry or first submission) [8, 11, 13]. This can be problematic. Students can take a quick look at the assignment, maybe even solve some problems they find simple, submit, and not come back until the assignment is almost due. Similarly, another way is to analyze the final time of submission [8, 11, 19]. If students can submit multiple times, this can also be problematic, as students who started early and completed most of the assignment may return late to add a few finishing touches. This is an acute concern in our scenario, since each assignment has extra tasks a student can complete to earn bragging rights (*e.g.,* a higher place on a leaderboard) and are thus incentivize to work on the assignment past the time needed to perform well.

A distinctly different way to measure earliness is to count how many days a student worked on the assignment. This connects with the notion of the *spacing effect*, which suggests students distributing work over multiple days improves performance [13, 27, 28]. Students who start very late will necessarily spend less days on an assignment. Nevertheless, research has shown students who start very early and work many days can still perform worse, and students who work only a single day can still do very well [13].

Therefore, to create a good balance between earliness and workload, we measure earliness by recording the offset from the deadline of an assignment's relevant event (code submission or game connection) and compute the mean over all such events. A similar approach is also used by Rao [23]. We term this metric the *mean submission time* or *mean connection time* depending on the event used. We use two different events because students do not submit anything on the gamified assignment. Instead, task completion is monitored in-game and grades are automatically assigned when the relevant requirements are met. Students complete tasks by leveraging modifications they make to the game client. To apply these modifications, students need to refresh their browser and establish a new game connection. Therefore, like each new submission, each new connection serves as a test of a student's code, making it a great parallel to submission on the gamified assignment.

Measuring Time-on-Task

We measure time-on-task through the proxy metric of total interaction time. Interaction time is the time difference between log entries of the grading system. Total interaction time is the sum of all such time differences, but with a caveat: to account for breaks, we exclude interaction times above a certain threshold (the break threshold) when summing them – an approach also taken by previous work [14]. A longer threshold has a higher chance of false negatives (including shorter breaks as long work sessions) while a shorter threshold has a higher chance of false positives (excluding long work sessions as breaks). From experimenting with different thresholds, we observed that the break threshold setting has a major impact on the resulting analysis.

To determine an appropriate threshold, we compared the distribution of total interaction times for different thresholds with the students' self-reported time-on-task from their responses on a post-assignment survey. Since our objective time-on-task metric only measures the time the students were actively coding and not time they may have spent thinking, researching, or otherwise focused on the assignment, we expect the measured time-on-task to be less than the self-reported time-on-task. Considering this, we chose a break threshold of 30 minutes, as this resulted in interaction times that best approximated the self-reported time-on-task without overshooting it (see Figure 1). Our break threshold is higher than previous work [14] because we have a coarser metric for activity (log entries versus keystrokes).

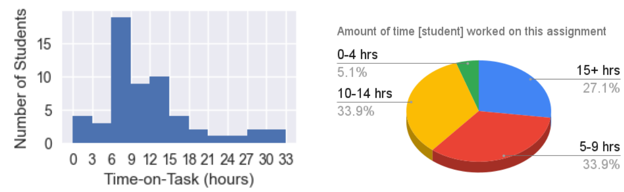


Figure 1: A comparison of the observed time-on-task of students for the conventional assignment with a 30 minute break threshold (left) versus the self-reported time-on-task on the survey (right).

4 DATA ANALYSIS

For RQ1, we wish to assess whether our observations align with previous research on time management. Thus, following common wisdom [8, 13, 23], we hypothesize that *a significant number of students will work on their assignments close to the deadline* **(Hypothesis 1a)**, and we expect that *students who work close to the deadline tend to perform worse* **(Hypothesis 1b)**. In particular, due to the danger of running out of time, *students who work in the last few days perform the worse* **(Hypothesis 1c)**. Also, for time-on-task, we expect *time-on-task will have a significant positive effect on student performance* **(Hypothesis 1d)**.

For RQ2, we theorize that academic stress is a leading cause of poor time management. It has been shown that gamification can increase engagement [4] and thereby reduce stress [3]. Similarly, we expect the preparatory assignment's low risk nature to also reduce stress. Thus, we hypothesize that *student time management on the gamified assignment will align more closely with the preparatory assignment than with the conventional assignment* **(Hypothesis 2)**.

Finally, for RQ3, we expect *earliness will have a significant impact on time-on-task if and only if many students delay their work too late* **(Hypothesis 3)**, since students who work too late will run out of time, reducing their time-on-task.

Earliness. If students submit primarily submit close to the deadline, the distribution of submission times will be exponential (*i.e.,* many times close to the deadline of 0 and few out on the tail). Thus, to see if they did so (Hypothesis 1a), we analyze whether the mean submission times fit such a distribution. Since we provided in-class time to work on the preparatory assignment when it was handed out, we removed this time period from the data used in our analyses.

We applied an Anderson-Darling statistical test to determine whether we can reject the null hypothesis that the mean submission times come from an exponential distribution. We used an A-D test for this because, of the three standard goodness of fit tests (Kolmogorov-Smirnov, Anderson-Darling, and Cramer-von-Mises), the A-D test best handles distributions which may have fat tails (*e.g.,* exponential distributions). We found that, for the preparatory assignment, we could reject the A-D hypothesis at a significance level of 1% (A^2 = 2.56 > critical value = 1.94). For the conventional assignment, we could not reject the hypothesis even at a significance level of 15% (A^2 = 0.71 < critical value = 0.91). This indicates that **students submitted the conventional assignment close to the deadline, while they submitted earlier on the preparatory assignment,** providing evidence for Hypothesis 1a on the conventional assignment but not the preparatory assignment.

To see if students' submission behavior affected their assignment grade (Hypothesis 1b), we computed the correlation coefficient (two-sided Pearson's r) between their mean submission time and their final grade on the assignment. We found a statistically significant positive correlation for the conventional assignment ($r = 0.44$, $p = 0.0007$), but a weaker and less significant correlation for the preparatory assignment ($r = 0.40$, $p = 0.002$). Additionally, students worked earlier (by the A-D test) and performed significantly better (by Welch's t-test: $t = 6.58, p < 0.0001$) on the preparatory assignment than on the conventional assignment – the mean grade improved from a B to an A. This supports Hypothesis 1b and suggests that **the earlier students work, the better they tend to do, but with diminishing returns** (since improving on an A is not that useful). Figure 2 helps illustrate all this.

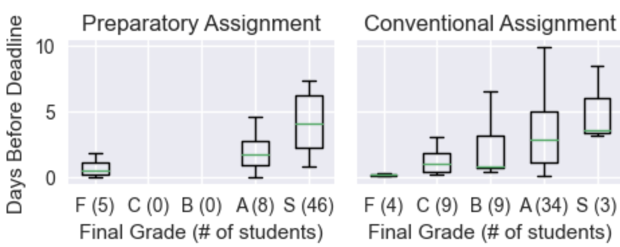


Figure 2: A box plot of students' mean submission times categorized by their final grade on the assignment.

To determine whether submitting in the last few days was particularly significant (Hypothesis 1c), we split the students in two based on their mean submission time and a threshold (12, 24, 36, 48, 60, and 72 hours before the deadline) then computed the correlation coefficient (two-sided Pearson's r) between mean submission time and grade of the two subsets for each threshold. We also performed a two-sided Welch's t-test of two independent samples (which, unlike Student's t-test has no assumption of equal variance) between the two subsets to test whether the average grade differed significantly.

For the preparatory assignment, there was no significant difference between the subsets at any threshold. This aligns with the observation that students were not frequently submitting during the last few days of the preparatory assignment and with its overall weak correlation between mean submission time and grade. In contrast, for the conventional assignment, we found that a split on 48 or 60 hours led to similar sized subsets that differed significantly. 28 students had a mean within the 48 hours and 29 did not (and vice versa for 60) and there was a statistically significant ($t_{48} = 4.20$, $t_{60} = 4.05$, $p < 0.001$) difference in performance between the subsets but an insignificant, weak correlation ($r < 0.15$, $p > 0.50$) between mean submission time and performance within them.

In other words, students who mostly submitted within two days of the conventional assignment's deadline performed worse (had a median grade of B) whereas those who submitted earlier did better (had a median grade of A). Futhermore, within each group, how early or late one submitted was not statistically significant. This provides evidence for Hypothesis 1c and supports our theory that **many students started the conventional assignment too late and ran out of time, hurting their grade. Conversely, students worked earlier on the preparatory assignment, giving them more than enough time to do well on the assessment.**

Gamification. To analyze whether gamified assignment was more like the preparatory or the conventional (Hypothesis 2), we compared them. Figure 3 plots the mean connection time as a histogram and as a boxplot categorized by grade, paralleling Figure 4b & 2.

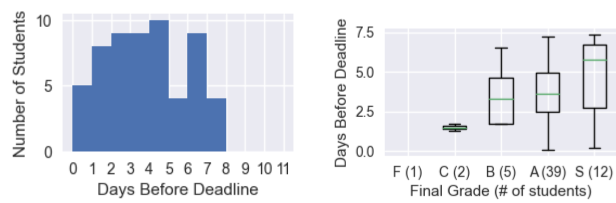


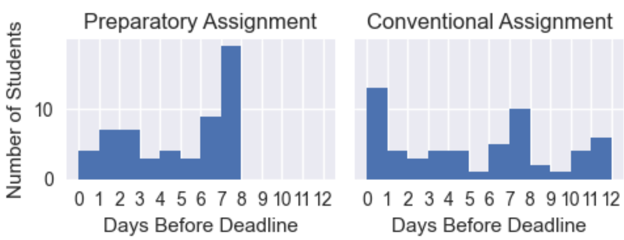
Figure 3: The mean time the students connected to the game on the gamified assignment as a histogram (left) and as a boxplot categorized by grade (right). The weekend (Saturday/Sunday) is days 3-5.

The data, as expected, is observed to more closely align with that of the preparatory assignment than the conventional assignment. Specifically, the mean submission time across students was 3.6 days on the preparatory assignment and 2.9 days on the conventional assignment, whereas the mean connection time across students was 3.8 days on the gamified assignment (i.e., similar to the preparatory). Also, like the preparatory assignment, the distribution across days is not exponential even at a 15% significance level by an A-D test (A^2 = 5.00 > critical value = 1.94).

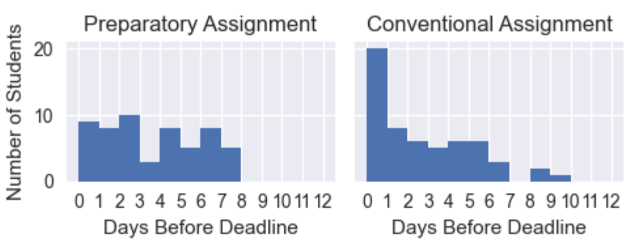
We also computed the correlation coefficient (two-sided Pearson's r) between the mean connection time and the grade on the gamified assignment. The correlation ($r = 0.27$, $p = 0.039$) was much weaker than the correlation between mean submission time and grade on both the preparatory assignment ($r = 0.40$, $p = 0.002$) and the conventional assignment ($r = 0.44$, $p = 0.0007$).

All in all, student time management on the gamified assignment was more like the preparatory assignment than the conventional assignment even though it was normally graded. This supports Hypothesis 2 and indicates that the low risk nature of the preparatory assignment is not necessary to reproduce the improvement in student time management seen there. In fact, a highly gamified yet normally graded assignment can produce even stronger results. This provides evidence for our theory that academic stress is the cause of students' poor time management and that **assignments with high engagement (more carrot) or low risk (less stick) can mitigate stress and thereby improve time management.**

Does Metric Matter? With some results in hand, we wished to test whether our choice of metric for submission time was a good one, as there were many other alternatives – first submission time [8, 11, 13], last submission time [8, 11, 19], number of days worked [13, 27, 28], etc. To begin, we compared the distribution of first submission times (the most popular metric of earliness) with that of mean submission times times (our metric). See Figure 4.



(a) First Submission Time



(b) Mean Submission Time

Figure 4: Histograms of submissions times for the preparatory and conventional assignment. Assignments are distributed about a week (7-12 days) before the deadline. The weekend (Saturday/Sunday) is days 4-6 on the preparatory and 2-4 on the conventional.

The first submission times suggested most everyone worked early on the preparatory assignment. The mean submission times, however, showed that workloads were more uniformly distributed across the period. Similarly, while first submission times suggested that students' earliness was relatively uniformly distributed across the conventional assignment, the mean submission time indicated that most of their work was crammed into the the final few days.

Next, we computed the correlation coefficients (two-sided Pearson's r) between the popular metrics of earliness and the student's grade and compared them to results we obtained for mean submission time. On the conventional assignment, where the correlation for mean submission time was strong and statistically significant ($r = 0.44$, $p = 0.0007$), the correlations for first submission time ($r = 0.40$, $p = 0.002$), last submission time ($r = 0.29$, $p = 0.03$), and days worked ($r = 0.31$, $p = 0.02$) were all weaker but still significant. The preparatory assignment was similar. The correlation for mean submission time was the strongest ($r = 0.40$, $p = 0.003$), and the correlations for first submission time ($r = 0.30$, $p = 0.03$) and last submission time ($r = 0.31$, $p = 0.02$) were weaker. Days worked even had a statistically insignificant *negative* correlation ($r = -0.18$, $p = 0.19$). Finally, on the gamified assignment, where mean submission time had a weak and barely significant correlation ($r = 0.25$, $p = 0.065$), the correlations for first submission time ($r = 0.09$, $p = 0.52$), last submission time ($r = 0.14$, $p = 0.31$), and days worked $r = 0.20$, $p = 0.12$) were all weaker and statistically insignificant. In summary, **the metric of mean submission time was a stronger predictor of student performance than others common in the literature on all assignment styles**. To our knowledge, this is a new result, as few others use this metric and those that did (*e.g.,* Rao [23]), did not compare it with others.

Time-on-Task. To analyze time-on-task, we computed the correlation coefficient (two-sided Pearson's r) between a student's total interaction time and grade. A positive correlation for total interaction time (*i.e.,* more time-on-task, higher grade) would support **Hypothesis 1d** and follow previous research (*e.g.,* Leinonen *et al.* [14]). As expected, we find a statistically significant positive correlation on the conventional assignment ($r = 0.36, p = 0.007$). However, on the preparatory assignment, we find a statistically insignificant and slightly *negative* correlation ($r = -0.12, p = 0.37$), which is quite extraordinary. Figure 5 helps illustrate the difference.

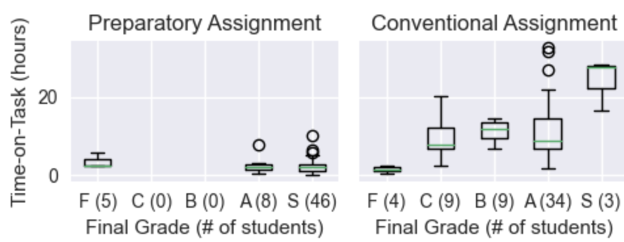


Figure 5: A box plot of students' time-on-task categorized by their final grade on the assignment.

While our results are contrary to common wisdom [14], our theory that earliness and time-on-task are interrelated (RQ3) may help explain them. To verify that earliness impacts time-on-task, we computed the correlation coefficient (two-sided Pearson's r) between mean submission time and total interaction time. On the preparatory assignment, where few students started late, we found a statistically significant negative correlation ($r = -0.36, p = 0.008$). Conversely, on the conventional assignment, where many students started late, we found a statistically significant positive correlation ($r = 0.31, p = 0.02$). This supports Hypothesis 3 and provides evidence for our theory that **students running out of time is a significant confounding factor in time-on-task analysis**.

However, it is not the only factor. We also computed the correlation coefficient (two-sided Pearson's r) between mean submission time and total interaction time for the gamified assignment and found a statistically significant positive correlation ($r = 0.28$, $p = 0.031$). Despite few students starting late, how early one worked still had a significant impact on time-on-task, contradicting Hypothesis 3. Furthermore, we also found a statistically significant positive correlation between total interaction time and grade ($r = 0.29, p = 0.03$). Lower performing students spent less time on the assignment, despite working early enough to spend more.

We have two explanations for why this is. First, many lower performing students simply chose to spend less time, presumably satisfied with their grade. For example, despite B students working days before the deadline, some only spent a few hours on the assignment (see Figure 6). Second, in previous work, we found that the gamified assignment was highly engaging and students continued playing the game long after what was required to achieve full marks [17]. As students who were heavily engaged were also shown to perform better, increased engagement increases time-on-task and performance, producing a correlation between the latter two. Furthermore, the earlier a heavily engaged students starts, the longer they can play, producing a correlation between earliness and time-on-task. Therefore, **engagement is another significant confounding factor in time-on-task analysis**.

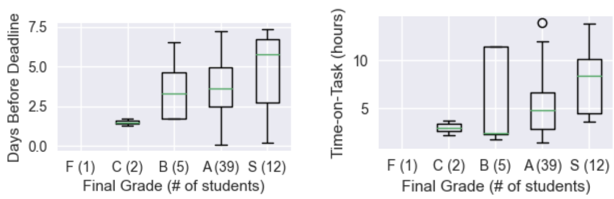


Figure 6: Boxplots of the student's mean connection time (left) and time-on-task (right).

5 DISCUSSION

Earliness of work and time-on-task are often cited as key factors in determining student performance [13, 14]. Our results indicate that, while earliness and time-on-task do often correlate with student performance, the correlations do not always hold, can even be inverted, and there are many confounding factors to consider.

Earliness. The earlier students work, the better they tend to do. However, truly poor performance primarily correlates with running out of time. If students give themselves the time to work they need, how early they start is less important (*i.e.,* it merely allows refinement of an already high score).

Like Leinonen *et al.* [13], we note that students can still work relatively late and succeed. Therefore, placing too much emphasis on earliness can be unwise. Instead, **we suggest analyzing why students delay work too late and considering how to stop it.**

Stress. Students know procrastination is a problem and research has shown that verbal reinforcement of this fact does not significantly change their behavior [23]. Instead, prior work highlights that incentives must be built into the design of assessments to improve student time management [19]. Existing interventions focus on coercing students to start early and work often using strategies like intermittent scheduled feedback [6], limited daily submissions [11], and time management visualizations [10].

We, however, conjecture that most students already know how to manage their time well; they simply do not do so. We base this on our results, where student's time management varied across assignments and was actually quite good on some. Instead, we theorize that students delay work because of stress and dread. This stress can be reduced by providing less stick (lower risk) or more carrot (fun and engagement). Hence, students worked early on the low-risk preparatory assignment and the highly engaging gamified assignment, but delayed the stressful conventional assignment.

To our knowledge, there is little research investigating the use of assignment or course design to reduce stress and thereby improve time management and avoid procrastination. In fact, the intervention in Irwin *et al.* [11] notably *increased* student frustration. In contrast, there is substantial psychology research on academic stress, its causes, and its relation with time management and procrastination [2, 20, 24], but there is little insight on how one should structure a course to reduce it. Therefore, **we recommend further research into this area and suggest designing assignments with more carrot and/or less stick.**

Time-on-Task. Earliness, engagement, and/or skill disparities can confound time-on-task statistics. Students who start late can run out of time, spending less time and receiving lower grades on an assignment than they otherwise would have. Furthermore, even if they have ample time to do more, some students may choose to spend only the time needed to achieve the minimum grade they require. Conversely, if students enjoy their work and are motivated to continue even after achieving the highest marks, they will spend more time on the assignment than necessary. Finally, if students who know the material well can finish quickly while those who are still learning need more time, the correlation between time-on-task and performance may even be negative.

Considering these caveats, **we find time-on-task to be an ambiguous metric and we recommend that time-on-task be decomposed into its constituent factors rather than analyzed as a whole.** For example, in Leinonen *et al.* [14], the machine learning model for predicting performance from time-on-task was much more accurate than simple linear correlations. It may be that the model learned similar caveats to improve its predictions.

6 CONCLUSION

Students can struggle with time management, and we have examined the impact that different assignment styles can have on this behavior. On a conventional assignment, students followed common wisdom: most students worked on the assignment close to the deadline and earliness and time-on-task were strongly correlated with performance. However, on a preparatory or gamified assignment, many students started and finished early, and the correlation with performance was weaker and less significant. Furthermore, on our preparatory assignment, time-on-task was negatively (albeit insignificantly) correlated with performance, clearly defying standard expectations.

Our results suggested that student time management is more complex than previously understood, so we presented the theory that one key element in this equation is academic stress. If an assignment causes too much stress, students delay it; otherwise, they organically manage their time better. Gamified assignments and preparatory assignments reduce stress either by increasing engagement (more carrot) or reducing risk (less stick). In contrast, on a conventional assignment, stress makes them delay work on it, sometimes for so long that they run out of time and underperform.

Regardless of whether our explanation is correct, we hope instructors can use our results to organically improve student time management in their own courses. Expanding on the idea that preparatory assignments help, one might preface a complex task with simpler practice tasks, *e.g.,* following an approach like that of Denny *et al.* [5]. Alternatively, one may seek to increase student engagement through gamification [4]. Or, like us, one may use a combination of the two [17, 26].

REFERENCES

[1] Theresa Beaubouef and John Mason. 2005. Why the High Attrition Rate for Computer Science Students: Some Thoughts and Observations. *SIGCSE Bulletin (Inroads)* 37, 2 (Jun 2005), 103–106. https://doi.org/10.1145/1083431.1083474

[2] Robert T. Brown. 1992. Helping Students Confront and Deal with Stress and Procrastination. *Journal of College Student Psychotherapy* 6, 2 (1992), 87–102. https://doi.org/10.1300/J035v06n02_09

[3] Jerusha O. Conner and Denise C. Pope. 2013. Not Just Robo-Students: Why Full Engagement Matters and How Schools Can Promote It. *Journal of Youth and Adolescence* 42, 9 (01 Sep 2013), 1426–1442. https://doi.org/10.1007/s10964-013-9948-y

[4] Luma da Rocha Seixas, Alex Sandro Gomes, and Ivanildo José de Melo Filho. 2016. Effectiveness of gamification in the engagement of students. *Computers in Human Behavior* 58 (2016), 48–63. https://doi.org/10.1016/j.chb.2015.11.021

[5] Paul Denny, Andrew Luxton-Reilly, Michelle Craig, and Andrew Petersen. 2018. Improving Complex Task Performance Using a Sequence of Simple Practice Tasks. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (Larnaca, Cyprus) *(ITiCSE 2018)*. Association for Computing Machinery, New York, NY, USA, 4–9. https://doi.org/10.1145/3197091.3197141

[6] Paul Denny, Jacqueline Whalley, and Juho Leinonen. 2021. Promoting Early Engagement with Programming Assignments Using Scheduled Automated Feedback. In *Australasian Computing Education Conference* (Virtual, SA, Australia) *(ACE '21)*. Association for Computing Machinery, New York, NY, USA, 88–95. https://doi.org/10.1145/3441636.3442309

[7] John Edwards, Joseph Ditton, Dragan Trninic, Hillary Swanson, Shelsey Sullivan, and Chad Mano. 2020. Syntax Exercises in CS1. In *Proceedings of the 2020 ACM Conference on International Computing Education Research* (Virtual Event, New Zealand) *(ICER '20)*. Association for Computing Machinery, New York, NY, USA, 216–226. https://doi.org/10.1145/3372782.3406259

[8] Stephen H. Edwards, Jason Snyder, Manuel A. Pérez-Quiñones, Anthony Allevato, Dongkwan Kim, and Betsy Tretola. 2009. Comparing Effective and Ineffective Behaviors of Student Programmers. In *Proceedings of the Fifth International Workshop on Computing Education Research Workshop* (Berkeley, CA, USA) *(ICER '09)*. Association for Computing Machinery, New York, NY, USA, 3–14. https://doi.org/10.1145/1584322.1584325

[9] Arto Hellas, Petri Ihantola, Andrew Petersen, Vangel V. Ajanovski, Mirela Gutica, Timo Hynninen, Antti Knutas, Juho Leinonen, Chris Messom, and Soohyun Nam Liao. 2018. Predicting Academic Performance: A Systematic Literature Review. In *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (Larnaca, Cyprus) *(ITiCSE 2018 Companion)*. Association for Computing Machinery, New York, NY, USA, 175–199. https://doi.org/10.1145/3293881.3295783

[10] Kalle Ilves, Juho Leinonen, and Arto Hellas. 2018. Supporting Self-Regulated Learning with Visualizations in Online Learning Environments. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) *(SIGCSE '18)*. Association for Computing Machinery, New York, NY, USA, 257–262. https://doi.org/10.1145/3159450.3159509

[11] Michael S. Irwin and Stephen H. Edwards. 2019. Can Mobile Gaming Psychology Be Used to Improve Time Management on Programming Assignments?. In *Proceedings of the ACM Conference on Global Computing Education* (Chengdu,Sichuan, China) *(CompEd '19)*. Association for Computing Machinery, New York, NY, USA, 208–214. https://doi.org/10.1145/3300115.3309517

[12] Richard N. Landers and Amy K. Landers. 2014. An Empirical Test of the Theory of Gamified Learning: The Effect of Leaderboards on Time-on-Task and Academic Performance. *Simulation & Gaming* 45, 6 (2014), 769–785. https://doi.org/10.1177/1046878114563662

[13] Juho Leinonen, Francisco Enrique Vicente Castro, and Arto Hellas. 2021. Does the Early Bird Catch the Worm? Earliness of Students' Work and Its Relationship with Course Outcomes. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1* (Virtual Event, Germany) *(ITiCSE '21)*. Association for Computing Machinery, New York, NY, USA, 373–379. https://doi.org/10.1145/3430665.3456383

[14] Juho Leinonen, Francisco Enrique Vicente Castro, and Arto Hellas. 2022. Time-on-Task Metrics for Predicting Performance. *ACM Inroads* 13, 2 (May 2022), 42–49. https://doi.org/10.1145/3534564

[15] Therese H Macan, Comila Shahani, Robert L Dipboye, and Amanda P Phillips. 1990. College students' time management: Correlations with academic performance and stress. *Journal of Educational Psychology* 82 (1990), 760–768. https://doi.org/10.1037/0022-0663.82.4.760

[16] Mac Malone and Fabian Monrose. 2021. Game & Match: The Effects of Various Gamified Elements in a Computer Security Course. Univeristy of North Carolina at Chapel Hill. IRB #21-1879..

[17] Mac Malone, Yicheng Wang, Kedrian James, Murray Anderegg, Jan Werner, and Fabian Monrose. 2021. To Gamify or Not? On Leaderboard Effects, Student Engagement and Learning Outcomes in a Cybersecurity Intervention. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, USA, 1135–1141. https://doi.org/10.1145/3408877.3432544

[18] Mac Malone, Yicheng Wang, and Fabian Monrose. 2021. An Online Gamified Learning Platform for Teaching Cybersecurity and More. In *Proceedings of the 22st Annual Conference on Information Technology Education* (SnowBird, UT, USA) *(SIGITE '21)*. Association for Computing Machinery, New York, NY, USA, 29–34. https://doi.org/10.1145/3450329.3476859

[19] Joshua Martin, Stephen H. Edwards, and Clfford A. Shaffer. 2015. The Effects of Procrastination Interventions on Programming Project Success. In *Proceedings of the Eleventh Annual International Conference on International Computing Education Research* (Omaha, Nebraska, USA) *(ICER '15)*. Association for Computing Machinery, New York, NY, USA, 3–11. https://doi.org/10.1145/2787622.2787730

[20] Ranjita Misra and Michelle McKean. 2000. College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction. *American journal of Health studies* 16, 1 (2000), 41.

[21] Dale Parsons and Patricia Haden. 2006. Parson's programming puzzles: a fun and effective learning tool for first programming courses. In *Proceedings of the 8th Australasian Conference on Computing Education-Volume 52*. 157–163.

[22] Andrew Petersen, Michelle Craig, Jennifer Campbell, and Anya Tafliovich. 2016. Revisiting Why Students Drop CS1. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research* (Koli, Finland) *(Koli Calling '16)*. Association for Computing Machinery, New York, NY, USA, 71–80. https://doi.org/10.1145/2999541.2999552

[23] Dhananjai M. Rao. 2019. Experiences With Auto-Grading in a Systems Course. In *2019 IEEE Frontiers in Education Conference (FIE)*. 1–8. https://doi.org/10.1109/FIE43999.2019.9028450

[24] Fuschia M. Sirois. 2014. Procrastination and Stress: Exploring the Role of Self-compassion. *Self and Identity* 13, 2 (2014), 128–145. https://doi.org/10.1080/15298868.2013.763404

[25] Mark Trueman and James Hartley. 1996. A comparison between the time-management skills and academic performance of mature and traditional-entry university students. *Higher Education* 32, 2 (01 Sep 1996), 199–215. https://doi.org/10.1007/BF00138396

[26] Valdemar Švábenský and Jan Vykopal. 2018. Challenges Arising from Prerequisite Testing in Cybersecurity Games. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, MD, USA) *(SIGCSE '18)*. Association for Computing Machinery, New York, NY, USA, 56–61. https://doi.org/10.1145/3159450.3159454

[27] Matthew M. Walsh, Kevin A. Gluck, Glenn Gunzelmann, Tiffany Jastrzembski, and Michael Krusmark. 2018. Evaluating the Theoretic Adequacy and Applied Potential of Computational Models of the Spacing Effect. *Cognitive Science* 42, S3 (2018), 644–691. https://doi.org/10.1111/cogs.12602

[28] Iman YeckehZaare, Elijah Fox, Gail Grot, Sean Chen, Claire Walkosak, Kevin Kwon, Annelise Hofmann, Jessica Steir, Olivia McGeough, and Nealie Silverstein. 2021. Incentivized Spacing and Gender in Computer Science Education. In *Proceedings of the 17th ACM Conference on International Computing Education Research* (Virtual Event, USA) *(ICER 2021)*. Association for Computing Machinery, New York, NY, USA, 18–28. https://doi.org/10.1145/3446871.3469760