2023

Due Tomorrow, Do Tomorrow: Measuring and Reducing Procrastination Behavior Among Introductory Physics Students in an Online Environment

Zachary Felker
University of Central Florida

DUE TOMORROW, DO TOMORROW: MEASURING AND REDUCING PROCRASTINATION BEHAVIOR AMONG INTRODUCTORY PHYSICS STUDENTS IN AN ONLINE ENVIRONMENT

by

ZACHARY FELKER
B.S. University of Michigan, 2018
M.S. University of Central Florida, 2021

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Physics
in the College of Science
at the University of Central Florida
Orlando, Florida

Summer Term
2023

Major Professor: Zhongzhou Chen

ABSTRACT

This work is focused on the measurement and prevention of procrastination behavior among college level introductory physics students completing online assignments in the form of mastery-based online learning modules. The research is conducted in two studies. The first study evaluates the effectiveness of offering students the opportunity to earn a small amount of extra credit for completing portions of their homework early. Unsupervised machine learning is used to identify an optimum cutoff duration which differentiates taking a short break during a continuous study session from a long break between two different study sessions. Using this cutoff, the study shows that the extra credit encouraged students to complete assignments earlier. The second study examines the impact of adding a planning-prompt survey prior to a string of assignments. In the survey, students were asked to write a plan for when and where they would work on their online homework assignments. Using a *difference in differences* method, a multilinear modeling technique adopted from economics research, the study shows that the survey led to students completing their homework on average 18 hours earlier and spreading their efforts on the homework over time significantly more. On the other hand, behaviors associated with disengagement, such as guessing or answer-copying, were *not* impacted by the introduction of the planning prompt. These studies showcase novel methods for measurement of procrastination behavior, as well as evaluating the effectiveness of the designed interventions to help students avoid waiting until the last minute to make progress on assigned tasks.

Dedicated to myself.

After all, why not? Why shouldn't I keep it?

ACKNOWLEDGEMENTS

An endeavor such as this is impossible without the support of many.

First and foremost, I extend my gratitude to Professor Zhongzhou Chen, whose patience, guidance, advice, support, and allocation of grant funds have made my graduate career possible.

To my parents I express my most sincere appreciation; without them I could not have dreamed of achieving what I have. The same goes for my many friends and family members.

For my intellectual formation I am grateful to the faculty and staff of Rockford Public Schools, and of my most beloved *alma mater,* the University of Michigan.

I am thankful for those in my field who came before me. I am especially thankful for the cautionary example of L. L. Ainsworth, whose 1979 paper never fails to summon fury that motivates me to continue [1]; and for the inspiration of Michael Yeomans and Justin Reich, whose 2017 paper provided inspiration for the design of my main experimental study [2].

Many thanks are due to the Instructional Systems and Technology team at the University of Central Florida's Center for Distributed Learning for developing the Obojobo platform and providing the log data for analysis.

All research herein was generously supported by the people of the United States through the National Science Foundation Division of Undergraduate Education grant number 1845436.

Finally, I acknowledge the support of my dog, Mushu, who ensures I get out of bed each day whether I want to or not.

To those I have inevitably neglected to mention here: please accept my thanks nonetheless.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1:
INTRODUCTION

As a sufferer of ADHD since childhood, I have experienced myself the importance of external support and structure in helping students actively manage their time and avoid procrastination. Disorders such as ADHD that impact executive function can paralyze a student who has every desire to begin a task. In an extreme example, the policies of a particularly permissive professor combined with poor mental health led to me being forced to rush through half a semester of galactic astrophysics homework over the course of four days. It is my personal experience that the rushed pace and high stress level associated with cramming is not conducive to effective learning.

Today's online learning technologies enable unprecedented flexibility that allow more students to work at their own pace. While studies have shown that some level of self-pacing can be good, too much flexibility in pacing can be detrimental by placing the burden of time management on the student, as in the case of my galactic astronomy class [3,4]. Some instructors, such as Ainsworth, place the blame on students by arguing that certain demographics are simply not capable of time management [1]. This view is, put mildly, misguided. Fortunately, online learning platforms also provide instructors with unprecedented ability to implement various kinds of supports, as well as gather data on the effectiveness of those interventions. In particular, online learning technology enables instructors to more easily implement interventions that provide scaffolding for students to properly spread their work, as well as enable researchers to evaluate the effectiveness of those interventions through the analysis of log data.

In this dissertation, we examine the impact of two different interventions on students' work distribution via two novel approaches to quantitatively measuring procrastination behavior based on log data from online homework assignments. We showcase these techniques in the context of evaluating the impact and effectiveness of different instructional interventions that show promise in supporting students' ability to space out their work instead of cramming the night the assignment is due. In the first study, we evaluate the effectiveness of extra credit incentives on student work distribution. In the second, we examine the impact of a planning prompt survey on student procrastination behavior.

1.1. Effects of Procrastination on Academic Performance

Studies overwhelmingly show that procrastination behavior is associated with reduced academic performance [5–9]. For example, Agnihotri, Baker, and Stalzer found that habitual procrastination as measured by a procrastination index is associated with a 21-fold higher risk of failing the course [10]. Sabnis, Yu, and Kizilcec also found that procrastination is more common among males, racial minorities, and first-generation students [11].

According to an extensive review by Dunlosky et al., theoretical explanations for the observed benefit to spaced work—as opposed to procrastination—can be sorted in to two general categories: cognitive and behavioral [12].

1.1.1. Cognitive

The cognitive aspects of spacing are the effects we see as a result of the delay between study sessions itself. There is a strong consensus that when encountering information more than once, spacing the exposures in time dramatically improves learning of the information. There are a few theories why this is the case [12].

The *deficient processing* theory argues that the spacing effect is due to metacognition. The lack of difficulty in recalling information just seen, the theory argues, can lead students to incorrectly believe that they know the material better than they actually do, which hinders subsequent learning [12,13].

Another theory, referred to as "*reminding*" by Dunlosky et al., holds that when a student is presented with the material for a subsequent time, they are reminded of the previous encounter with it. Reminding the student prompts retrieval, which is more strenuous the more time has passed since the previous learning event; it is well documented that retrieval leads to improved understanding [12,14].

Dunlosky et al. further note that in addition to the above theories subsequent learning opportunities will benefit from consolidation of the prior opportunities to the extent that enough time has passed for the consolidation to occur [12]. Literature related to the cognitive effects of spaced study concentrates on recall of the same information or skill later.

Literature related to the cognitive effects of spaced study concentrates on recall of the same information or skill later. In this dissertation, we study spaced study in the context of learning new material in the domain of physics. In physics, new knowledge builds extensively on existing knowledge, leading to a high degree of repetition for existing concepts. For example, a simple conservation of momentum problem may present momentum as a new concept while still requiring students to retrieve knowledge of forces, vectors, coordinate systems, etc. This enables students to enjoy the benefits of spaced practice documented in the literature despite learning new material; students could potentially benefit from one or more of the above effects when retrieving existing knowledge that recurs later.

1.1.2. Behavioral

Behavioral mechanisms of spaced versus massed learning are the result of behavioral differences exhibited by students engaged in spaced learning as opposed to cramming.

The behavioral mechanism rests on the fact that cramming is often the result of rushing to meet a deadline. When work is spaced, time pressure is significantly reduced, allowing students to spend longer on the material to be learned. Study time is likely positively associated with learning outcome, leading to a time-mediated improvement in learning outcome for spaced work [15]. In addition to time-on-task, it has been shown that students who space their study are more likely to engage in other demonstrably effective learning strategies such as self-testing, although it is unclear whether use of learning strategies is caused by spacing or whether both stem from a student's skill in self-regulated learning [16].

1.2. Theoretical Framework: Self-Regulated Learning

Self-regulation is "the self-directive process by which learners transform their mental abilities into academic skills … [and] refers to self-generated thoughts, feelings, and behaviors that are oriented toward attaining goals" [17]. Self-regulated learning (SRL) is integral to how students manage their time and space their work. It is self-regulation skills that allow students to actively set goals for spaced learning, plan for achieving those goals, follow their plan, and reflect on their effectiveness after the fact.

There are several competing SRL models. Of these, the Zimmerman model is the simplest as well as one of the oldest [18]. Zimmerman models SRL in three interdependent phases in a cycle: forethought, performance, and self-reflection. In Zimmerman's words: "The forethought phase refers to processes and beliefs that occur before efforts to learn; the performance phase refers to processes that occur during behavioral implementation, and self-reflection refers to processes that occur after each learning effort”. [17]

1.2.1. Forethought

Processes in the forethought phase belong to two categories: task analysis and self-motivation. Task analysis involves identification of effective learning skills and strategies, strategic planning for implementing the strategies, and setting specific goals for the implementation of the strategic plan. Self-motivation arises from student beliefs about learning and the material to be learned [17]. Procrastination can happen because of an ineffective forethought phase, leading to a failure to plan for effective work distribution. It could also be caused by a plan to procrastinate due to lack of self-motivation or self-confidence, or incorrect task analysis, such as underestimating the difficulty of the task.

1.2.2. Performance

The performance phase also has two categories: self-control and self-observation. A student exercises self-control when they put into action the strategic plan that they envisioned in the forethought phase. This is propelled by the motivation engendered in the previous phase. Self-observation is a learner’s tracking of personal learning events and experimentation to assess the cause of such events and results of such experiments [17]. Procrastination could also result from ineffective performance-phase self-regulation through to lack of executive function. Students who had a plan to avoid procrastination may fail to execute that plan, and thus end up procrastinating.

1.2.3. Self-Reflection

There are once again two categories in the self-reflection phase: self-judgement and self-reaction. Self-judgement contains self-evaluation, in which a student compares their performance against some standard. It also contains causal attribution, through which the learner assigns beliefs about what events, strategies, or influences contributed to the performance. Self-reaction looks forward to future learning events [17]. The connection between self-reflection and procrastination is less obvious. It could be that the action of procrastination is the result of self-reflection on previous failures. The student may react defensively by not studying at all in the future to lessen the emotional impact of the anticipated future poor performance.

1.2.4. Relationship Between Phases

Notice that each phase of the model has two categories of cognitive tasks. One of these looks forward while the other is retrospective. In the forethought phase, *self-motivation* is based on beliefs about learning and about one's self-efficacy previously formed in self-reflection. *Task analysis* (forethought) anticipates and prepares for the subsequent performance phase, where *self-control* (performance) realizes the plans made. During the performance phase a learner is constantly engaging in *self-observation*, gathering information that they use for *self-judgment* in self-reflection. The cycle closes as judgements made about performance generate *self-reaction*, which leads to new beliefs to feed future *self-motivation*.

Note also that the cycle operates over a wide range of timescales—from reading individual sentences to entire college careers. Thus, a student is to be likely in multiple phases at the same time with respect to different tasks. For example, making a schedule is both a task to be performed and forethought for an upcoming task. It could also be an opportunity to engage in reflection on past tasks.

1.3. Measuring Procrastination Behavior

Early studies seeking to measure procrastination behavior relied on surveys or self-reporting by students [5,6,9,19,20]. More recent studies have turned to analysis of log data from online learning platforms to measure student work distribution [10,11,15,21–31].

There has yet to emerge a consensus on how to measure procrastination behavior. One method is to use a cutoff to separate student efforts into sessions—if a gap in work is longer than the cutoff, that gap represents a break of work into two sessions rather than a break within a session. For example, Miyamoto et al. used a 30-minute cutoff to evaluate spaced work in the context of a massive open online course (MOOC) [15]. They found that more study sessions correlated with increased completion rate, even after controlling for total time spent. While promising, this method relies on an entirely arbitrary cutoff time. In our first study we use a mixture-modeling method to find an appropriate cutoff based on the data [30]. This method treats procrastination as equivalent to massed study.

A more recent approach has been comparing a student's submission time against their peers. This can be used to create a procrastination index which captures how late students work, as opposed to how many sessions students complete.

Agnihotri, Baker, and Stalzer used the time at which 75% of students had begun work on a task to set a procrastination "deadline" for each assignment. Their study used data from two multi-institution datasets from the Connect online learning platform with data from more than 100,000 students over 3700 course sections. They found that students who habitually fall in the last 25% of students beginning assignments have 21-times the failure rate for courses. [10]

Sabnis, Yu, and Kizilcec instead examined the submission times and based their procrastination index on the assumption that the last 10% of submissions come from students who procrastinated. Their study used data aggregated from their university's learning management system over 2200 courses with 26,000 students. They found that students among the last 10% to submit an assignment were more likely to be male, underrepresented minorities, or first-generation college students. They found a very weak negative association between their procrastination index and course performance, which was uniform across student groups. [11]

Approaches such as these make it impossible to meaningfully detect an overall decrease in procrastination behavior, as a fixed fraction of students are labeled as exhibiting the behavior definitionally. This makes it difficult to gauge the effectiveness of instructional interventions. Other than our work, there has been little progress in measuring procrastination in a way that allows comparison between groups [21,30,31].

One notable exception is Nieberding and Heckler, who used students' mean completion time as a proxy for cramming behavior. They used latent profile analysis to sort students into "[four] completion time classes that clearly distinguish students between their mean completion time, their week-to-week completion time patterns, assignment completion rates, mean course grades, and proportion of women". They found that completion time is associated with non-exam course achievement, but that exam score differences associated with late completion time were completely mediated by the non-exam scores. They also found that ACT scores were correlated with course exam scores but not with non-exam grades. This led them to suggest that ACT scores and completion time can be considered orthogonal metrics in predicting course achievement. Their results regarding gender differences in procrastination behavior match those of Sabnis, Yu, and Kizilcec [11], finding that women complete assignments 8 hours earlier on average and earn corresponding higher grades in non-exam course components [29].

Our second study builds on the measurement techniques used by Nieberding and Heckler by measuring procrastination using an ensemble of metrics such as mean and standard deviation of completion times and measures of student disengagement. We also bring to bear the *differences in differences* method of analysis to probe the specific impact of our intervention on the measured behavior [21,31].

1.4. Reducing Procrastination Behavior

Compared to studies attempting to measure procrastination behavior, fewer have studied how to reduce it.

Ackerman and Gross found that providing incentives for early completion could significantly reduce self-reported procrastination. Their study asked students on which assignments they procrastinated and about various assignment factors that influenced the tendency of students to put off work. They found that, among other factors, perceived rewards offered for getting an early start encouraged students to do so [5]. Studies from much earlier found similar results regarding the efficacy of explicit rewards for early progress [19,20]. We report similar results in our first study, which examines the effectiveness of offering small amounts of extra credit for early progress [30].

Cavanaugh, Lamkin and Hu sent a checklist to students to encourage early submission. They divided their 56 graduate education students into a treatment and a control group and found that giving students the checklist of tasks led to students submitting their assignments on average almost a full day earlier. Their results are limited by both the small sample size and the way in which they measured submission time. They only recorded the date of submission, not the time. This limits the precision to one day increments. [32]

In blended and online learning settings, *nudging* is a strategy often used to influence student behavior—see Damgaard and Nielsen for a more comprehensive review [33]. Nudging in the form of goal setting activities or reminder emails or texts have been frequently used to fight procrastination and distraction. For example, Patterson found that asking students to set a goal for limiting distracting internet time increased course engagement and completion in a massive open online course (MOOC) [34]. Huang et al. examined multiple forms of email "calls to action" and found that while descriptive norms lead to reduced procrastination, deadline reminders can actually backfire and result in increased dropouts [35]. Fouh, Lee, and Baker found that nudging emails induced students to use more free late days (allow for late submission without grade penalty) on homework assignments, but the change was short lived and students quickly reverted to previous behavior [36].

Finally, Yeomans and Reich sent planning prompts to students taking a massive open online course (MOOC). Providing students with this survey, which asked them when and where they planned to make progress on course content, increased completion of the online course by 29%. They also found that students who mentioned specific times rather than a more general timeframe were unlikely to complete the course. While this work did not directly address procrastination, the planning prompt approach inspired our implementation of a similar survey aimed at helping students space their work more effectively. [2]

1.5. Dissertation Outline

Our work focuses on bolstering the two areas of weakness in the literature identified above: few studies have examined the efficacy of interventions on procrastination behavior, and few studies on the topic deploy measurement techniques that can resolve improvements in work spacing on the scale of an entire course.

First, in Chapter 2, we will discuss the instructional context of our studies. In Chapter 3 we present a 2020 study that measured the impact of offering small amounts of extra credit for early progress on assignments [30]. Chapter 4 presents our study published in 2022 and 2023 that evaluates the effectiveness of using a planning prompt survey similar to the one deployed by Yeomans and Reich [2,21,31]. Finally, Chapter 5 concludes the dissertation.

CHAPTER 2:
INSTRUCTIONAL CONTEXT

All data for this work was collected using online homework modules deployed in sections of introductory mechanics, Physics 2048: Physics 1 with Calculus, at the University of Central Florida taught between 2018 and 2022. Sizes for sections of the course range from 120 to 288 students, with about 25%–35% female and 30%–40% under-represented minority students (self-report data for both). All of the sections are taught with a traditional lecture format. Class attendance is either required or optional depending on the instructor.

Beginning in 2018, we created online learning modules to replace both the textbook and the homework for the course. These are designed to provide free and open-source instruction to the students while also recording all student interaction in the form of event logs. It gathers this clickstream data to facilitate physics education research [22,37]. This is made possible by the Obojobo platform on which the modules are built. Obojobo is an open-source platform developed by UCF's Center for Distributed Learning. Both the platform and the course modules are free of charge to UCF students.

2.1. Module Design

Online learning modules implemented over the course of several years are central to our research. These modules, which replace both the homework and the textbook for the course, cover either a single concept or single type of problem and consist of three parts: assessment, instruction, and practice. These modules accounted for 30% of students' overall grade.

2.1.1. Assessment

The assessment portion of the module is designed to test students' understanding of the material contained in that module. It typically consists of one or two multiple-choice questions—though later produced modules contain questions which ask a student to type in a number. There are always at least three different versions of the assessment which cycle as a student makes attempts at the assessment. The question versions are designed to be of equal difficulty and often feature the same mathematical steps to solve, with only the numbers and the context of the problem changed. We created the incorrect choices for the multiple-choice questions by solving the problems while making errors commonly seen in students solving problems. An example would be use of the cosine instead of the sine or having one's calculator set to radians instead of degrees.

Students are allowed 5 attempts to complete the assessment. If a student's highest score occurs on their fourth or fifth attempt, their score is capped at 90% of the credit for the module. This is partially because on the fourth attempt a student will generally see the exact problem that they encountered in their first attempt. The solution to this version of the problem is also typically given in the practice section of the module.

For all modules involved in the studies of this dissertation, we require students to complete one assessment attempt before accessing any other module content. This both serves as a formative assessment and allows students who already understand the material to skip to the next module in the sequence [37,38].

2.1.2. Instructional Content

As mentioned above, the learning modules replace both the textbook and the homework. The instructional content contains text and figures adapted from *University Physics Volume 1* from OpenStax CNX, a free online textbook available through creative commons license [39]. We designed each module to focus on a specific topic so that the instructional content for most modules is under two pages of text.

2.1.3. Practice Problems

The practice problem portion of a module contains one or more questions in the same format as presented in the assessment. Once a student gives an answer for a practice question, they can view the correct answer; a full solution to the problem (if available); and, in the case of multiple-choice questions, a feedback message that either praises the student for their correct answer or tells them what error was used to generate the incorrect answer they selected. For most modules, students first see a simpler practice problem compared to the assessment before encountering questions identical to those on the first assessment attempt of the module. Seeing the same problems again after instruction allows the student to address mistakes they may have made during their first attempt.

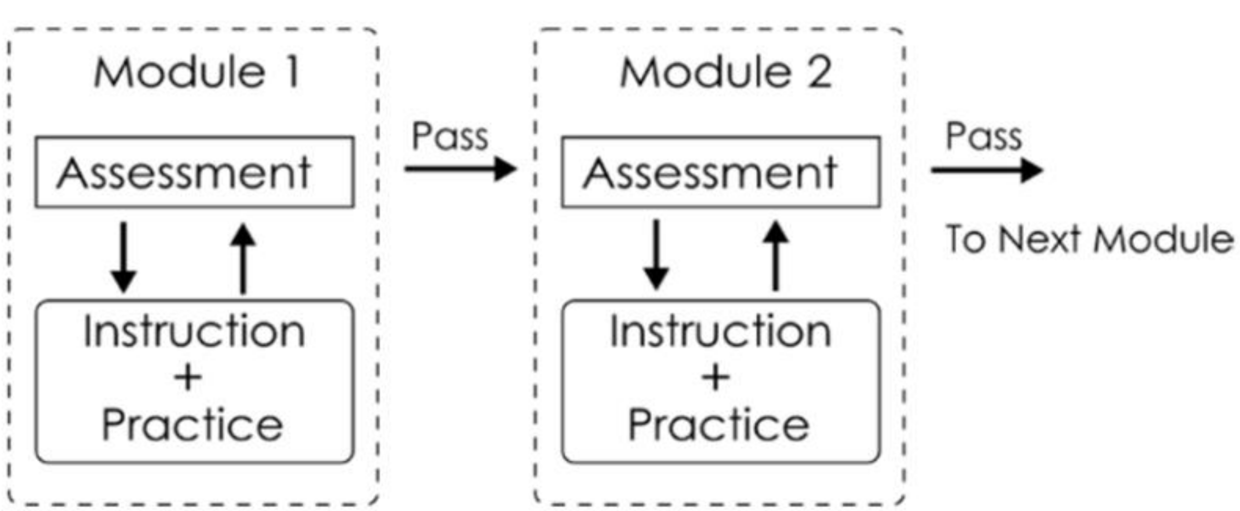

Figure 1: Diagram Showing Module Structure. Students must begin each module by accessing and submitting the assessment at least once. After the initial attempt, the instruction and practice section of the module is unlocked. Once a student passes the assessment by answering the question(s) correctly, they may proceed to the next module. They may also proceed by exhausting their five assessment attempts.

2.1.4. Mastery-Based Module Sequences

Mastery-based learning modules are assigned to students in sequences of 7–16 modules. Students must complete each module in the sequence in the given order; modules (after the first) remain locked until a student completes the previous module by either attaining a perfect score on the assessment or exhausting their 5 assessment attempts. This is called a “mastery-based” design because it allows a student to progress only when they have mastered the material to be learned. The assessment-first structure of the modules fosters this design, allowing students to advance quickly through a sequence until they encounter a concept they have not mastered. See Figure 1.

The module sequences each cover one to two weeks’ worth of course content—roughly equivalent to a chapter in a textbook. Module sequences are made available to students at around the time of the first lecture on the topic covered. They are due approximately two weeks later, at 23:59 on the Saturday following the last lecture on the topic. Prior to the Fall term of 2020, no credit was given for module progress after the due date. For the Fall 2020 semester and all subsequent semesters (as of Spring 2023), the due date policies have been softened: submission of assignments after the due date now receives a 13-percentage-point deduction per day, rounded up to the nearest day. This attenuates to no credit being possible 7 days after the posted due date.

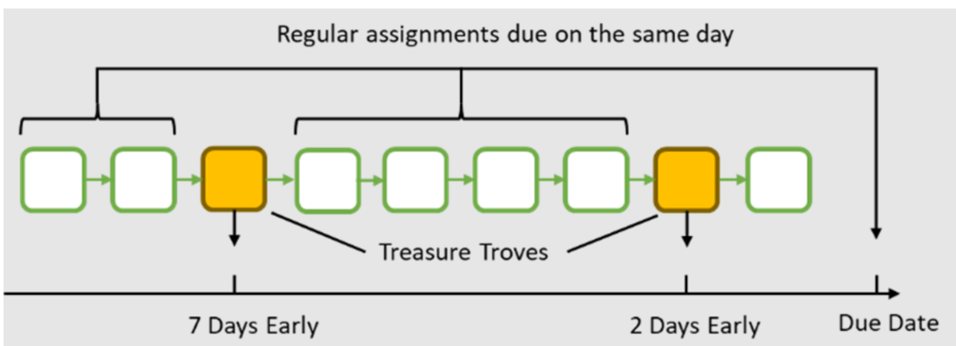

Figure 2: Illustration of Module Sequence Structure. White squares represent learning modules containing course content. These all share the same due date. Yellow squares represent treasure trove extra credit opportunities. Each of these is due before the sequence due date.

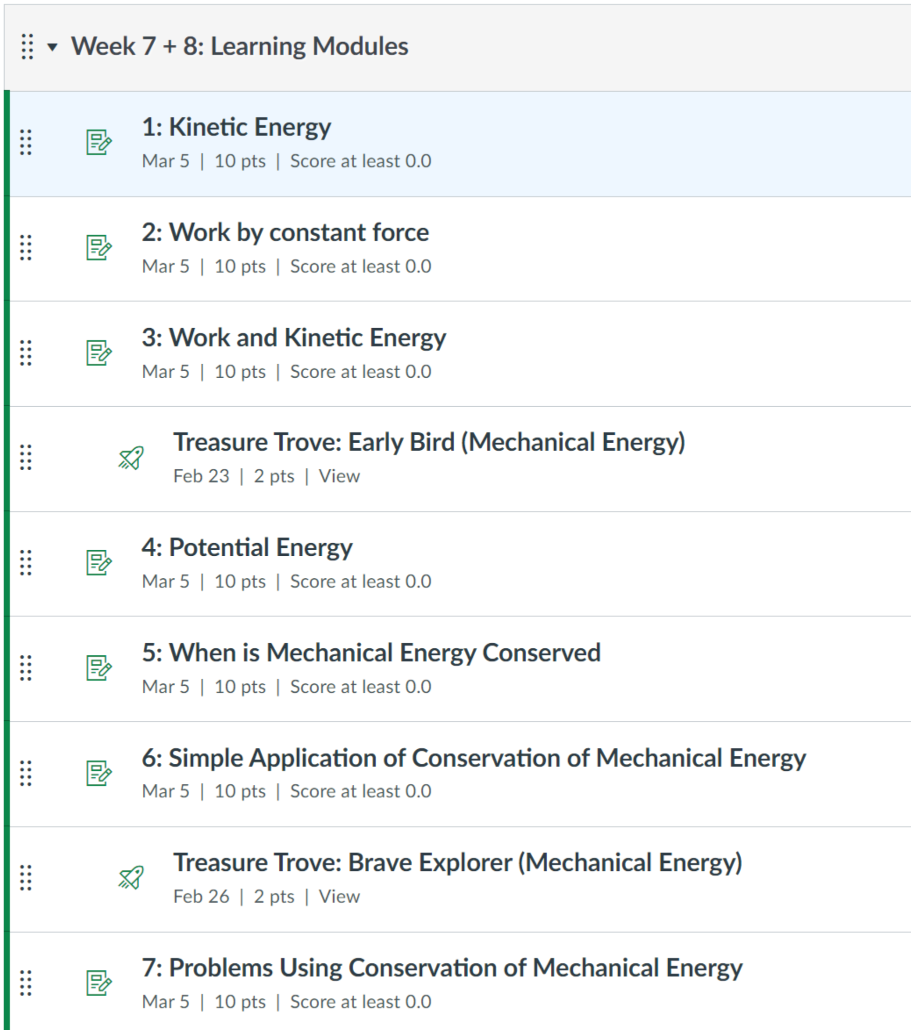

Figure 3: Screenshot of Energy Sequence Showing Extra Credit Opportunities. The sequence of learning modules shown was due at 23:59 on 5 March. The treasure trove quizzes expired earlier to encourage students to make early progress on the sequence. Each module after the first was locked until the student completed the one prior in the sequence.

2.1.5. Extra-Credit for Early Progress

Students can earn extra credit by completing certain numbers of modules a specified number of days prior to the due date. For example, on the sequence covering mechanical energy, students who complete the first 3 of the 10 modules 10 days before the due date can earn 2 points of extra credit; those who complete the first 6 modules 7 days in advance can earn and additional 2 points of extra credit; and those who completed 9 modules 2 days ahead can earn another 3 points of extra credit. This is implemented in the form of "treasure trove" Canvas quizzes embedded in the module sequence with due dates in advance of the sequence due date as shown in Figure 2 and Figure 3. These quizzes contain a single question: "Do you want extra credit?" The extra credit quizzes were implemented in the Spring of 2019. See Figure 4. All semesters thereafter include them.

The setting that allows soft due dates on the modules cannot be turned off for the treasure troves. This means that students can earn some extra credit even if they do not follow the treasure trove schedule. None of the instructors of the course sections involved in the study announced this fact to students.

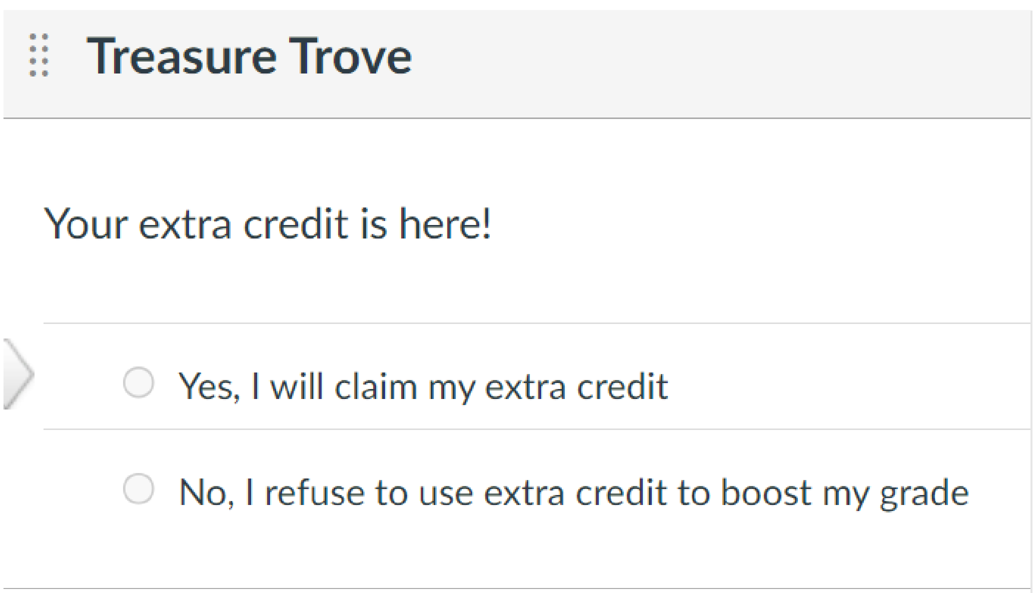

Figure 4: Screenshot of Treasure Trove Extra Credit Quiz. Students who arrive at the quiz are asked whether they would like extra credit. Answering in the affirmative earns extra credit if the deadline for the treasure trove—set in advance of the sequence due date—has not elapsed.

CHAPTER 3:
IMPACT OF EXTRA CREDIT INCENTIVES

The contents of this chapter are adapted from results published in the proceedings of the *2020 Physics Education Research Conference* [30]. See Appendix C for the cover page of the publication.

3.1. Introduction

Many college students frequently procrastinate on assignments and cram before due dates and exams. Several studies have shown that those behaviors are frequently associated with lower academic performance or learning outcomes. While earlier research relied on surveys or self-reports [5,6,9,19,20], more recent studies have utilized the log data from online learning platforms to measure quantitatively students' work distribution [15,25,27,28,40]. Using log data allows for a much more accurate and objective measurement of students' work distribution.

In addition to observing and measuring the level of cramming or procrastination, few studies have tried to answer the more important question of how to reduce such behavior and incentivize students to distribute their work. For example, Cavanaugh, Lamkin and Hu [32] sent a checklist to students to encourage early submission; Yeomans and Reich sent planning prompts to students taking a massive open online course (MOOC) [2]. Ackerman and Gross [5] found that providing incentives for early completion could significantly reduce procrastination, while two much earlier papers measured the effectiveness of such incentives [19,20]. One way of providing this type of incentive is by giving extra credit for early completion of assignments, which rewards early submission and distribution of work but does not penalize those who are unable to do so. Advancement in online learning technology makes it easy to implement flexible forms of extra credit on major learning management systems such as Canvas.

This study examines the effectiveness of offering such extra credit on students' work distribution and study time as they complete online homework in the form of a sequence of *online learning modules* [22,37,41]. It also showcases a novel method to measure student work distribution.

The first research question (**RQ1**) that we answer is: How can we accurately measure students' distribution of work over time? We adopt an approach similar to that of Miyamoto et al. by grouping learning events into "study sessions" separated by a time gap [15]. Compared to the more common approach of analyzing event frequencies [25,28,40,42], study sessions better capture students' work distribution by providing information on not only the timing of work but also the duration of work. Unlike the Miyamoto study—which used a 30-minute minimum gap duration—we performed a more careful analysis on the distribution of gaps between learning events to estimate the separation between different study sessions [15]. In addition, we also exclude those sessions that were likely generated by students quickly browsing through the learning materials. Based on the identified study sessions, we will answer **the second research question** (**RQ2):** Can a small amount of extra credit incentivize students to start working on assignments earlier?

In addition, we examine two potential side effects of using extra credit incentives to reduce cramming and procrastination. First, research on student self-regulated learning (SRL) suggests that some learners may focus more on achieving extrinsic goals, in this case acquiring the extra credit [43–45]. For those learners, it is possible that they will replace one cramming session with multiple earlier cramming sessions to earn the extra credit without increasing their study time on learning materials. Second, students with a higher level of self-regulatory skill are not only more likely to earn the extra credit through proper planning and distribution of work, but are also likely to perform better in other aspects of the course such as exams [46]. Therefore, it is possible that extra credit will be earned primarily by already high performing students, having the effect of "making the rich richer". To measure the extent of those side effects, we will explore the answer to the following two research questions:

RQ3: To what extent do extra credit incentives lead to an increase in study time on the assigned learning materials?

RQ4: Are our extra credit opportunities benefiting predominantly high-performing students?

In 3.2, Methods, we will describe in detail the implementation of extra credit, the collection and selection of log data as well as analysis methods to answer each of the four research questions. In 3.3, Results, we present the outcome of the analysis, followed by a discussion on the answers to our four research questions in 3.4, Discussion and Future Directions.

3.2. Methods

3.2.1. Implementation of Online Homework Assignments and Extra Credit

Online Homework Assignments in the form of online learning module sequences are created on the free and open source Obojobo platform, developed by the Center for Distributed Learning at the University of Central Florida, which is integrated with the Canvas Learning Management System via Learning Technology Interoperability standards. Each module covers a single concept or one type of problem using an instructional component containing text and practice problems, and an assessment component containing 1 or 2 multiple choice problems. Several modules form a module sequence on a given topic and students must complete each module in order. The sequence in the current study consists of 10 modules on the topic of conservation of mechanical energy. The sequence was assigned as homework to be completed over two weeks in both Fall 2018 and Fall 2019 semesters. Students in each semester enrolled in the same calculus-based college introductory physics course taught in lecture format by the same instructor.

Extra credit for early completion: In Fall 2019 semester, students who completed the first 3, 6 and 9 out of 10 modules of the Energy sequence could each access one of three additional Canvas quizzes named “Treasure Troves”. Each contained only a single question asking if students want to claim their extra credit. The three quizzes expire 10, 7 and 2 days before the assignment due date, and are worth 2, 2, and 3 points in extra credit, respectively. All “Treasure Trove” quizzes in the course (across all module sequences) contain a total of 21 points and are worth 5% of course credit.

3.2.2. Identifying Valid Study Sessions from Event Logs (RQ1)

Collection of Online Event Log Data: A subset of all log events are selected for analysis. These events correspond to students navigating between instructional pages and assessments, viewing and answering questions, and entering and leaving the modules.

Identifying study sessions by clustering of log events: Our method for identifying students’ study session from the log data is based on the following hypothesis: Longer breaks between study sessions and shorter breaks within a study session arise from different and independent student behavior. The former are closely related to the students’ decision to distribute work over time, while the latter stem from the student taking breaks or running errands during a continued period of work. We measure these breaks using the time elapsed between a “viewer-exit” event, which triggers when either the module is closed or the browser window is minimized or remain inactive for more than 10 minutes, and the following “viewer-enter” event, triggered when the student returns. Therefore, the distribution of time elapsed between consecutive exit and enter events should contain two or more separate distributions, which can be separated by fitting the data using multi-component mixture model, similar to the analysis in several earlier papers [22,38,47,48].

Based on the fitting result (as shown in Figure 5), we estimate the maximum separation for two consecutive log events to be considered to belong to the same study session. Students' study sessions can then be identified by clustering all events that took place within the maximum separation into one cluster, using the DBSCAN clustering algorithm from Scikit-Learn [49].

Selecting valid study sessions: Some of the study sessions identified via this method arise from students briefly glancing over the homework assignments. To ensure that the identified study sessions corresponds to actual engagement with the learning material, we filter out sessions that are either too short, contain too few events, or have a time density of events that is too low, according to the distribution of study sessions (Figure 6) and methods explained in the next section.

3.2.3. Answering Research Questions 2–4

To answer Research Question 2, regarding whether extra credit can motivate students to distribute their work, we compare the distribution of the start dates of all valid study sessions with respect to the assignment due date and examine if the distributions are significantly different before and after the implementation of extra credit incentives (Figure 7).

To examine the relationship between distribution of work and engagement with learning resources (RQ3), we create two proxy variables for each student. A student's distribution of work is captured by $\bar{d}$, the weighted average start date of all study sessions with respect to the due date:

$$\bar{d} = \frac{\sum_s m_s d_s}{\sum_s m_s}$$

where m_s is the number of modules accessed in study session s, and d_s the session start date relative to due date. The summation is over all valid study sessions s by the student. A student's engagement with learning resources is estimated by $\bar{t}$, the average study time per module:

$$\bar{t} = \frac{\sum_s t_s}{M}$$

where t_s is the duration of study session s and M is the total number of unique modules accessed by the student prior to the assignment due date. We then divide students into four categories based on the combined population median of each variable, as shown in Figure 9 and listed in Table 1, and compare the fraction of students in each category between Fall 2018 and Fall 2019.

To answer RQ4, which addresses who primarily benefits from extra credit, we compare the difference in normalized scores (z-scores) on a midterm exam administered before the homework assignment, between students in each of the four categories. We hypothesize that in the absence of extra credit, students who have lower exam scores are also more likely to "cram" close to the due date and engage less with the learning resources. If those students were incentivized to distribute their work, then the gap in normalized score between different categories will be reduced. Likewise, the gaps would increase if only high-performing students were motivated by the extra credit.

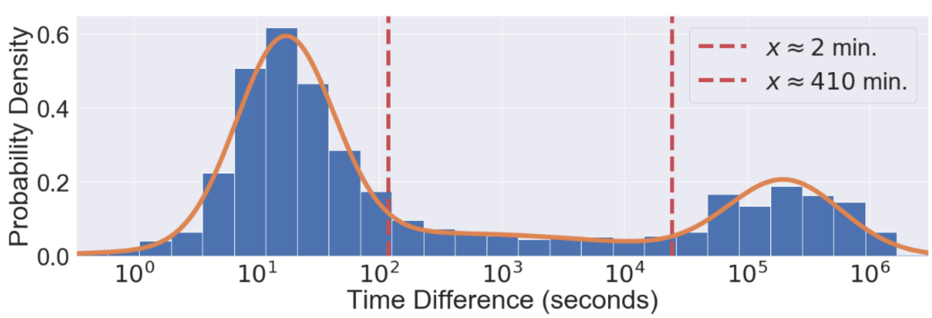

Figure 5: Distribution of Time Elapsed Between Viewer Enter and Viewer Leave Events. Plotted in orange is the three-component log-Gaussian mixture model fit. Dashed lines at $10^{2.08}$ s (2 min.) and $10^{4.39}$ s (410 min.) show the equiprobability boundaries between the three components.

3.3. Results

3.3.1. Identifying Valid Study Sessions (RQ1)

We plot in Figure 5 the distribution of time elapsed between all exit and enter events from the same student on the assigned modules for both semesters on a log scale. The data set is best fit with a three-component normal mixture model according to the Bayesian Information Criterion [50]. A reasonable interpretation is that the three components correspond to three types of behaviors: short interruptions less than a minute during study (such as advancing from one module to the next), medium length interruptions (such as answering a phone call or taking a lunch break) that happen less frequently, and the separation between two study sessions longer than 7 hours that likely stem from deliberate distribution of work over multiple days. Therefore, in the current study we adopt ε = 410 minutes as the minimum gap separating two study sessions.

We identify a total of 689 study sessions for Fall 2018 semester and 811 sessions for Fall 2019 semester, which are plotted in Figure 6 according to their duration and event densities. Of these, 79% reside within the boundaries: duration $\geq$ 120 seconds, event density $\geq 10^{-2.5}$ events per second, and event number $\geq$ 10 events. These boundaries are approximately tangential to the 0.15 contour line of the best fit 2D log-Gaussian distribution shown in the figure.

Therefore, we consider those 1191 study sessions within the boundaries valid study sessions. Among which, 542 were from Fall 2018 semester and 649 from Fall 2019 semester. The difference in proportion of valid sessions between years is not statistically significant.

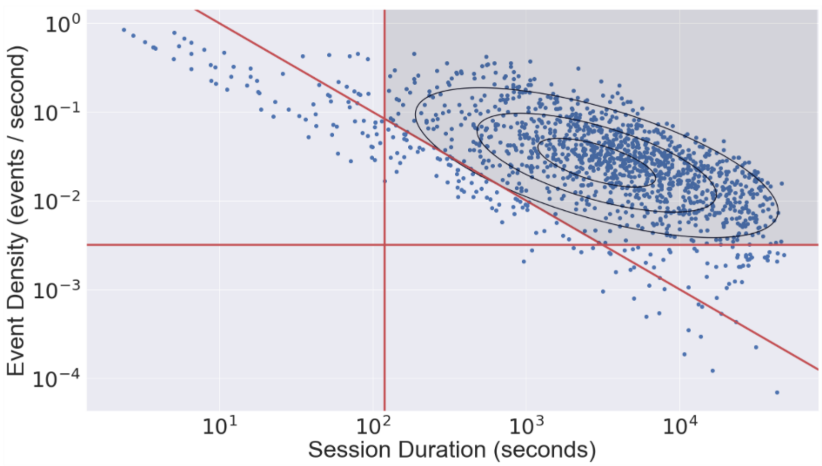

Figure 6: Event Density Versus Session Duration of All Sessions in Both Semesters (Fall 2018 and Fall 2019). Contour lines are based on a 2D log-normal fit and correspond to probability densities 0.15, 0.30, and 0.45. Note that this Gaussian fit is normalized in log-space and not for the axes as labeled. Red lines correspond to criteria for validity: duration ≥ 120 seconds, event density ≥ 10-2.5 events per second, and event number ≥ 10 events. The irregular shaded pentagon upper right contains “valid” sessions.

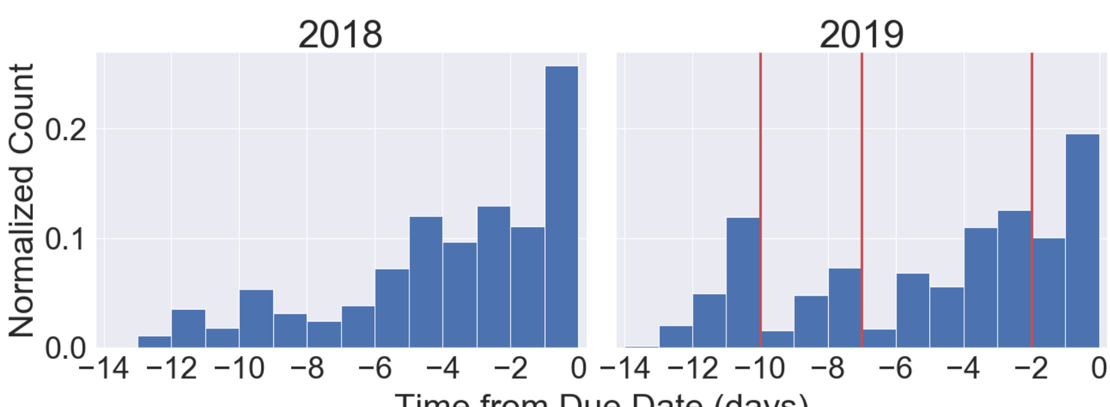

Figure 7: Distribution of Study Session Start Dates Without and With Extra Credit Incentives (2018 and 2019, Respectively). Sessions without extra credit are closer on average to the due date. Red lines indicate the extra credit deadlines in 2019: 10, 7, and 2 days before the due date. When extra credit is offered, we see clear peaks in graph corresponding to the first two extra credit deadlines: 10 and 7 days before the due date.

3.3.2. Impact on Session Start Dates and Study Time (RQ2 & RQ3)

In Figure 7 we show the distribution of start dates measured with respect to the due date for all valid sessions from each semester. The distributions differ significantly (Mann-Whitney U test, $p < 0.01$), with the Fall 2018 start dates grouped closer to the due date and Fall 2019 starting dates having two more peaks. The two additional peaks correspond to the expiration date of the first two "Treasure Trove" extra credit quizzes (10 and 7 days before due date).

To quantify the impact of the extra credit on session start dates and study time, we plot in Figure 8 the distribution of the two proxy variables, $\bar{d}$ and $\bar{t}$ for all students in each semester. As can be clearly seen, the distribution shifted left, indicating an earlier start for the assignment. To quantify this shift, we divide the data into four behavior categories based on the combined population medians for each variable, as listed in Table 1. In Figure 9A we plot the fraction of students in each category for both semesters. In the Fall 2019 semester, population in both LE and SE categories increased by almost 10 percentage points, whereas the population in LL and SL categories decreased by a similar amount. The differences between the two semesters are statistically significant ($p < 0.01$, extended Fisher's exact test). However, when we merge the data into two categories: long and short (as opposed to four), the test is no longer significant ($p > 0.5$). This indicates that there is no significant impact on the study time.

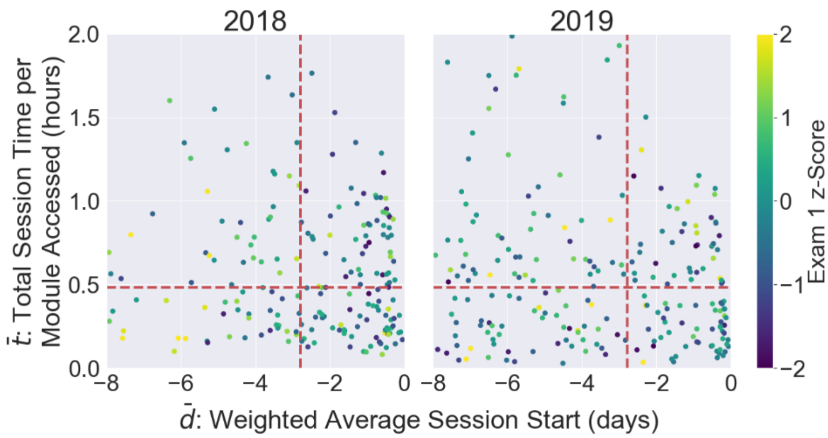

Figure 8: Total Session Time per Accessed Module ($\bar{t}$) Versus Weighted Average Session Start Date ($\bar{d}$), Colored by Normalized Exam 1 Score. A dotted line represents the combined median for each variable (lines have same position for both subplots). These lines divide the field into four behavior categories, starting upper right and moving counterclockwise: long late, long early, short early, and short late.

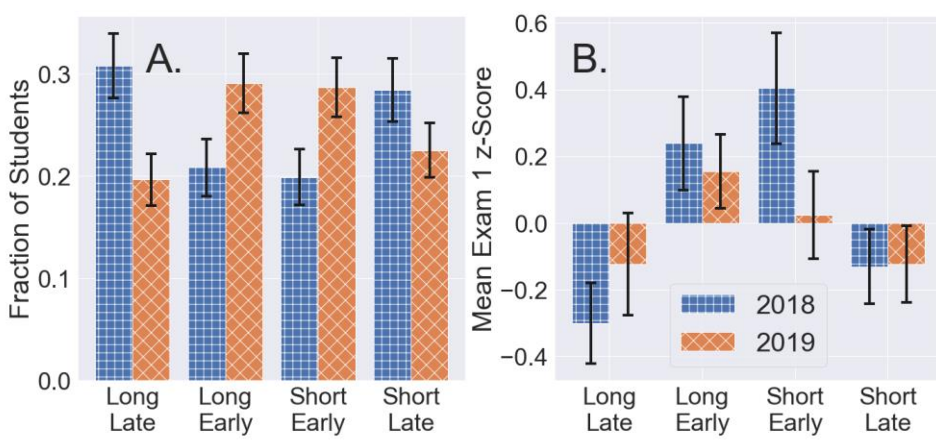

Figure 9: Fraction of Students and Exam 1 *z*-Scores for Students in Each Category and Year. Error bars extend one standard error from the mean.

Table 1: Definition of the Four Behavior Categories and Number of Students in Each.

Category Name	**Definition**	**No. of students (2018)**	**No. of students (2019)**
Long Late (LL)	$\bar{d} \geq -2.82$ days $\bar{t} \geq 0.482$ hours	65	48
Long Early (LE)	$\bar{d} < -2.82$ days $\bar{t} \geq 0.482$ hours	44	71
Short Early (SE)	$\bar{d} < -2.82$ days $\bar{t} < 0.482$ hours	42	70
Short Late (SL)	$\bar{d} \geq -2.82$ days $\bar{t} < 0.482$ hours	60	55

3.3.3. Work Distribution and Exam Performance (RQ4)

To answer RQ4, we plot the average exam scores for each of the four behavior categories in Figure 9B. In the Fall 2018 semester, there is a significant gap between the two early categories (SL and LL) and two late categories (ANOVA, $df = 3, F = 5.8, p < 0.01$), which is in agreement with some earlier studies as well as common belief that low-performing students are also prone to more cramming and procrastination [6,9]. In the 2019 semester, the differences between the four categories are no longer statistically significant (ANOVA, $df = 3, F = 1.1,\ p = 0.34$). This means one of two things: either extra credit induced lower performing students to work earlier, or the extra credit led students who worked earlier to perform worse on the exam. This suggests that the extra credit motivated students who scored lower on the midterm exam to move work earlier, placing more of them in the early categories. RQ4 asks whether extra credit benefits primarily high-performing students, “making the rich richer”. We see here that while the extra credit is more likely to be earned by high performers (since they are still overrepresented in the early categories) the extra credit benefits low-performing students by motivating them to complete the assignments earlier than they otherwise would have.

3.4. Discussion and Future Directions

In this study, we first analyze the distribution of gaps between study events and estimated the minimum duration of separation between two distinct study sessions (410 minutes). Compared to the 30-minute cutoff used in an earlier study [15], this longer gap duration better reflects students' deliberate choice to distribute work on different days, rather than taking a break or temporarily attending to a different task during a period of continued work. Compared to simpler metrics such as work done on each day, our clustering method can correctly identify study sessions that start just before midnight as a single session that spans two calendar days—a common occurrence with college students.

Based on the identified valid study sessions, we observe the following effects of extra credit on students' distribution of work: First, we see a significant reduction in cramming prior to the assignment due date, as more students begin work earlier in order to collect the extra credit. Several students commented in the course evaluation survey that the extra credit motivated them to start working early. Second, the reduction in cramming is *not* accompanied by a significant increase in the number of students with longer study time. We find no evidence that some students spent more time on assignments as a result of starting earlier. Finally, the reduce in score gap on an earlier exam between the four categories suggest that the change in work distribution is not limited to high-performing students, as more students who scored below average also started the assignment earlier compared to previous years.

Our results suggest that assigning extra credit for completing parts of the assignment early can be an effective method to encourage better work distribution for both high and low performing students. Since extra credit can be easily implemented on most existing learning management systems such as Canvas, it serves as a valuable tool for both instructors and students.

One of the most prominent questions that need to be answered in follow up studies is whether better work distribution leads to better learning gains from the assignments, and how extra credit affects this relationship. In addition, it will also be valuable to examine how the total weight of extra credit assigned impacts both its effects and its side effects. Finally, the analysis methods developed for this chapter will allow us to study how students work habits change over the semester.

CHAPTER 4:
IMPACT OF PLANNING PROMPT SURVEY

The contents of this chapter are adapted from results published in the journal *Physical Review: Physics Education Research* and as a work-in-progress in the *Proceedings of the Ninth ACM Conference on Learning @ Scale* [21,31]. See Appendix C for the cover pages of the publications.

4.1. Introduction

In the previous study, we examined the effectiveness of credit incentives in promoting early work and reducing procrastination, by offering small amounts of extra credit to students who complete portions of assigned homework in advance of the due date [30]. Our previous analysis showed that those assignments resulted in a small but measurable decrease in procrastination overall. A natural next step is to examine another instructional intervention to reduce procrastination behavior.

In a 2017 study conducted in three HarvardX massive open online course (MOOC) sections, Yeomans and Reich asked students at the beginning of the course to form a plan indicating where and when they would work on the course material, and what they would do to ensure they would carry out their plan. They found that merely asking students to write a plan, with no consequence on either the content or quality of the plan, increased course completion rate by 29% [2].

Inspired by the results of Yeomans and Reich, we designed a similar intervention with the intent to increase the number of students earning early completion extra credits and improve engagement with learning materials. The intervention was designed to be implemented at the beginning of two-week sequences of online homework modules in a college introductory level physics course, together with the extra credit incentives for early completion. The intervention first asked students three Likert-scale questions on how strongly they intended to earn each extra credit reward, followed by an open-ended question asking them to write down their plan for completing the sequence of learning modules assigned. Students who clicked the "submit" button on the survey were awarded 0.1% of regular course credit, regardless of their answer to the survey question.

We expected that our intervention would have a similar impact on college students as Yeomans and Reich observed on MOOC students. More specifically, we hypothesized that the planning prompt intervention will lead to (i) an increase in the average amount of extra credit earned by students, (ii) more students completing portions of the homework earlier compared to the due date, and (iii) a reduction in disengaged forms of interaction with the assignment [31].

To examine these hypotheses, we implemented the planning prompt intervention in three large (100–250 students) sections of the targeted college physics course. Two of those sections were taught in Spring 2022 and one was taught in the Fall 2021. Each of the three had a different instructor. Since randomly assigning a for-credit assignment to part of the students in a class is both technologically difficult and could disrupt instruction, we chose to assign the planning prompt intervention to the entire course section at the beginning of three different homework sequences—one in each section.

Most previous studies on procrastination categorized a subgroup of students as "procrastinators" by comparing their submission time to that of their peers. Agnihotri, Baker, and Stalzer defined their procrastination index based on the 75th percentile of submission times [10]. Nieberding and Heckler used cluster analysis to divide students into groups with similar procrastination behavior [29]. Sabnis et al. simply ranked students in order of assignment submission time and used rank divided by number of submissions as a measurement of procrastination [11]. However, any intervention to improve the work distribution will likely have an impact on more students than just the "procrastinators." In addition, members of such a "procrastination" group could change between different assignments, making longitudinal comparison of the impact more complicated. Therefore, in this study we use several metrics to evaluate the impact of our planning prompt intervention on the work distribution and level of engagement of the entire student population. We accomplish this by creating a multilinear model for each procrastination metric.

To eliminate the impact of extraneous factors such as differences in student population, semester, instructor, and instructional methods between different sections, we employ the analysis method of *difference in differences*, initially used by John Snow in his famous efforts to pinpoint the source of London's cholera outbreaks, and first formalized by economists Ashenfelter, Card, and Krueger [51]. Card later won the 2021 Nobel Prize in Economic Sciences for his work in formalizing natural experiments, laying the groundwork for the analysis presented in this study. Our implementation of the analysis method, adopted from Cunningham's *Causal Inference* [51], isolates the effect of the treatment from extraneous factors by constructing a linear regression model that includes data from before and after the treatment and is fitted to data from multiple semesters.

In the simplest case of a difference in differences design, data on a given observable, Y, is collected from two groups—control and treatment—over a period of time before and after the treatment took place. The observed value for the control group after the treatment took place can be written as $Y_c = C + \Delta_t$, in which C is the baseline observed value at the beginning of the observation, and Δ_t is the change of the variable over time independent of the treatment. The observed value for the treatment group during the same period can be written as $Y_T = T + \Delta_t + D$, where T is the beginning value of the observable for the treatment group, and D is the effect of the treatment itself. The basic assumption behind the differences in differences design is that $(C + \Delta_t) - (T + \Delta_t)$ remains constant over time. In other words, the trends in the observable of the two groups are parallel (also known as the *parallel trends hypothesis*), and the difference in the initial observed value is due to extraneous factors unrelated to the treatment. As a result, the treatment effect, D, can be separated from the extraneous factors by taking the differences in the two trends after the time of the treatment, thus the name "differences in differences." In this study, we assume that the change in observables between each week's assignments shows parallel trends, and the treatment effect can be detected as a difference in the trends specific to the week of the treatment and estimated using linear regression [51].

4.2. Experimental Setup

4.2.1. Instructional Conditions

Data used in our analysis was collected from four different sections of the same calculus-based college introductory level physics course: two sections from the Spring 2022 semester (22A, and 22B), one section from the Fall 2021 semester (21), and one semester from the Fall 2020 semester (20). Section 22A had 288 registered students, of which 24% were female and 40%

belonged to minority groups traditionally under-represented in STEM fields (URMs). Section 22B had 285 students, with 24% female and 34% URM. Section 21 had 121 students, with 34% female and 34% URMs. Section 20 had 254 students, with 26% female and 39% URMs. All four sections were taught using the same textbook and online homework assignments, following the same weekly schedule of homework release and due dates. Sections 22A, 21, and 20 were taught by the same instructor, and 22B by a different instructor. Section 20 was taught in an online only format due to campus closure resulting from the COVID-19 pandemic, while the other three sections were taught in a blended format with in-person lectures and online homework and learning resources.

4.2.2. Implementation of Online Homework and Extra Credit

Homework assignments in this study take the form of online learning module sequences implemented on the Obojobo learning modules platform and embedded in the Canvas Learning Management System. Each sequence consists of 7–16 learning modules which cover the material for one to two weeks of the course. Students must complete each module in the sequence in the given order. Each individual module consists of an instructional component containing text and practice problems, and an assessment component containing one or two problems focusing on a single concept or single type of problem. A module is considered complete when the student either answers all the problems(s) in the assessment component correctly or uses up all five attempts.

Data for the study was collected from five sequences of 7–16 modules each that were administered in all four sections. These are the 2nd, 3rd, 5th, 6th, and 8th of nine sequences assigned to students during the semester. The four remaining sequences each had complications making the data unsuitable for use: Sequence 1 is a brief review of vectors and had no associated extra credit; Sequence 4 was two half-sequences with the midterm occurring between them, each having only one extra credit opportunity; Sequence 7 was chosen to host another experiment in the 2022 semester; finally, sequence 9 was not yet created for the Fall 2020 semester.

For all five sequences included in our analysis, the module sequence was due two weeks after the first lecture on the topic—at 23:59 on the Saturday following the last lecture on the topic. In each section, students who completed a certain number of modules before a date earlier than the due date of the whole sequence could earn “treasure trove” extra credit points. For example, on the 5^{th} sequence, students who completed the first 3 of the 10 modules 10 days before the due date could earn 2 points of extra credit; those who completed the first 6 modules 7 days in advance could earn and additional 2 points of extra credit; and those who completed 9 modules 2 days ahead can earn another 3 points of extra credit. Each module is worth ten points before considering any extra credit. The five homework sequences had similar extra credit scheme, which is uniform across the four course sections.

Each extra credit opportunity was implemented as a for credit quiz on Canvas, embedded in between modules in the homework sequence. Each quiz was made accessible after students completed all the prior modules, and each quiz simply asked whether the student wanted the extra credit. The name “treasure trove” represented the idea that students can access those extra credit rewards after completing several learning modules. The due dates of the treasure trove quizzes were set individually according to the treasure trove schedule, and earlier than the sequence due date. A total of 47 extra credit points were made available to students and were worth 5% of the total course grade. Submission of module assessments after the due date received a 13-percentage-point deduction per day, rounded up to the nearest day. The late submission penalty also applied to the extra credit assignments. This setting was implemented in all four sections included in the current analysis.

4.2.3. Planning Prompt Intervention

The planning prompt survey was given to students at the beginning of each module sequence chosen to receive the intervention in the form of a Canvas graded survey. It contained one piece of informative text and four questions. The informative text outlined the treasure trove extra credit schedule of the upcoming sequence, while questions 1–3 asked students to indicate on a Likert scale how strongly they intended to earn each treasure trove, ranging from “Definitely Not” to “Definitely Yes.” Question 4 is largely modeled after Yeomans and Reich’s study [2], and asked students to formulate a plan by stating the following:

> *Research in education has shown that spacing homework over multiple days is beneficial for learning.*
>
> *We want to know about what plans you have to complete this sequence. Write down some of your plans to learn. For example, try to specify:*
>
> *1. When and where will you work on the modules?*
> *2. What specific steps you will take to ensure you complete the modules according to your schedule?*

Completion of the survey contributed to 4% of the homework sequence credit—or about 0.1% of the total course credit. The survey was only graded for completion, so submitting a blank survey would still have earned the full points. Students were not required to submit the survey prior to accessing the learning module sequence. The due date of the survey was set at the same time as the first extra credit due date.

Section 21 received the planning prompt survey before the 5th homework sequence of the semester. Section 22A received the survey prior to the 6th sequence of the semester, while 22B received it before the 8th sequence. Section 20 occurred before the beginning of the experiment.

4.3. Data Analysis

4.3.1. Measures of Procrastination Behavior

We choose five metrics to gauge the impact of the planning prompt intervention on students' completion of homework assignments. Three of these metrics measure the level of students' work distribution and procrastination: treasure trove extra credit points earned for early completion, average excess time of module completion, and standard deviation of module completion times. The other two metrics reflect a student's engagement with the homework: fraction of brief last assessment attempt and fraction of disengaged study state.

The treasure trove extra credit points earned by each student for each sequence is the most direct measure of how closely the students' work schedule aligned with the recommended schedule. Four of the five sequences offered 7 possible extra credit points, while sequence 2 offered a maximum of 8. To make sequence 2 directly comparable to the others, we multiply these extra credit scores by $7/8$—making all sequences have a maximum score of 7 points [31].

Average excess time per module is a measure of how early a student completed the modules on average. Excess time is defined as the amount of time between a student's first passing attempt on a given module (or last attempt if the student did not have a passing attempt) and the due date for the sequence. If a student completes a module after the due date by no more than 7 days, a negative excess time is recorded. If a student does not finish their work on a module within 7 days after the due date (or never attempts the module at all) we consider them to have completed it exactly 7 days after the due date—after which no credit can be earned—to reduce the impact on the average. This happens about 6% of the time [31].

The standard deviation of module completion times measures to what degree students spaced out their work. We use the sample standard deviation of the timestamps corresponding to the submission of a student's last attempt on each module in a given sequence. This directly measures how spread out a student's submissions are over time. We exclude the rare case where a student retakes an assessment they have already passed, so the "last attempt" is either the student's final attempt or the first attempt on which they earn a passing score, whichever comes first.

Brief last attempt fraction is a proxy for disengaged problem-solving behavior on the learning module assignments. A brief attempt is defined as an attempt on the assessment that is considered too short to be an authentic problem-solving effort. Previous studies on online problem solving and online learning modules have connected brief attempts to unintended problem-solving behavior such as guessing or answer copying [38,52]. Cutoff times between "brief" and "normal" attempts for each online learning module are independently estimated using a mixture-model fitting method developed in an earlier study [24]. The cutoff times range between 15 seconds to around 200 seconds for different modules, depending on the complexity of the problem in the assessment. We choose to focus on the last attempt on each module because previous studies revealed that some students intentionally make a brief first attempt in order to access the learning material [53]. Brief last attempts are more likely to indicate disengagement from the learning process and are more likely to happen among students who are under time pressure to finish the assignment. We find that brief last attempts are correlated with less excess time ($R = -0.14$). The brief last attempt fraction is simply the fraction of modules within a sequence that a student finished with the last attempt being brief. Modules that a student did not attempt are also counted as having a brief last attempt.

Finally, the disengaged study state fraction measures the likelihood of a student not actively engaged with learning from the modules. A student is categorized to be in a disengaged state under the following two situations (a) failed the assessment three or more times without having accessed the learning materials in between or (b) neither pass the module nor use up all their attempts. Previous studies have shown that the frequency of being in a disengaged state is correlated to lower overall learning outcome in the course [38]. Both disengaged study states and brief last attempts have been used in previous studies as indicators of lack of engagement with the learning materials [53].

4.3.2. Multilinear Modeling

For each metric outlined above, we create the following linear model adopting the technique outlined in Cunningham [51]:

$$\bar{x} = \beta_2 M_2 + \beta_3 M_3 + \beta_5 M_5 + \beta_6 M_6 + \beta_8 M_8 + \\ \beta_{21} S_{21} + \beta_{22A} S_{22A} + \beta_{22B} S_{22B} + \beta_T T \tag{1}$$

Here $\bar{x}$ is the estimate for one of the metrics on a given module sequence in a given section as a function of the sequence, the section, and whether the treatment is applied. M_2 through M_8 are module sequence dummy variables for the five sequences included in our analysis; they are equal to 1 when the observation comes from the corresponding module sequence and 0 otherwise. The three variables S_{21}, S_{22A}, and S_{22B} indicate the three sections that had a planning prompt intervention. Each of these variables is equal to 1 if the data comes from its corresponding section and zero otherwise. Data from 2020 is represented by all zeros in these variables. T is 1 for module sequences that received the planning prompt treatment and 0 otherwise. Each β is a fit coefficient that is the same for all sequences and all sections. The coefficient β_T is the estimated effect size of the planning prompt intervention. We determine these coefficients using a least-squares multilinear regression algorithm. We calculate the best fit for each of the five metrics using the above model and examine whether β_T is significantly different from zero for each metric. This is equivalent to testing for a nonzero effect size of our intervention.

For this study, we do not include any possible cross terms representing the interaction of multiple independent variables. This first-order approach imposes three implicit assumptions on the model. First, in the absence of a treatment intervention, each metric exhibits parallel trends across the four sections. That is, factors extraneous to the intervention—such as differences in student population, instructor, spring versus fall semesters, and online versus in person learning—on average impact each sequence uniformly across different semesters. Those uniform impacts of extraneous factors can then be captured by the terms β_{21}, β_{22A}, and β_{22B}. To the best of the instructors' knowledge, there were no extraneous factors, such as a natural disaster or a special event that would impact student learning behavior on a particular week in both semesters. All the sections were taught using the same curriculum following mostly the same schedule, and there were no known major shifts in student population between different semesters. Second, this model assumes that our planning prompt intervention only impacts student behavior on the sequence for which it is administered and is independent of the content of the sequence or the extraneous factors. In other words, the impact of the planning prompt intervention is largely the same across all three sections, which is reasonable given that the intervention is context independent. Third, the model assumes that the planning prompt treatment has a similar impact size each time it is applied. That is, the impact of the treatment is independent from the content and the section.

In addition, we also compare the primary model to two alternative models. The first alternative model assumes that the treatment had no impact on student behavior, created by setting $T = 0$ for all sequences in the original model. The second alternative model examines the hypothesis that the treatment may have a lasting effect on student behavior on the next homework sequence. This model is created by addition of a post-treatment term ($\beta_P P$) to the original model to represent the lingering effect of the intervention. $P = 1$ for the module sequence that immediately follows the sequence with the planning prompt survey, and $P = 0$ otherwise. We compare the two alternative models with the original model using an ANOVA that is modified to be robust against heteroskedasticity—the case where the variance of the dependent variable changes with the independent variable(s) (HC1 standard error computation) [54]. This test gives identical results as a conventional Type 1 ANOVA when variance is constant. Comparing the primary and first alternative models in this way provides another form of evidence of the treatment's effectiveness. Comparison to the second alternative tests whether the treatment has a lasting impact beyond its associated module sequence.

To ensure those in our treatment group were exposed to the intervention, students in the three treatment sections who did not submit the survey prior to its due date are excluded from our analysis. This is because the survey due date is the same as the due date for the first treasure trove extra credit opportunity. Across the three sections, 470 out of 621 students (76%) submitted the survey on time and were included in our analysis. There was no significant difference in the proportion of students who completed the survey on time between the class sections.

Since data from those students are excluded from all sequences in the analysis, self-selection effects would be present for all sequences, not just the treatment sequence. In other words, while students who answered the survey could be less likely to procrastinate, this effect would be the same for all sequences studied and be accounted for by the β_{21}, β_{22A}, and β_{22B} terms of Equation 1, separated from the treatment effect (β_T) term.

4.4. Results

In Figure 10 and Figure 11, we plot the means of each metric for each of the five sequences, connected by solid lines. The models for each metric are plotted with dotted lines. Purple crosses show the values that the model predicts would have been observed had we not introduced the planning prompt intervention ($\beta_T T \equiv 0$; note that this differs from the alternative model where the fitting is performed without the $\beta_T T$ term). Figure 10 includes data for extra credit points earned, average excess time per module, and standard deviation of module completion time. For these three metrics, we expect the planning prompt treatment to have a positive effect. Figure 11 shows data for a fraction of modules on which the student's last attempt was brief, and fraction of modules for which students show study patterns associated with disengagement. The expected effect is negative for these variables.

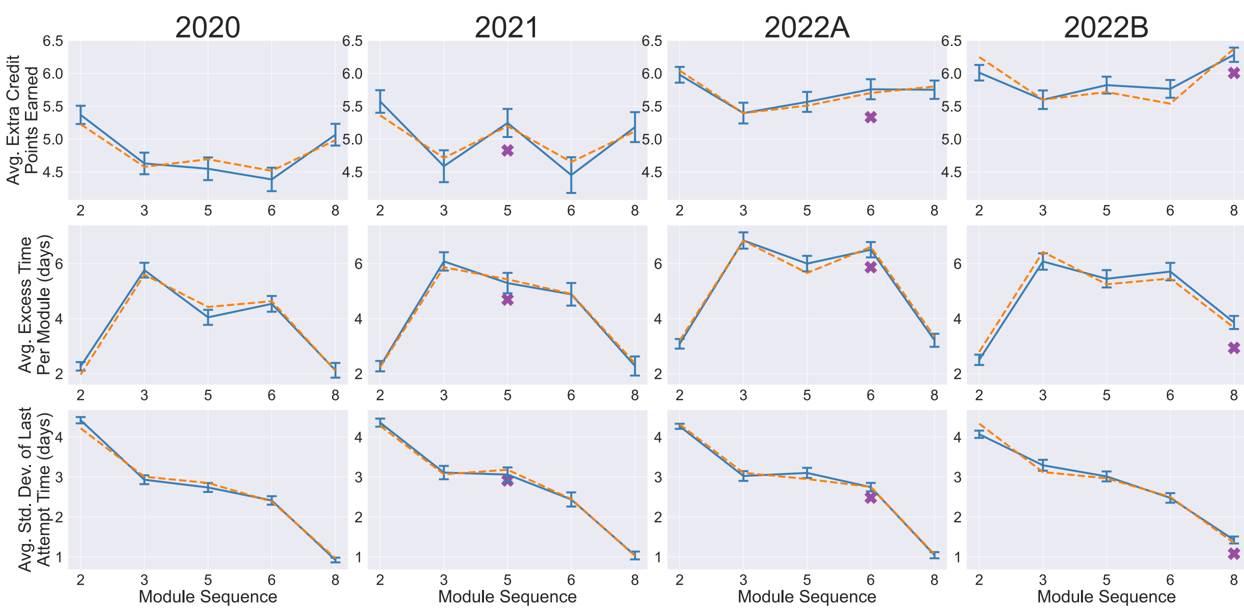

Figure 10: Observed and Modeled Work Distribution Metrics. Blue lines indicate observed averages, with error bars for 1 standard error of the mean. Orange dotted lines show the modeled values from Equation 1. Purple crosses show the model's prediction of each metric if the planning-prompt intervention had not been implemented ($\beta_T T \equiv 0$). The gap between the orange dotted line and the purple cross thus indicates the estimated effect size. The purple marks appear only for sequences that received the intervention, which sequence this is differs by class section.

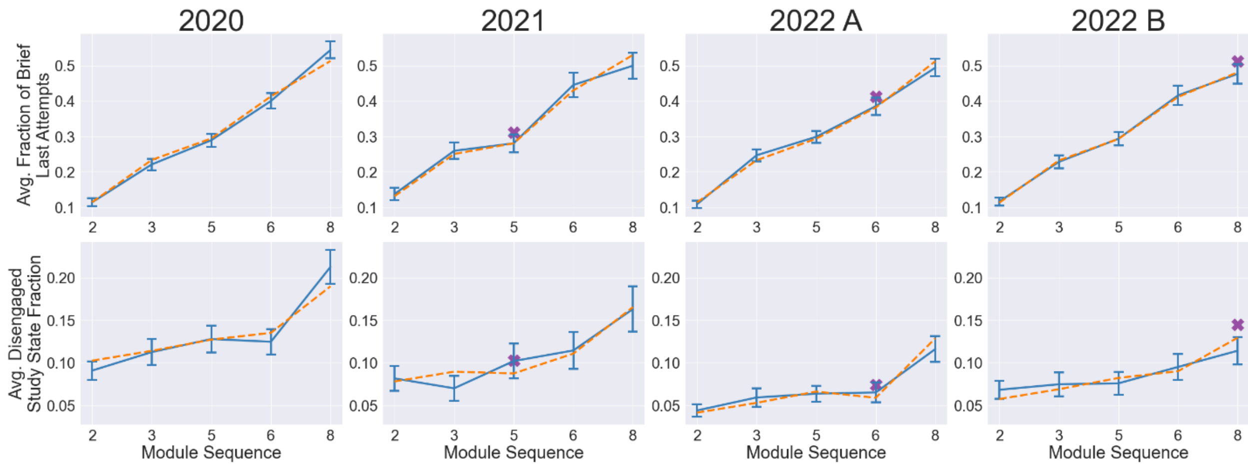

Figure 11: Observed and Modeled Engagement Metrics. Blue lines indicate observed averages, with error bars for 1 standard error of the mean. Orange dotted lines show the modeled values from Equation 1. Purple crosses show the model's prediction of each metric if the planning-prompt intervention had not been implemented ($\beta_T T \equiv 0$). The gap between the orange dotted line and the purple cross thus indicates the estimated effect size. The purple marks appear only for sequences that received the intervention, which sequence this is differs by class section.

Table 2: Model Coefficients for Various Procrastination Metrics. The coefficients listed correspond to those in Eq. (1); p values for treatment coefficients come from significance tests for differences from 0. The test used is a t test with modification to make it robust against heteroskedasticity (unequal variances).

Coefficient	**Extra Credit Points Earned**	**Excess Time per Module (days)**	**Std. Dev. of Last Attempt Completion Time (days)**	**Brief Last Attempt Fraction**	**Disengaged Study State Fraction**
$\boldsymbol{\beta_T}$ (Treatment)	**0.3690****	**0.7355*****	**0.2679****	**-0.0307**	**-0.0153**
$\boldsymbol{\beta_{21}}$ (Course Section 21)	0.1345	0.2672	0.0622	0.0169	-0.0245
$\boldsymbol{\beta_{22A}}$ (Course Section 22A)	1.0232	0.8247	0.1163	-0.0014	-0.0451
$\boldsymbol{\beta_{22B}}$ (Course Section 22B)	0.8172	1.2333	0.0979	-0.0009	-0.0612
$\boldsymbol{\beta_2}$ (Sequence 2 of 9)	5.2258	1.9729	4.2202	0.1143	0.1025
$\boldsymbol{\beta_3}$ (Sequence 3 of 9)	4.5767	5.6006	3.0088	0.2336	0.1140
$\boldsymbol{\beta_5}$ (Sequence 5 of 9)	4.6924	4.4263	2.8526	0.2944	0.1273
$\boldsymbol{\beta_6}$ (Sequence 6 of 9)	4.5155	4.6353	2.3873	0.4132	0.1353
$\boldsymbol{\beta_8}$ (Sequence 8 of 9)	4.9834	2.1146	0.9664	0.5134	0.1900

***Significant with p < 0.005 after Bonferroni adjustment. **Significant with p < 0.01 after Bonferroni adjustment. *Significant with p < 0.05 after Bonferroni adjustment. Only β_T (bold) tested for significance.

4.4.1. Model Fit

Table 2 gives the coefficients associated with the multilinear models for the five metrics and for which metrics the treatment term is significantly different from 0. The test is a modified *t*-test that is robust against heteroskedasticity—the case where variances of the dependent variable change with the independent variable(s) (HC1 standard error computation) [54]. The test used gives results identical to Student's *t* when variance is constant.

We find that for all three metrics on procrastination and work distribution (extra credit points earned, average excess time per module, and standard deviation of module completion time) the linear model predicts a significant positive treatment effect. On the sequence with the treatment, students earned an average of 0.369 more extra credit points (of 7 possible), completed modules on average 17.7 hours earlier, and spread their work out 6.3 hours more. The other two metrics that reflect students' level of engagement with the material: brief last attempt fraction and disengaged state fraction, do not show statistically significant impact at the $\alpha = 0.05$ level, though the direction of impact (negative) was as expected.

4.4.2. Comparison to Models Without Treatment Term

Using the ANOVA test mentioned above, we find that the model presented in Equation 1 fits the data significantly better than the alternative model without the treatment term (at the 0.05 level) for: extra credit points earned ($F = 9.43$, $p_A = 0.0108$), excess time per module ($F = 12.99$, $p_A = 0.0016$), and standard deviation of last attempt completion time ($F = 10.81$, $p_A = 0.0051$). The test is not significant for the remaining two metrics: brief last attempt fraction ($F = 3.87$, $p_A = 0.2463$) and disengaged study state fraction ($F = 1.92$, $p_A = 0.8319$). Here p_A is the adjusted *p*-value (Bonferroni adjustment).

4.4.3. Comparison to Models with Post-Treatment Term

We examine the possibility that the planning prompt treatment had an impact on the following sequence by adding a post-treatment term, $\beta_P P$, to the linear model. Here P is 1 for the sequence immediately after one that received the treatment and 0 otherwise. We find that for every metric, adding this independent variable term to the original model does not produce a significantly better model ($p > 0.29$; before p-value adjustment). β_P is also not significantly different from zero for any metric. This suggests that the effect of the planning prompt survey was short lived and only observable on the same sequence that immediately follows the intervention.

4.5. Conclusions and Discussion

We assigned students a planning prompt survey asking about their intention to earn extra credit for early progress on online homework and what their plan was for realizing that intention. The survey was assigned prior to three different online learning module homework sequences in three different sections. We find that for the homework sequence immediately after the survey, students who submitted the survey on time completed work roughly one day earlier, earned more extra credit, and spaced their work out more. The effects of the planning prompt were similar in nature to the MOOC study by Yeomans and Reich, but the magnitude of the impact was smaller. One possible reason is that MOOCs are low stakes with comparatively low completion rates. This allows for more room to improve with a planning prompt intervention. Another explanation could be that students in a MOOC have different self-regulatory ability as compared to students in traditional courses. A valuable follow up study could investigate the relation between students' self-regulatory skills and their likelihood of following through on their plans. Potential studies could further examine the generated plans for qualities such as length and specificity.

Similar to the observations of Fouh, Lee, and Baker [36], the impact of the planning prompt was limited to only the current sequence. In other words, after answering the planning prompt survey, a substantial fraction of students completed homework modules earlier and as a result earned more extra credit points on the following learning module sequence, while also spreading their work out more over time. However, data shows that they went back to their old working habits on the next module sequence. This means that a single planning prompt nudge can impact students' learning behavior over the course of two weeks but is not enough to enable students to form a habit of planning ahead and distributing their work. This is far from unexpected based on previous research [55]. We are currently planning on a follow up study to investigate whether repeated planning prompts could result in a long-lasting impact on students' cramming behavior.

Contrary to our expectations, we found that the survey intervention did not significantly change student engagement as measured by brief last attempt fraction and disengaged fraction (see 4.3.1 for precise definition). It is likely that for some of the students who would have waited until the due date to make many quick guesses on the module assessments, answering the planning prompt simply motivated them to guess earlier so that they can get extra credit. This also seems to suggest that while cramming and disengagement behavior are correlated, cramming is unlikely to be the main cause of disengagement under the instructional conditions of this study. Rather, both behaviors could be the outcome of ineffective self-regulation in some subset of students.

It is worth noting that our difference in differences model assumes that confounding factors including student population, instructor, and instructional environment impact different sequences uniformly (the parallel trend assumption). As mentioned earlier, those factors are captured in the terms β_{21}, β_{22A}, and β_{22B}, and separated from the treatment effect β_T. The model does not include cross terms such as $\beta_{21}\beta_1$ which would capture an extraneous impact specific to one of the sequences, such as a natural disaster, a holiday, or a special event. The authors are unaware of any such event over during the semesters covered by our data.

It is also unlikely that the observed treatment effect could be caused by such an extraneous event because the treatment is deliberately implemented on three different sequences. It is highly unlikely that three different extraneous events coincidentally occurred on all the sequences for which the intervention was implemented. The model also did not include cross-terms that indicate the treatment is more effective on certain sequences compared to others, or more effective in one semester than the others. That is also unlikely because the intervention does not depend on the content of the homework and is automatically presented to students via the learning management system.

While a quasiexperimental design may be less straightforward in determining causal inferences compared to a randomized experiment, it has the unique advantage of being much easier to implement in an authentic instructional setting. Comparing different instructional methods in different class sections is a common practice among many instructors, and the current study introduces a much more rigorous tool to analyze the outcome.

One shortcoming of the analysis presented here is the inclusion of only students who completed the survey, which restricts our conclusions to those who turned in the survey on time. Work is underway to include different levels of compliance with the intervention into the linear model. This will allow us to examine the impact of assigning the survey and the impact of students submitting it on time separately. We might see a treatment effect proportional to a student's level of compliance with the intervention. It is also plausible that students who do not complete surveys on time have different self-regulatory characteristics that would lead them to respond differently altogether to the intervention.

There are several valuable future directions that can follow from this work. First, it is worthwhile to investigate the relationship between the quality and specificity of a student’s plan, their self-regulatory skills, and their actual work distribution. Second, rather than using simple proxy measures for engagement, more sophisticated techniques such as process-mining and sequential cluster analysis could be employed to examine whether the nudging prompt impacted students’ learning strategy [56]. Third, from Figs. 1 and 2, we see that across all sections students space their work less and disengage more as the semester goes on. Those trends are very similar to the observation in an earlier paper in which students self-regulatory behavior declined as the semester progressed [53], perhaps due to both the increase in course content difficulty and the accumulated fatigue towards the end of the semester. It is valuable to explore possible interventions to reduce or reverse such a trend. Finally, future analysis could investigate whether and how interventions for reducing procrastination impact students with different backgrounds such as gender, race, and ethnicity.

CHAPTER 5:
CONCLUSION

This chapter concludes the dissertation by summarizing the findings of the included studies, discusses the place of these findings in the literature, and suggests opportunities for future research.

5.1. Findings

In the first study, we find that extra credit incentives significantly changed student behavior—strongly enough that the effect can clearly be seen in Figure 7. This change in student behavior is *not* associated with an increase in study time. The study also shows that the extra credit incentives narrowed the exam score gap between students who work early and those who work late. This means that the extra credit benefited under-performers as well as high achievers. We arrive at these results using a novel approach: using machine learning to determine an appropriate time cutoff to separate short study breaks from breaks between study sessions. This cutoff is found to be approximately 410 minutes for our data.

The second study evaluated the impact of a planning prompt survey on student procrastination behavior using the *differences in differences* analysis method. We found that, for students who submitted the survey on time, the intervention led to students completing their average module 18 hours earlier, earning 5% more extra credit, and spreading their progress out significantly more over time. The intervention is shown to have no impact on measures of engagement such as fraction of very short assessment submissions—indicative of guessing or cheating—and fraction of modules on which a student does not open the instructional content before taking the assessment multiple times. The study further found that the impacts of the survey did not persist beyond the module sequence immediately after it. Methodologically, a significant contribution is that this study utilizes a natural experiment technique. Natural experiment techniques are methods developed in the field of economics that correct for the inability to perform randomized controlled trials.

5.2. Connection to Existing Literature

Both studies advance the understanding of instructional techniques for combatting procrastination behavior, showing that both extra credit incentives and planning prompts can effectively help students to space their work. The first confirms a prior study by Ackerman and Gross that indicates incentivizing early progress leads to a reduction in procrastination behavior—though their work only measured student perception of rewards for starting early; we are unaware of any published work this century directly tying the inclusion of this kind of extra credit reward to early progress [5]. The second study extends the work of Yeomans and Reich by applying a similar intervention in the context of a sequence of online learning modules rather than to a massive open online course (MOOC) [2]. Unlike their study, we directly measure the impact of procrastination through metrics of work distribution and engagement.

A second major contribution of the studies is in their methodology. One previous study simply used calendar date of submission to track procrastination behavior [32]. Under this regime a student working from 11 P.M. to 1 A.M. would be misclassified as spacing their work over two days. Miyamoto et al. improved on this idea by using a 30-minute cutoff to separate log events into study sessions [15]. Our first study further refines this method by pinpointing an appropriate cutoff from the applicable data. In addition to the 30-minute cutoff being entirely arbitrary, our analysis shows that the data-derived cutoff is 14 times longer.

Our second study debuted the *difference in differences* method of analysis in the realm of physics education research. The linear modeling technique is used extensively in economics to make grounded claims about natural experiments [51]. Our application of the method to online learning data is new and may lead others to adopt similar techniques in the future.

5.3. Connection to Theory

Recall that, in Chapter 1, we adopt Zimmerman's three-phase model of self-regulated learning as a theoretical framework. In the model, student self-regulation takes place in a three-phase cycle: forethought, performance, and self-reflection. See 1.2 for details.

We assumed at the outset of this research that procrastination is primarily a failure in student self-regulation—reasonable given literature showing that students do not regularly delay beginning on an assignment as part of an intentional strategy [29]. In each of our studies, we anticipated that the intervention—extra credit or planning prompt surveys—would improve both work spacing and engagement by scaffolding students' self-regulated learning. Extra credit for early progress provided an incentive that fed student self-motivation (one of two parts of the forethought phase). The planning prompt survey was more direct: it demanded that they engage in both self-motivation and task analysis (the other part of forethought). It did this by asking students to indicate whether they intended to earn each of the extra credit incentives (motivation), then prompting students to create a plan for realizing their intention (task analysis). We anticipated that scaffolding self-regulation would in turn lead to increased engagement and reduced procrastination. This is not what we observe. We instead see a reduction in procrastination behavior without a corresponding increase in engagement.

A possible explanation for this outcome is that spaced work and engagement exist as separate self-regulatory cycles. Were this the case, it would make sense that we only scaffolded the former. Both interventions were targeted towards getting students to work earlier and in more sessions. The extra credit was awarded to students who completed the modules early, with no dependence on a student's engagement level. The survey similarly focused on where and when a student intended to complete the assignments, without prompting them to do so in a more engaged fashion.

A future study could test this explanation by performing a pair of experiments symmetrical to the ones presented here focused on disengagement instead of procrastination. First, one would offer extra credit for hitting some metric of engagement. One would then introduce a planning prompt asking students for their plan to remain engaged. Analyzing the impact of these interventions on both spaced study and engagement would elucidate the relationship between the two behaviors with respect to self-regulation.

5.4. Future Work

The work presented herein supplies several additional opportunities for future study.

Regarding the first study, follow up studies are needed to examine whether better work distribution leads to better learning gains from assignments, and how extra credit affects this relationship. In addition, it will also be valuable to examine how the total weight of extra credit assigned impacts both its effects and its side effects. The analysis methods developed for the study will also enable research on how students work habits change over the semester.

Regarding the second study, the responses students provided in the survey provide an untapped resource in understanding the link between intent, planning, and behavior as concerns procrastination behavior. To date, no work has yet examined the survey response data. Another avenue of extending the second study is the introduction of some kind of compliance metric so data can be included for students who did not fill out the survey, or who did it late. This would allow the results to be more strongly generalized. Additional metrics could also be added to understand the relationship of ethnicity, socioeconomic status, and gender to procrastination behavior and to the effectiveness of a planning prompt survey. The methods deployed in the study will allow future analysis of many kinds of uncontrolled experiments in the learning sciences.

Further, the studies call into question the assumption that procrastination is a significant cause of reduced engagement in learning, as posited by the proposed reduced engagement mechanism of the spaced study effect [12]. We found that reducing procrastination did not significantly impact engagement, at least not based on our metrics. This suggests that procrastination and insufficient engagement in learning might be correlated by not causal. In other words, they are both reflections of underlying students' internal states, such as motivation and goal orientation. Future research is needed to investigate how to change those underlying traits.

Finally, the analysis methods of the two studies could potentially be combined and improved in future research. Number and distribution of study sessions as used in the first study could be analyzed more robustly through the use of multilinear models such as those used in the second. Also, more precise measurements of students' level of engagement can be conducted. For example, Zhang, Taub, and Chen identified different learning tactics using process mining of students' trace data while completing our learning modules [53]. An immediate future work opportunity is to study the differences in the adoption of learning strategies among students who procrastinate versus those to space out their work. Furthermore, one could study how the interventions change students' adoption of learning tactics utilizing a natural experiment design.

APPENDIX A:
DESCRIPTION OF ONLINE LEARNING MODULE DEVELOPMENT

DESCRIPTION OF ONLINE LEARNING MODULE DEVELOPMENT

The online learning modules from which data is taken for this work were painstakingly developed by our team at the UCF physics department. The design process for a module began by identifying learning goals for the module. Once this was accomplished, we began to create the assessment component of the module.

First, we devised one to three questions which assess mastery of the learning goals. For each assessment problem we created three different isomorphisms. Two problems are isomorphic is they have the exact same solution path and difficulty. These could be rotations of variables—exchanging knowns and unknowns—or three different question contexts with different numbers. We then created a set of "distractors"—three to six incorrect answers chosen to match plausible student errors that appear as incorrect choices on the multiple-choice assessment—for each isomorphic form of each assessment problem. We completed the assessments by using Inkscape to create figures to match the context of each form of each assessment question.

Once the assessment component was complete, we created the instruction and practice components as a single unit. The instructional text was, to the extent possible, adapted from the text of *University Physics Volume 1* from OpenStax CNX [39]. If necessary, we would write the instructional text from scratch. Practice problems were either placed in line with the text or at the end of the text as appropriate for the specific module. At the end each module, we placed the first set of assessment question(s), the ones a student sees on their first assessment attempt. Since students were required to make one assessment attempt before opening the instructional materials, they were guaranteed to have seen the problems once before encountering them as practice questions. Finally, each practice problem received choice feedback and a solution. The feedback either praised the student for being correct or gave them specific hints on what they may have done incorrectly. For example, “You used sine instead of cosine. Try again.” The solution gave a step-by-step guide for solving the practice problem, accessible only after a student selected an option and submitted their answer.

Note that each assessment question had three isomorphic forms while the students were allowed five attempts at the assessment. We stopped at three to save manpower, and discouraged students from simply submitting three attempts to arrive at the first set again (for which they had solutions in the practice section) by assigning a ten-percentage-point penalty to the grade earned on fourth and fifth attempts.

Once all this content was generated and inserted into the module, we began the process of quality control. Every word was checked by at two sets of eyes, and each version of each assessment was completed by at least three different people to ensure that the answers were correct. During this stage we also checked for consistency in style and formatting, following an *ad hoc* style guide I personally drafted as situations arose that needed a consistent style. This document is reproduced in Appendix B.

During the entire process of module creation, our team stayed in consistent communication via Microsoft teams meetings and chats. Version control was achieved using a git repository.

In this way we produced nine sequences of 6–12 high-quality mastery-based online learning modules from 2018–2021. These modules entirely replaced both the homework and the textbook for Physics 2048 at no direct cost to the student (unlike the previous textbook and online homework platform).

APPENDIX B:
ONLINE LEARNING MODULE STYLE GUIDE

ONLINE LEARNING MODULE STYLE GUIDE

1. When not contradicted here, use the APS style guide: https://cdn.journals.aps.org/files/styleguide-pr.pdf
2. English
    a. To do: Header levels and capitalization
    b. Capitalization:
        i. Use title case for specific physical laws: “Newton’s Second Law of motion”. M is lower case because we decided it wasn’t part of the name of the law. Similarly, use “Step 5” for a specific named step.
        ii. People’s names are, of course, in title-case. Units named after people are always lower case when written out, but capital in abbreviation. “The newton is named for Newton. The symbol for the newton is \mathrm{N}.”
            1. The degree Celsius is an edge case, as the “d” in “degrees” is considered the first letter.
            2. The liter is abbreviated “L” to avoid confusion with I or 1.
            3. This section (2.b.ii) should be the same as what SI/ISO recommends.
    c. Misc.
        i. Use the Oxford/serial comma. We aren’t savages.
        ii. Words like “rewrite” and “massless” do not need to be hyphenated. This is largely a judgement call. Examples of cases: “x axis” (no hyphen); “free-body” (hyphen); “non-equilibrium” (hyphen);

3. LaTeX

   a. Use latex for all variables, vectors, and unit abbreviations, even if the only thing in the latex equation is “\mathrm{N}”. Numerals which represent physical quantities should always be in latex as well.

   b. Units:

      i. General form for units is “10 \; \mathrm{ N \, m / s^2}”.

      ii. The “\;” space after the number is present for all units (including the % sign and degrees Celsius), except for plane angle degrees, minutes, and seconds.

      iii. To do: standard order for units to appear (“N m” not “m N”)

   c. Vectors:

      i. Use arrow notation for all non-unit vectors.

      ii. Unit vectors get a hat, no arrow. Unit vectors i and j retain their dot under the hat. Currently, no space (such as \,) is inserted between a numeric value and a unit vector.

      iii. Bold vectors in addition to hats and arrows inside figures if clearer or more convenient.

      iv. When vectors are not bold, they are italicized (this is default).

   d. Multiplication signs:

      i. See pg 21 of the APS style guide. Exception: between two numeric values use \times. Treat trig functions as numerical values when a numerical argument is given. ex. [10 \;\mathrm{N} \times \sin(60 \degree)]

   e. Trig Functions: \sin, \cos, \tan, \sec, etc makes spacing look nice and removes italics.

4. Figures

    a. Any letter choices should have *just* the letter i.e. “A” as opposed to “A.”

5. Code and ID conventions

    a. Each module sequence is assigned a short—1–2 letter—code unique to that module sequence. Ex: M for momentum, RK for rotational kinematics. Modules in the sequence are tagged with a 2 digit number starting at 01 for the first in the sequence. Ex: RK03. This tag appears at the beginning of each ID so it occurs in no more than one module. In this document “XX01” will stand for the module tag. The module tag is included at the beginning of the module title.

    b. IDs are built by adding a tag for each nested layer decreasing in size until reaching the object to be identified. Separate these tags with periods. Example: the *text* for *choice 3* of the *second practice problem* on the *first page* of the *module* is labeled with the ID: “XX01.pg1.Q2.choice3.content”.

    c. Each page is tagged by “pg1”, “pg2”, … in order. The assessment page gets identified by “assessment”. Thus page 5 has ID “XX01.pg5”.

    d. Questions

        i. Assessment Page:

            1. All questions in the assessment sit inside a master question bank. “XX01.assessment.QBmaster” is the ID of the master question bank.

            2. If each assessment attempt will have four graded questions, there will be four question banks within QBmaster labeled QB1, QB2, QB3, and QB4. (XX01.assessment.QBmaster.QB1 …)

3. In a similar way, for each desired ungraded (survey) question, there is a question bank within QBmaster labeled surveyBank1, surveyBank2, …

4. For each version of the assessment, there are questions Q1, Q2, … in each of the question banks. “XX01.assessment.QBmaster.QB2.Q3” is the second graded question on the third attempt assessment. “XX01.assessment.QBmaster.surveyBank1.Q2” is the first survey question that appears on the second attempt assessment.

ii. Non-Assessment Pages: on non-assessment pages, there are no question banks. Questions are labeled in the order they appear on the page: Q1, Q2, … Example: “XX01.pg1.Q2”

iii. Choice options for questions are labeled choice1, choice2, … regardless where the question appears.

iv. It is preferable to make the correct answer choice1 and enable shuffle for the choices if reasonably practical.

APPENDIX C:
COVER PAGES OF PREVIOUSLY PUBLISHED WORKS

The impact of extra credit incentives on students' work habits when completing online homework assignments

Zachary Felker and Zhongzhou Chen
Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando, Florida, 32816-2385

The current paper examines the effectiveness of offering a small amount of extra credit as an incentive to encourage proper work distribution and reduce procrastination and cramming among college students completing introductory physics homework assignments in the form of online learning modules. Students' distribution of work over time is systematically measured by clustering clickstream log events into study sessions according to a cutoff determined empirically using mixture model analysis. Significantly more study sessions are initiated well before the assignment due date when extra credit is offered compared to data from a previous semester. Using two proxy variables designed to capture the distribution and duration of work, we found that in addition to starting the assignments earlier, students also spent a longer time on the assignments. Finally, the benefit of extra credit in encouraging work distribution is not limited to high-performing students, as shown by a reduction in score gap between early and late starters on a midterm exam administered prior to the release of the homework assignment.

2020 PERC Proceedings edited by Wolf, Bennett, and Frank; Peer-reviewed, doi.org/10.1119/perc.2020.pr.Felker

Using a Planning Prompt Survey to Encourage Early Completion of Homework Assignments

Zachary Felker
Department of Physics
University of Central Florida
Orlando, Florida USA
felker@knights.ucf.edu

Zhongzhou Chen
Department of Physics
University of Central Florida
Orlando, Florida USA
zhongzhou.chen@ucf.edu

ABSTRACT
In an earlier study we showed that small amounts of extra credit offered for early progress on online homework assignments can reduce cramming behavior in introductory physics students. This work expands on the prior study by implementing a planning prompt intervention inspired by Yeomans and Reich's similar treatment. In the prompt we asked students to what degree they intended to earn extra credit offered for early work on the module sequence, and what their plan was to realize their intentions. The survey was assigned for ordinary course credit and due several days before the first extra credit deadline. We found that students who completed the prompt earned on average 0.6 more extra credit points and completed the modules an average of 1.1 days earlier compared to a previous semester. We detect the impact of the survey by creating a multilinear model based on data from students exposed to the intervention as well as students in a previous semester. Data from five homework sequences are included in the model to account for differences between the two semesters that cannot be attributed to the planning prompt intervention.

CCS CONCEPTS
•Applied computing~Education~E-learning

•Applied computing~Education~Computer-assisted instruction

KEYWORDS: self-regulated learning, SRL, planning prompt, extra credit, online learning

ACM Reference format:

Zachary Felker and Zhongzhou Chen. 2022. Using a Planning Prompt Survey to Encourage Early Completion of Homework Assignments. *In Proceedings of the Ninth ACM Conference on Learning @ Scale (L@S '22). June 1–3, 2022, Roosevelt Island, NY.* ACM, New York, NY. https://doi.org/10.1145/3491140.3528297

ACM 978-1-4503-9158-0/22/06...$15.00
https://doi.org/10.1145/3491140.3528297

1 INTRODUCTION
It is well documented that college students often procrastinate on assignments and, as a result of such procrastination, cram against due dates or before exams. Previous studies, as well as our own research, show that procrastination is associated with diminished academic performance [1,4,6,7]. A likely cause for this negative impact on academic performance is that students who cram due to procrastination do not have enough time to properly engage with the intended learning process. Accordingly, in this paper we focus on early work as the antithesis of procrastination.

In a previous paper, we showed that offering small amounts of extra credit to students who complete portions of assigned work in advance of the due date is associated with a small but measurable decrease in measured procrastination behavior. Yet a significant fraction of students did not take advantage of the extra credit opportunity [4].

In a study conducted on massive open online courses, Yeomans and Reich found that merely asking students to write a plan for how they would distribute their work dramatically increased the course completion rate, regardless of the contents of the plan. In the study, students were asked at the beginning of the course to indicate where and when they would work on the course material, and what they would do to ensure they would carry out their plan [9].

This work applies a similar intervention to a two-week sequence of online homework modules in a large, mixed mode, introductory physics course in addition to the extra credit we evaluated previously. We added Likert scale questions at the beginning of the planning prompt survey asking the student how strongly they intended to earn each extra credit reward for early completion of modules. We then asked them to write down their plan for completing the sequence of modules—and earning the extra credit if they intended on doing so.

We predicted that the intervention would prompt students to form concrete intent about whether they would seek the extra credit, and then create a plan for achieving their intentions. *We hypothesize that the planning prompt thus will lead students to work on the module sequence earlier and as a result earn more of the extra credit offered.*

Reducing procrastination on introductory physics online homework for college students using a planning prompt intervention

Zachary Felker [*] and Zhongzhou Chen [†]

Department of Physics, University of Central Florida, 4111 Libra Drive, Physical Sciences Building 430, Orlando, Florida 32816-2385, USA

(Received 15 December 2022; accepted 9 February 2023; published 30 March 2023)

We examine the effectiveness of a planning prompt intervention to reduce procrastination on online homework for college students. The intervention asked students to indicate their intention to earn small amounts of extra credit for completing assignments earlier and form a plan to realize their intentions. Students' learning behavior is measured by five data metrics collected from gradebook and student interaction logs from four sections of the same college level physics course, including three metrics that capture how students space their work on assignments and two that measure the level of student engagement when completing assignments. To separate the impact of extraneous factors from the treatment effect, we employ a "difference in differences" method—initially developed in economics—and construct multilinear models for each of the five metrics. Our models show that by simply asking students to form a plan prior to the assignment, students on average earned 5% more extra credit, completed homework significantly earlier, and spread out their work significantly more. However, the intervention did not significantly change students' level of engagement with the learning materials, nor did it change students' work distribution on the next assignment.

DOI: 10.1103/PhysRevPhysEducRes.19.010123

I. INTRODUCTION

Many studies have shown that a significant fraction of college students often procrastinate on assignments and cram against due dates or before exams. Studies, including our own, have also shown that procrastination is often associated with reduced academic performance [1–4]. For example, Agnihotri, Baker, and Stalzer found that habitual procrastination as measured by a procrastination index is associated with a 21-fold higher risk of failing the course [5]. Studies have also shown that procrastination is associated with maladaptive learning strategies and reduced levels of self-regulation in online learning [6,7]. Sabnis *et al.* also found that procrastination is more common among males, racial minorities, and first-generation students [8]. Nieberding and Heckler further showed that procrastination behavior is directly associated with nonexam grades, and that a large majority of students who procrastinate did not report planning to delay work on assignments [9]. Therefore, effective methods to discourage procrastination and encourage early work could have substantial positive impact on student learning. In an earlier study, we examined the effectiveness of credit incentives in promoting early work and reducing procrastination, by offering small amounts of extra credit to students who complete portions of assigned homework in advance of the due date [1]. As explained in more detail in Sec. II B, these so-called "treasure trove" extra credit assignments encourage students to break a sequence of 7–11 online homework modules into 2–3 portions and finish each portion earlier than the due date. Students who followed the suggested schedule can open the treasure trove assignments and get a small amount of extra credit. Our previous analysis showed that those assignments resulted in a small but measurable decrease in procrastination overall.

Aside from credit incentives, another widely used strategy to influence student behavior in blended and online learning settings is nudging [see Damgaard and Nielsen for a more comprehensive review] [10]. In particular, nudging in the form of goal setting activities or reminder emails or texts have been frequently used to fight procrastination and distraction. For example, Patterson found that asking students to set a goal for limiting distracting internet time increased course engagement and completion in a massive open online course (MOOC) [11]. Huang *et al.* examined multiple forms of email "calls to action" and found that while descriptive norms lead to reduced procrastination, deadline reminders can actually backfire and result in increased dropouts [12]. Fouh, Lee, and Baker found that

[*] felker@knights.ucf.edu
[†] zhongzhou.chen@ucf.edu

APPENDIX D:
IRB LETTERS

UCF

UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

EXEMPTION DETERMINATION

October 28, 2019

Dear Zhongzhou Chen:

On 10/28/2019, the IRB determined the following submission to be human subjects research that is exempt from regulation:

Type of Review:	Initial Study, Category
Title:	Searching for better online instructional design to develop STEM expertise
Investigator:	Zhongzhou Chen
IRB ID:	STUDY00000994
Funding:	Name: National Science Foundation (NSF), Grant Office ID: 1066167, Funding Source ID: 1845436
Grant ID:	1066167;

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made, and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request so that IRB records will be accurate.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Gillian Bernal
Designated Reviewer

UNIVERSITY OF CENTRAL FLORIDA

Physics Department

4111 Libra Drive
Physical Sciences Bldg. Room 430
Orlando, FL 32816-2385

July 7, 2023

To whom it may concern,

The research presented in the dissertation entitled "Due Tomorrow, Do Tomorrow: Measuring and Reducing Procrastination Behavior Among Introductory Physics Students in an Online Environment" by Zachary Felker was conducted under IRB STUDY00000994 according to approved procedures.

Zhongzhou Chen
Assistant Professor of Physics
College of Sciences
University of Central Florida
Phone: (217) 721-8411
e-mail: Zhongzhou.Chen@ucf.edu

REFERENCES

[1] L. L. Ainsworth, "Self-paced instruction: an innovation that failed," Teach. Psychol. **6**, 42 (1979).

[2] M. Yeomans and J. Reich, "Planning prompts increase and forecast course completion in massive open online courses," in *ACM Int. Conf. Proceeding Ser.* (Association for Computing Machinery, 2017), pp. 464–473.

[3] O. T. Anderson and R. A. Artman, "A self-paced, independent study, introductory physics sequence—description and evaluation," Am. J. Phys. **40**, 1737 (1972).

[4] G. D. Putt, "Testing the mastery concept of self-paced learning in physics," Am. J. Phys. **45**, 472 (1977).

[5] D. S. Ackerman and B. L. Gross, "My instructor made me do it: task characteristics of procrastination," J. Mark. Educ. **27**, 5 (2005).

[6] S. H. McIntyre and J. M. Munson, "Exploring cramming: student behaviors, beliefs, and learning retention in the principles of marketing course," J. Mark. Educ. **30**, 226 (2008).

[7] F. M. Sirois and R. Kitner, "Less adaptive or more maladaptive? a meta-analytic investigation of procrastination and coping," Eur. J. Pers. **29**, 433 (2015).

[8] J.-C. Hong, Y.-F. Lee, and J.-H. Ye, "Procrastination predicts online self-regulated learning and online learning ineffectiveness during the coronavirus lockdown," Pers. Individ. Dif. **174**, 110673 (2021).

[9] S. A. Nonis and G. I. Hudson, "Performance of college students: impact of study time and study habits," J. Educ. Bus. **85**, 229 (2010).

[10] L. Agnihotri, R. S. Baker, and S. Stalzer, "A procrastination index for online learning based on assignment start time," in *Proc. 13th Int. Conf. Educ. Data Mining, EDM 2020*, edited by A. N. Rafferty, J. Whitehill, V. Cavalli-Sforza, and C. Romero (2020), pp. 550–554.

[11] S. Sabnis, R. Yu, and R. F. Kizilcec, "Large-scale student data reveal sociodemographic gaps in procrastination behavior," in *Proc. Ninth ACM Conf. Learn. @ Scale* (ACM, New York, NY, USA, 2022), pp. 133–141.

[12] J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham, "Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology," Psychol. Sci. Public Interes. **14**, 4 (2013).

[13] H. P. Bahrick and L. K. Hall, "The importance of retrieval failures to long-term retention: a metacognitive explanation of the spacing effect," J. Mem. Lang. **52**, 566 (2005).

[14] S. K. Carpenter, "Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval.," J. Exp. Psychol. Learn. Mem. Cogn. **35**, 1563 (2009).

[15] Y. R. Miyamoto, C. A. Coleman, J. J. Williams, J. Whitehill, S. Nesterko, and J. Reich, "Beyond time-on-task: the relationship between spaced study and certification in moocs.," J. Learn. Anal. **2**, 47 (2015).

[16] M. K. Hartwig and J. Dunlosky, "Study strategies of college students: are self-testing and scheduling related to achievement?," Psychon. Bull. Rev. **19**, 126 (2012).

[17] B. J. Zimmerman, "Becoming a self-regulated learner: an overview," Theory Pract. **41**, 64 (2002).

[18] E. Panadero, "A review of self-regulated learning: six models and four directions for research," Front. Psychol. **8**, 1 (2017).

[19] L. L. Lamwers and C. H. Jazwinski, "A comparison of three strategies to reduce student procrastination in psi," Teach. Psychol. **16**, 8 (1989).

[20] R. A. Reiser, "Reducing student procrastination in a personalized system of instruction course," Educ. Commun. Technol. **32**, 41 (1984).

[21] Z. Felker and Z. Chen, "Reducing procrastination on introductory physics online homework for college students using a planning prompt intervention," Phys. Rev. Phys. Educ. Res. **19**, 010123 (2023).

[22] Z. Chen, G. Garrido, Z. Berry, I. Turgeon, and F. Yonekura, "Designing online learning modules to conduct pre- and post-testing at high frequency," in *2017 Phys. Educ. Res. Conf. Proc.*, edited by L. Ding, A. Traxler, and Y. Cao (American Association of Physics Teachers, Cincinnati, OH, 2018), pp. 84–87.

[23] M. Taub, A. M. Banzon, T. Zhang, and Z. Chen, "Tracking changes in students' online self-regulated learning behaviors and achievement goals using trace clustering and process mining," Front. Psychol. **13**, (2022).

[24] Z. Chen, "Measuring the level of homework answer copying during covid-19 induced remote instruction," Phys. Rev. Phys. Educ. Res. **18**, 010126 (2022).

[25] J. Park, K. Denaro, F. Rodriguez, P. Smyth, and M. Warschauer, "Detecting changes in student behavior from clickstream data," Proc. Seventh Int. Learn. Anal. Knowl. Conf. - LAK '17 21 (2017).

[26] F. Rodriguez, S. Kataoka, M. Janet Rivas, P. Kadandale, A. Nili, and M. Warschauer, "Do spacing and self-testing predict learning outcomes?," Act. Learn. High. Educ. 146978741877418 (2018).

[27] D. T. Seaton, G. Kortemeyer, Y. Bergner, S. Rayyan, and D. E. Pritchard, "Analyzing the impact of course structure on electronic textbook use in blended introductory physics courses," Am. J. Phys. **82**, 1186 (2014).

[28] G. Kortemeyer, "Work habits of students in traditional and online sections of an introductory physics course: a case study," J. Sci. Educ. Technol. **25**, 697 (2016).

[29] M. Nieberding and A. F. Heckler, "Patterns in assignment submission times: procrastination, gender, grades, and grade components," Phys. Rev. Phys. Educ. Res. **17**, 013106 (2021).

[30] Z. Felker and Z. Chen, "The impact of extra credit incentives on students' work habits when completing online homework assignments," in *2020 Phys. Educ. Res. Conf. Proc.*, edited by S. Wolf, M. B. Bennett, and B. W. Frank (American Association of Physics Teachers, Virtual Conference, 2020), pp. 143–148.

[31] Z. Felker and Z. Chen, "Using a planning prompt survey to encourage early completion of homework assignments," in *Proc. Ninth ACM Conf. Learn. @ Scale* (ACM, New York, NY, USA, 2022), pp. 446–450.

[32] T. W. Cavanaugh, M. L. Lambkin, and H. (Helen) Hu, "USING a generalized checklist to improve student assignment submission times in an online course," Online Learn. **16**, 39 (2012).

[33] M. T. Damgaard and H. S. Nielsen, "Nudging in education," Econ. Educ. Rev. **64**, 313 (2018).

[34] R. W. Patterson, "Can behavioral tools improve online student outcomes? experimental evidence from a massive open online course," J. Econ. Behav. Organ. **153**, 293 (2018).

[35] N. Huang, J. Zhang, G. Burtch, X. Li, and P. Chen, "Combating procrastination on massive online open courses via optimal calls to action," Inf. Syst. Res. **32**, 301 (2021).

[36] E. Fouh, W. Lee, and R. S. Baker, "Nudging students to reduce procrastination in office hours and forums," in *25th Int. Conf. Inf. Vis.* (IEEE, 2021), pp. 248–254.

[37] Z. Chen, S. Lee, and G. Garrido, "Re-designing the structure of online courses to empower educational data mining," in *Proc. 11th Int. Conf. Educ. Data Min.*, edited by K. E. Boyer and M. Yudelson (Buffalo, NY, 2018), pp. 390–396.

[38] Z. Chen, M. Xu, G. Garrido, and M. W. Guthrie, "Relationship between students' online learning behavior and course performance: what contextual information matters?," Phys. Rev. Phys. Educ. Res. **16**, 010138 (2020).

[39] W. Moebs, S. J. Ling, and J. Sanny, "University physics volume 1," *University Physics Volume 1* (OpenStax, Houston, Texas, Texas, 2016).

[40] F. Rodriguez, R. Yu, J. Park, M. J. Rivas, M. Warschauer, and B. K. Sato, "Utilizing learning analytics to map students' self-reported study strategies to click behaviors in stem courses," in *Proc. 9th Int. Conf. Learn. Anal. Knowl.* (ACM, New York, NY, USA, 2019), pp. 456–460.

[41] Z. Chen, K. M. Whitcomb, and C. Singh, "Measuring the effectiveness of online problem-solving tutorials by multi-level knowledge transfer," in *2018 Phys. Educ. Res. Conf. Proc.* (American Association of Physics Teachers, 2019), pp. 1–4.

[42] D. T. Seaton, Y. Bergner, G. Kortemeyer, S. Rayyan, I. Chuang, and D. E. Pritchard, "The impact of course structure on etext use in large-lecture introductory-physics courses," in *2013 Phys. Educ. Res. Conf. Proc.* (American Association of Physics Teachers, 2014), pp. 333–336.

[43] P. H. Winne, "Self-regulated learning," *Self-Regulated Learning*, Second Edi (Elsevier, 2015).

[44] P. R. Pintrich, "A conceptual framework for assessing motivation and self-regulated learning in college students," Educ. Psychol. Rev. **16**, 385 (2004).

[45] M. Boekaerts, "Self-regulated learning: where we are today," Int. J. Educ. Res. **31**, 445 (1999).

[46] P. R. Pintrich and E. V. De Groot, "Motivational and self-regulated learning components of classroom academic performance," J. Educ. Psycology **82**, 33 (1990).

[47] D. L. Schnipke and D. J. Scrams, "Modeling item response times with a two-state mixture model- a new approach to measuring speededness," J. Educ. Meas. **34**, 213 (1999).

[48] C. L. Barry, S. J. Horst, S. J. Finney, A. R. Brown, and J. P. Kopp, "Do examinees have similar test-taking effort? a high-stakes question for low-stakes testing," Int. J. Test. **10**, 342 (2010).

[49] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: machine learning in python," J. Mach. Learn. Res. **12**, 2825 (2011).

[50] G. Schwarz, "Estimating the dimension of a model," Ann. Stat. **6**, 461 (1978).

[51] S. Cunningham, "Causal inference: the mixtape," *Causal Inference: The Mixtape* (Yale University Press, New Haven, 2021).

[52] M. W. Guthrie, T. Zhang, and Z. Chen, "A tale of two guessing strategies: interpreting the time students spend solving problems through online log data," in *2020 Phys. Educ. Res. Conf. Proc.* (American Association of Physics Teachers, 2020), pp. 185–190.

[53] T. Zhang, M. Taub, and Z. Chen, "Measuring the impact of covid-19 induced campus closure on student self-regulated learning in physics online learning modules," in *LAK21 11th Int. Learn. Anal. Knowl. Conf.* (ACM, New York, NY, USA, 2021), pp. 110–120.

[54] J. G. MacKinnon and H. White, "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties," J. Econom. **29**, 305 (1985).

[55] J. D. Bransford and D. L. Schwartz, "Rethinking transfer: a simple proposal with multiple implications," Rev. Res. Educ. **24**, 61 (1999).

[56] T. Zhang, M. Taub, and Z. Chen, "A multi-level trace clustering analysis scheme for measuring students' self-regulated learning behavior in a mastery-based online learning environment," in *LAK22 12th Int. Learn. Anal. Knowl. Conf.* (ACM, New York, NY, USA, 2022), pp. 197–207.