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

nudging emails promoted students to use more free late days (allow for late submission without grade penalty) on homework assignments, but the change was short lived and that students quickly revert to previous behavior [13].

In a 2017 study conducted in three HarvardX 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% [14].

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.

The design of our intervention is based on Zimmerman's three-phase cycle framework of self-regulated learning [15]. The framework posits that a learner first forms intentions and goals about the assignment and creates a plan to realize them in the *forethought phase*; then the learner puts their plan into action to the extent of their ability in the *performance phase*; finally, the learner forms opinions regarding learning and of their own self-efficacy in the *self-reflection phase*. Our planning prompt intervention serves as scaffolding for the forethought phase, by first asking students to form goal intentions about the assignment and extra credit incentive, then prompt them to form a concrete plan to achieve their goals.

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 [16].

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 by different instructors in Spring 2022, and one was taught in the Fall 2021 semester. Since randomly assigning a for-credit assignment to part of the students in a class is both technologically difficult and could cause much disruption to normal 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 *et al.* defined their procrastination index based on the 60th percentile of submission times [5]. Nieberding and Heckler used cluster analysis to divide students into groups with similar procrastination behavior [9]. 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 [8]. 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 [17]. 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 [18]. Our implementation of the analysis method, adopted from Cunningham's *Causal Inference* [17], 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 [17].

II. EXPERIMENTAL SETUP

A. 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.

B. 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 5th 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 assignments 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.

C. 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 contains one piece of informative text and four questions. The informative text outlines 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 [14], and asks 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 contributes to 4% of the homework sequence credit—or about 0.1% of the total course credit. The survey is only graded for completion, so submitting a blank survey would still have earned the full points. Students are 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.

III. DATA ANALYSIS

A. 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 [16].

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 [16].

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 [19,20]. 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 [21]. 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 [22]. 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 [19]. Both disengaged study states and brief last attempts have been used in previous studies as indicators of lack of engagement with the learning materials [22].

B. Multilinear modeling

For each metric outlined above, we create the following linear model adopting the technique outlined in Cunningham [17]:

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

Here $\bar{x}_{\mathrm{est}}$ 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}, 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) [23]. 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.

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 Eq. (1), separated from the treatment effect (β_T) term.

IV. RESULTS

In Figs. 1 and 2, 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

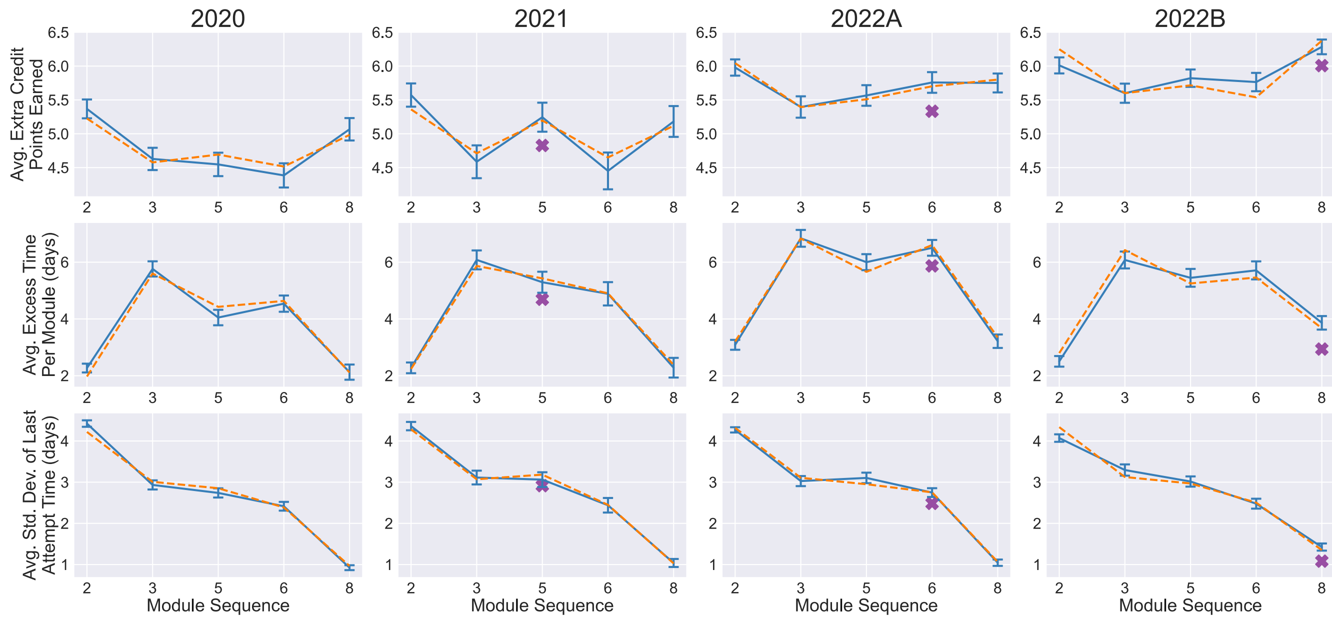

FIG. 1. Observed and modeled work distribution metrics. Blue lines indicate observed averages, with error bars for 1 standard error of the mean. Orange dotted lines show the modeled values from Eq. (1). Purple crosses show the model's prediction of each metric if the planning-prompt intervention had 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.

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 1 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 2 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.

A. Model fit

Table I 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) [23]. 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.

B. Comparison to models without treatment term

Using the ANOVA test mentioned above, we find that the model presented in Eq. (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).

C. 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

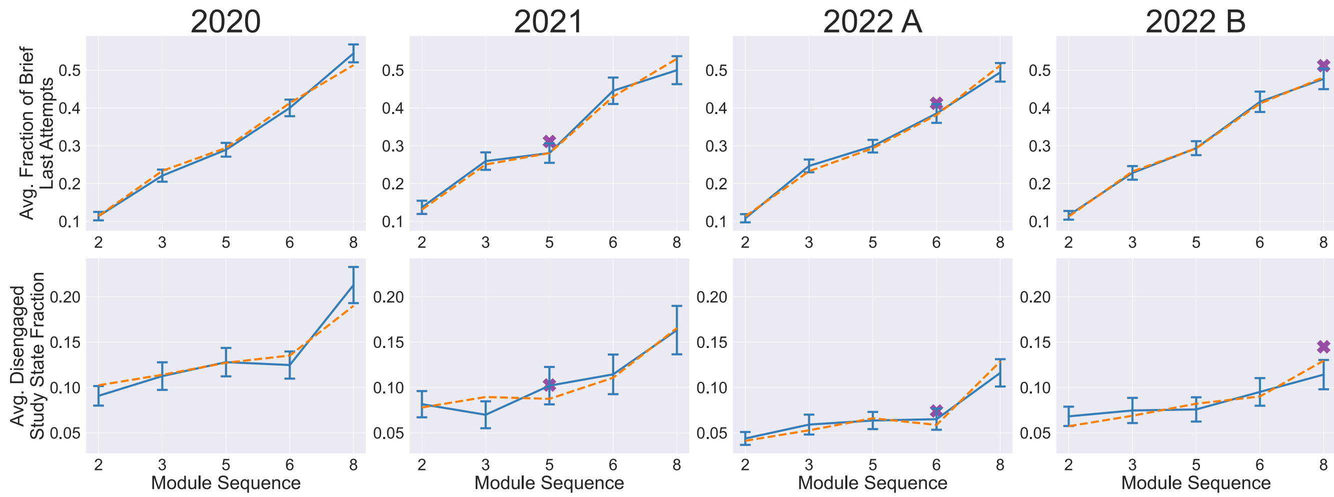

FIG. 2. Observed and modeled engagement metrics. Blue lines indicate observed averages, with error bars for 1 standard error of the mean. Orange dotted lines show the modeled values from Eq. (1). Purple crosses show the model's prediction of each metric if the planning-prompt intervention had not been 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 I. 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
β_T (Treatment)	**0.3690****	**0.7355*****	**0.2679****	**−0.0307**	**−0.0153**
β_{21} (Course Section 21)	0.1345	0.2672	0.0622	0.0169	−0.0245
β_{22A} (Course Section 22A)	1.0232	0.8247	0.1163	−0.0014	−0.0451
β_{22B} (Course Section 22B)	0.8172	1.2333	0.0979	−0.0009	−0.0612
β_2 (Sequence 2 of 9)	5.2258	1.9729	4.2202	0.1143	0.1025
β_3 (Sequence 3 of 9)	4.5767	5.6006	3.0088	0.2336	0.1140
β_5 (Sequence 5 of 9)	4.6924	4.4263	2.8526	0.2944	0.1273
β_6 (Sequence 6 of 9)	4.5155	4.6353	2.3873	0.4132	0.1353
β_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.

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.

V. 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 MOOC students are on average older (average age 30 or more), and half of them have a bachelor's degree or higher. Therefore, those students could have higher levels of self-regulatory skills that would enable them to follow through on their plans. A valuable follow up study could investigate the relation between students' self-regulatory skills and their likelihood of following through on their plans.

Similar to the observations of Fouh, Lee, and Baker [13], 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. However, data showed that they went back to their old working habits on the next module sequence. This shows that a single planning prompt nudge could 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. 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 measures of engagement, as measured by brief last attempt fraction and disengaged fraction. 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 the two semesters.

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. It would also be much harder to observe the treatment effect in the model if it only had an impact on one or two of the sequences.

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 very 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.

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 [24]. 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 [22], 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 revert 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.

ACKNOWLEDGMENTS

The authors are grateful 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. We are also grateful to Dr. Melanie Guldi for introducing us to natural experiments and the differences in differences method. This study is supported by National Science Foundation's Division of Undergraduate Education 1845436.

[1] Z. Felker and Z. Chen, The impact of extra credit incentives on students' work habits when completing online homework assignments, in *Proceedings of PER Conf. 2020, virtual conference*, pp. 143–148, 10.1119/perc.2020.pr.-Felker.

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

[3] 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).

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

[5] L. Agnihotri, R. S. Baker, and S. Stalzer, A Procrastination Index for Online Learning Based on Assignment Start Time, in *Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Virtual Conference*, edited by A. N. Rafferty, J. Whitehill, V. Cavalli-Sforza, and C. Romero (2020), pp. 550–554.

[6] 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. Differ. **174**, 110673 (2021).

[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] S. Sabnis, R. Yu, and R. F. Kizilcec, Large-scale student data reveal sociodemographic gaps in procrastination behavior, in *Proeedings of the 9th ACM Conf. Learn. @ Scale 133* (Association for Computing Machinery, New York, 2022), pp. 133–141, 10.1145/3491140.3528285.

[9] 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).

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

[11] 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).

[12] 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).

[13] E. Fouh, W. Lee, and R. S. Baker, Nudging Students to Reduce Procrastination in Office Hours and Forums, in *2021 25th International Conference Information Visualisation (IV)* (2021), pp. 248–254, 10.1109/IV53921.2021.00047.

[14] M. Yeomans and J. Reich, Planning Prompts Increase and Forecast Course Completion in Massive Open Online Courses, in *ACM International Conference Proceeding Series* (Association for Computing Machinery, New York, 2017), pp. 464–473, 10.1145/3027385.3027416.

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

[16] Z. Felker and Z. Chen, Using a planning prompt survey to encourage early completion of homework assignments, in *Proceedings of the Ninth ACM Conference on Learning @ Scale* (Association for Computing Machinery, New York, 2022), pp. 446–450, 10.1145/3491140.3528297.

[17] S. Cunningham, *Causal Inference: The Mixtape* (Yale University Press, New Haven, Connecticut, 2021).

[18] E. Nevelius, The Prize in Economic Sciences (2021), https://www.nobelprize.org/prizes/economic-sciences/2021/press-release/.

[19] 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).

[20] 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 *Proceedings of PER Conf. 2020, virtual conference*, pp. 185–190, 10.1119/perc.2020.pr.Guthrie.

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

[22] 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 International Learning Analytics and Knowledge Conference (LAK21). Association for Computing Machinery, New York, NY, USA* (2021), pp. 110–120, 10.1145/3448139.3448150.

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

[24] 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, *LAK22: 12th International Learning Analytics and Knowledge Conference* (Association for Computing Machinery, New York, 2022), pp. 197–207, 10.1145/3506860.3506887.