Not Too Early, Not Too Late: Encouraging Engagement in Education

Tracy Xiao Liu[*] Ulrike Malmendier[†] Stephanie W. Wang[‡] Shuhuai Zhang[§]

May 10, 2024

Abstract

A common reason why individuals fail to reach educational and other longterm goals is the difficulty of maintaining the required effort over time. While educators often attempt to prevent disengagement with motivating material at the beginning of a class, or with remedial sessions at the end of the class, we argue that the critical time for sustaining longterm effort falls instead in the middle. We use a simple model to capture increasing course difficulty and accumulating time-management frictions over the course of a semester. The model illustrates that the decline in student effort can be addressed by engaging students in a costly task that lower the cost of future effort at an intermediate time during the semester, just before students start "falling off." To study the effect of timing empirically, we conduct a field experiment that assigns a task aimed at engaging students with the class material at different times throughout the semester. We show that assigning tasks to low-performing students in the middle of the term, compared to early or late in the semester, improves their performance along several dimensions: attendance, homework grades, and exam grades. We also observe spillover effect to other courses. The resulting change in grades implies significant improvement in expected wages post-graduation.

[*]liuxiao@sem.tsinghua.edu.cn

[†]ulrike@berkeley.edu

[‡]swwang@pitt.edu

[§]zhangshh.20@pbcsf.tsinghua.edu.cn

1 Introduction

A key challenge in the pursuit of educational or other longterm goals is that it can be difficult to maintain the required effort throughout the duration of a program. In university settings, for example, student attendance and homework submissions tend to fall off over the course of a semester.[1] Similarly, exercise and dietary regimens suffer from high attrition rates (Kuijpers et al., 2013; Vandelanotte et al., 2016), with fitness programs losing up to 50% within the first twelve months (Dishman et al., 1985). In the domain of addiction recovery, a known challenge is the decline in the effectiveness of standard drug-abuse treatments (Higgins et al., 2004; Petry et al., 2000). Another example is consumer engagement in energy conservation, where initial efforts to reduce energy usage tends to wane over time (Allcott, 2011; Allcott and Rogers, 2014).

In all of these domains, researchers and policy makers have explored various approaches to counteract the declining engagement. In the context of education, in particular, the strong evidence on the returns to education[2] has motivated researchers to implement numerous variations in the size and type of both monetary and non-monetary incentives to improve engagement and performance in the short- and in the long-run (Lavecchia et al., 2016; Levitt et al., 2016; Gneezy et al., 2011; Kizilcec et al., 2020; Gneezy and Rustichini, 2000; Clark et al., 2020; Fryer, 2011, 2016). Many of these studies, however, show null or small effects, which has been surprising given the extended search for effective intervention, as discussed in Burgess et al. (2021), Angrist and Lavy (2009), and Clark et al. (2020), among others.

One aspect that has gotten less attention in the design of these interven-

[1] Such patterns have been documented for undergraduate classes in the US, Denmark, Singapore, and many other countries, both offline (Kassarnig et al., 2017; Yeo et al., 2023) and online (Seaton et al., 2014; Qiu et al., 2016; Ginda et al., 2019; Gong et al., 2021).

[2] Card and Krueger (1992) estimate a 7.44% rate of return to education for young cohorts (age 30-39 in 1979) across all U.S. states. The estimated rate tends to be even higher in emerging economies, for example, 12.7% return to secondary education in China (Chen et al., 2020), and more than 25% return to upper secondary and tertiary education in African countries (Barouni and Broecke, 2014).

tions is the time series pattern of optimal incentives. In this paper, we argue that the timing of the intervention over the course of a semester is of first-order importance in addressing the decline in educational engagement. Existing interventions typically fail to account for the time when students start to "fall off" over the course of a semester and reward instead the final achievement or the result of specific tests. Similarly, common educational measures focus on providing incentives or information at the beginning of a class, or on remedial sessions at the end of the class. We argue that the critical time in sustaining longterm effort falls instead in the middle.

To see why the timing may be important, consider the various factors contributing to the decay phenomenon in the educational context. Typical reasons include the increasing difficulty of the course material, accumulating demands on students' time over the course of a semester, their misperception of time demands, and procrastination (Reschly and Christenson, 2012; Fredricks et al., 2004; Skinner and Pitzer, 2012). For any combination of these factors, the timing of incentives play an important role: Early in the semester, they will not make much of a difference since students are still in the low-effort or low-friction part of the learning experience. Late in the semester, some students might have given up on following the course material and incentives might not be able to remedy the disengagement. Vice versa, other students who are still engaged at this late stage are likely highly motivated and able to handle the course material, and the additional incentives will not have a significant additional effect.

We first illustrate the importance of timing in a simple theoretical model of educational investment over time. The model assumes that the cost of engagement (e. g., attending lectures and doing homework) increases over the semester. The cost can become prohibitively high if a student starts to miss lectures or does not study the class material, as it may be too difficult to catch up and later class content tends to build on prior material. The model also allows for this cost to be higher for some students ("low ability") than others ("high ability").

In this setting, the level of engagement among "low ability" students declines over the semester because costs increase more than the expected benefit from a good performance at the end of the term without any intervention.

We then consider an intervention that asks students to complete a task that engages them with the course material and requires attendance for a randomly chosen lecture. The student incurs a utility loss if the task is not completed. The model illustrates the effects of intervention timing: If the randomly chosen lecture is early in the semester, the intervention will have no effect because engagement has not yet started to decline. If the randomly chosen lecture is in the middle of the semester, the incentive to avoid the utility loss from not completing the task is strong enough to affect students whose engagement would have dropped otherwise and "them over the hump." Finally, if the randomly chosen lecture is later, the cost of engaging in the assigned lecture will outweigh the utility loss from not completing the task. Thus, lecture assignment in the middle of the semester maximizes student engagement under this framework.

We implemented a corresponding field experiment mirroring the theoretical analysis in three undergraduate STEM classes at a large Chinese university in the 2020 Fall semester. In each of these classes, the instructor announced at the beginning of week 3 that they would be assigned the additional task of taking notes for one of the remaining lectures and of sending their notes to the TAs by the end of the lecture day. For each lecture, one of the note-takers would be chosen at random to post their notes in the class' WeChat group.

For the 571 students who participated in the experiment, we obtain basic demographics as well as academic information (major, academic performance). In addition, we obtain the full transcripts (as of the end of the experiment semester) for all students in the two larger classes (90% of students).

We find that students who were assigned the task during the middle of the term, compared to earlier or later weeks, had better engagement as measured by attendance and homework submissions. "Low-ability" (defined as those with below median pre-intervention GPA) students who were assigned to the middle of the semester also achieved higher exam performance than those assigned to earlier or later weeks. In practice, it may or may not be possible to implement a task for which one can assign all participants to the optimal intervention date.

If not, an important question would be how to identify those who would most benefit from the intervention and those who do not need it, such as the high-ability students in our setting.

The data also suggests spillover effects as students who are assigned to the middle weeks also performed better in their other courses during the same term compared to those who were assigned to other weeks. Finally, we calculate longer-term effects of the intervention using the Chinese College Students survey data on labor market outcomes. We calculate that, compared to an assignment in the earlier weeks, an assignment in the middle weeks increases students' monthly wage after graduation by about 15.6%.

Our study highlights an important but often neglected dimension of designing an intervention aimed at increasing investment in education: the timing of the intervention relative to the predicted decay in engagement. Often interventions are rolled out concurrently with the announcement or start of the class. Others take place at "crunchtime" to prepare students for the final exam. Our experiment illustrates that, when interventions are announced at the beginning but assignment is delayed to a later time corresponding to a typical decay horizon, people may change their behavior in the interim in anticipation of the costly task. Our results suggest that a somewhat delayed intervention may be optimal to induce the effort needed over time to reduce the cost of completing the intervention task. Beyond the realm of education, the same mechanism might be applicable to other domains where repeated effort is required for future benefits (Bryan et al., 2010).

Related Literature. Improving student engagement and performance is a central challenge in education, not only in the US (Rumberger, 2011) but across the world (see, e. g., Lamb et al. (2015); Goss and Sonnemann (2017) for Australia, Kingdon (2007) for India, Canadian Council on Learning (2007) for Canada, Ministry of Education (2020) for South Korea, and Li and Lerner (2011) for China). In addition, researchers have argued that the problem has worsened over time. For example, Babcock and Marks (2011) documents that US college students have decreased the time they invested in studying from 40 hours per week in 1961 to 27 hours in 2003.

In response to this challenge, a large literature proposes intervention designs to increase learning and learning outcomes, both in the US and internationally. Studies have analyzed, for example, variations in incentives that combine learning software and financial incentives (Fryer et al., 2012), the effect of providing information about the returns to education (Fryer, 2016), tutoring (Jensen, 2010; Fryer and Howard-Noveck, 2020), or access to peer advising and merit scholarships (Angrist et al., 2009). Internationally, research has assessed the effect of the Mexican Progresa program on school attendance (Schultz, 2004), variations in the design of conditional cash transfers on school attendance in Colombia (Barrera-Osorio et al., 2011) or awards for the completion of high-school exit exams in Israel (Angrist and Lavy, 2009). Several of these studies, and many of the studies discussed above, report limited effects of financial and other incentives on student learning and educational outcomes. We contribute to the existing studies by shining light on a dimension less studied so far – the alignment of incentives with the gradual decay in educational engagement over the course of a semester or school year.

Our emphasis on timing is different from prior analyses of "immediate" versus "delayed" rewards, e. g., in Gneezy and List (2006); Bellemare and Shearer (2009); Ockenfels et al. (2015); Gilchrist et al. (2016). Building upon this strand of literature, which focuses on static decisions, we evaluate the effect of aligning the timing of an educational intervention with the timing of the decay on individual learning dynamics, both theoretically and experimentally. A more related laboratory experiment by Boosey and Goerg (2020) finds that subjects' output increases when the bonus is paid in the middle instead of upfront or at the end. Here, the proposed mechanism is that workers would increase their first-period effort to signal trustworthiness if the bonus is paid in the middle.

The rest of the paper is organized as follows. Section 2 introduces the theoretical framework. Section 3 presents our intervention. Section 4 discusses the results, and Section 5 concludes.

2 Model

We introduce a simple theoretical framework to illustrate the role of the cost of educational investment increasing over the course of a learning activity, and the importance of fine-tuning the timing of educational interventions.

Consider a semester that lasts 1 unit of time. At the beginning of the semester ($t = 0$), a student makes a plan and decides for each moment $t \in [0, 1]$ of the semester whether to exert effort ($e_t = 1$) or not ($e_t = 0$).

The cost of exerting effort at time t is $\frac{1}{a}c(t, \Delta)$, where Δ is the sum of effort not exerted by time t, i.e., the fraction of time that $e_t = 0$ up until time t, $\Delta = \int_{s=0}^{t} \mathbb{1}_{\{e_s=0\}} ds$. $a > 0$ captures a student's ability. We assume that $\frac{\partial c(t,\Delta)}{\partial t} > 0$, i.e., that the course is getting more difficult as time passes by. In addition, we assume that students who stop exerting effort at any time t during the course of the semester face higher effort costs subsequently. Under standard exponential discounting with discount rate $\delta \in (0, 1)$,

Assumption 1 (Student Cost Function)**.** $\forall 0 < \Delta < t \leq 1$, $\int_0^t e^{-\delta x} c(x, 0) dx - (\int_0^{\Delta} 0 dx + \int_{\Delta}^{t} e^{-\delta x} c(x, \Delta) dx) < 0$.

This assumption reflects that course content builds on past material and is a plausible assumption for STEM courses analyzed in our experiment.[3] In addition, the assumption implies that it is cost-minimizing for students to adopt a threshold strategy: keep exerting effort until t^s, and stop exerting effort for all remaining time afterwards. Therefore, $\forall t \leq t^s$, $\Delta = 0$ and $c(t, \Delta) = c(t)$. Additionally, we assume $c(t) \in [0, 1]$ and $c(0) = 0$.

We model students' academic performance as a binary outcome, Bad or Good, realized at $t = 1$, with the associated utilities normalized to 0 for outcome B (normalized to 0) and G (normalized to 1). The probability of achieving the good outcome is determined by the total effort exerted in the whole semester, $\int_0^1 e_t dt$, which is equal to t^s under the cost-minimizing strategy.

Note-Taking Activity: Now we introduce the experimental intervention: a

[3] In our data, missing one lecture decreases a student's probability of attending the subsequent lecture by 37%.

note-taking activity. The instructor randomly assigns a student to take notes at $t^N \in [0,1]$. If a student does not exert effort in that moment, thus failing to complete this assignment, she suffers an instantaneous utility loss S.

Altogether, given t^N, at $t = 0$, the total discounted utility of a student with ability a, $V(a, t^s, t^N)$, is defined below,

$$V(a, t^s, t^N) = e^{-\delta} t^s - \mathbb{1}_{\{t^s < t^N\}} e^{-\delta t^N} S - \frac{1}{a} \int_0^{t^s} e^{-\delta x} c(x) dx \tag{1}$$

Further define $U(a, t^s) = e^{-\delta} t^s - \frac{1}{a} \int_0^{t^s} e^{-\delta x} c(x) dx$, which is a student's utility of stopping at t^s, regardless of the impact of treatment.

Student's optimal stopping choice: Given t^N, a student with ability a chooses t^{s*} to maximize her total discounted utility,

$$t^{s*}(a, t^N) = \arg\max_{t^s} V(a, t^s, t^N)$$

Instructor's optimal choice of t^N: The instructor believes that the PDF of ability is $F(a)$. She would like to choose $t^N \in [0,1]$ to maximize the expected total effort exerted by a student,

$$t^{N*} = \arg\max_{t^N} \int_0^{\infty} t^{s*}(a, t^N) F(a) da \tag{2}$$

To solve instructor's maximization problem, we first analyze how students respond to a given t^N. For simplicity, we impose a technical assumption on the shape of the cost function to refine the number of solutions in the student's maximization problem. Assumption 2 simply does not allow for cost functions that are sometimes concave and sometimes convex across time. Common cost functions that are entirely linear, concave, or convex for all time periods all satisfy Assumption 2.

Assumption 2. $\forall t_x, t_y \in [0,1]$, $\frac{\partial^2 c(t)}{\partial t^2}|_{t=t_x} \times \frac{\partial^2 c(t)}{\partial t^2}|_{t=t_y} \geq 0$.

Lemma 1. *1. If $a \geq \bar{a}$, $t^{s*}(a, t^N) = 1$;*

2. If $a < \bar{a}$,

$$t^{s*}(a,t^N) \in \begin{cases} \{t_0(a), t^N, 1\}, & \textit{if } t^N > t_0(a) \\ \{t_0(a), 1\}, & \textit{if } t^N \le t_0(a) \end{cases}$$

where $t_0(a) = \min\{t \in (0,1) | \frac{\partial V(a,t^s,t^N)}{\partial t^s} = 0\}$.

Proof. First, $\frac{\partial V(a,t^s,t^N)}{\partial t^s} = e^{-\delta} - \frac{1}{a}e^{-\delta t^s}c(t^s)$. Let $\bar{a} = \frac{\sup e^{-\delta t^s}c(t^s)}{e^{-\delta}}$

1. When $a \ge \bar{a}$,

   $\frac{\partial V(a,t^s,t^N)}{\partial t^s} = e^{-\delta} - \frac{1}{a}e^{-\delta t^s}c(t^s) \ge 0$ and $-\mathbb{1}_{\{t^s<t^N\}}e^{-\delta t^N}S$ is non-decreasing in t^s, thus $V(a,t^s,t^N)$ is non-decreasing in t^s. Therefore, it is always optimal to stop at $t^s = 1$, i.e., $t^{s*}(a,t^N) = 1$.

2. When $a < \bar{a}$,

   $\exists t'$, $\frac{\partial V(a,t^s,t^N)}{\partial t^s}|_{t^s=t'} < 0$. Moreover, $\frac{\partial V(a,t^s,t^N)}{\partial t^s}|_{t^s=0} > 0$. The intermediate value theorem (IVT) guarantees that $\exists t \in [0,t']$, $\frac{\partial V(a,t^s,t^N)}{\partial t^s}|_{t^s=t} = 0$. Furthermore, Assumption 2 guarantees that $\frac{\partial V(a,t^s,t^N)}{\partial t^s}|_{t^s=t} = 0$ has at most 2 roots, and we denote the smallest root as $t_0(a) = \min\{t \in [0,1] | \frac{\partial V(a,t^s,t^N)}{\partial t^s} = 0\}$, which is the unique local maximum.

   First, we can rule out stopping at $t = 0$ because $V(a,t_0(a),t^N) \ge V(a,0,t^N)$.

   Second, we compare stopping at $t_0(a)$, t^N and the ending of the semester.

   (a) If $t^N > t_0(a)$, the sign of $V(a,t^N,t^N) - V(a,t_0(a),t^N)$ is indeterminate.[4] Similarly, the sign of $V(a,1,t^N) - V(a,t_0(a),t^N)$ and $V(a,1,t^N) - V(a,t^N,t^N)$ are indeterminate as well, thus $t^{s*} \in \{t_0(a), t^N, 1\}$.

   (b) If $t^N \le t_0(a)$, because $\forall t^s \in [0,t_0(a)]$, $\frac{\partial V(a,t^s,t^N)}{\partial t^s} > 0$, $V(a,t_0(a),t^N) > V(a,t^N,t^N)$, we can rule out $t^s = t^N$. However, the sign of $V(a,1,t^N) - V(a,t_0(a),t^N)$ remains indeterminate. Therefore, $t^{s*} \in \{t_0(a), 1\}$.

□

[4]For example, if $S = 0$, as $t_0(a)$ is the local maximum, $\exists \epsilon > 0$, and $X = \{t | 0 < t - t_0(a) < \epsilon\}$. Therefore, for $t^N \in X$, we have $V(a,t^N,t^N) - V(a,t_0(a),t^N) < 0$. On the other hand, $\forall$ $S > e^{\delta t^N}|e^{-\delta} - \frac{1}{a}|$, $V(a,t^N,t^N) - V(a,t_0(a),t^N) > 0$.

Lemma 1 has the following implications. First, students with high ability ($a \geq \bar{a}$) will always exert effort in all lectures regardless of t^N. Second, for students with low ability ($a < \bar{a}$), the optimal stopping point t^{s*} lies in a limited set. Additionally, whenever the timing of the task is too early ($t^N < t_0(a)$), t^N will not serve as a potential optimal stopping point. Last but not least, notice that since $t^{s*}(a, t^N) = \arg\max_{t^s} V(a, t^s, t^N)$, where $V(a, t^s, t^N)$ is a non-decreasing function in a, $t^{s*}(a, t^N)$ is non-decreasing in a as well. In other words, students with higher ability will never stop exerting effort at an earlier time than those with lower ability.

2.1 Linear Cost Function

To derive an explicit solution for the students' optimal stopping timing, we consider the linear cost function $c(t) = t$. With this linear form cost function, we show that low ability students may exert more effort with t^N assigned in the middle than they would have otherwise, and the effort is maximized when $t^N = \overline{t^N}(a)$.

Lemma 2. *With $c(t) = t$, $\bar{a} = 1$.*

1. *If $a \geq \bar{a}$, $t^{s*}(a, t^N) = 1$;*

2. *If $a < \bar{a}$,*

$$t^{s*}(a, t^N) = \begin{cases} t_0(a), & \textit{if } t^N > \overline{t^N}(a) \\ t^N, & \textit{if } t_0(a) < t^N \leq \overline{t^N}(a) \\ t_0(a), & \textit{if } t^N \leq t_0(a) \end{cases}$$

$t_0(a) = -\frac{1}{\delta} W_0(-a\delta e^{-\delta})$, where W_0 is the Lambert W function;

$\overline{t^N}(a) = \min\{t \in [t_0(a), 1) | V(a, t, t^N) - V(a, t_0(a), t^N) \leq 0\}$.

Proof. 1. When $a \geq \bar{a}$,

By Lemma 1, we have $t^{s*}(a, t^N) = 1$.

2. When $a < \bar{a}$, $\forall t^s \in [0,1]$, $\frac{\partial^2 V(a,t^s,t^N)}{\partial (t^s)^2} = -\frac{1}{a}(e^{-\delta t^s} - \delta e^{-\delta t^s} t^s) < 0$. Combining with Lemma 1, this implies that there exists $t_0(a) = -\frac{1}{\delta} W_0(-a\delta e^{-\delta}) \in [0,1]$, $\frac{\partial V(a,t^s,t^N)}{\partial t^s}|_{t^s = t_0(a)} = 0$.

   - If $t^N > t_0(a)$, Lemma 1 implies that $t^{s*} \in \{t_0(a), t^N, 1\}$. Similarly, since $\forall t^s \in [t_0(a), 1]$, $\frac{\partial V(a,t^s,t^N)}{\partial t^s} < 0$, $V(a,t^N,t^N) > V(a,1,t^N)$, $t^{s*}(a,t^N) \neq 1$, we only need to compare $V(a,t^N,t^N)$ and $V(a,t_0(a),t^N)$. Since $V(a,x,t^N) - V(a,t_0(a),t^N)$ is continuous and monotonically decreasing in $x > t_0(a)$,
     - If $t^N > \overline{t^N}(a)$, $V(a,t^N,t^N) - V(a,t_0(a),t^N) < 0$, $t^{s*}(a,t^N) = t_0(a)$.
     - If $t^N \in (t_0(a), \overline{t^N}(a)]$, $V(a,t^N,t^N) - V(a,t_0(a),t^N) \geq 0$, $t^{s*}(a,t^N) = t^N$.
   - If $t^N \leq t_0(a)$, Lemma 1 implies that $t^{s*} \in \{t_0(a), 1\}$. Since $\forall t^s \in [t_0(a), 1]$, $\frac{\partial V(a,t^s,t^N)}{\partial t^s} < 0$, $V(a,t_0(a),t^N) > V(a,1,t^N)$, $t^{s*}(a,t^N) = t_0(a)$.

□

With the linear cost function, we derive two key insights in Lemma 2. First, as we have discussed in Lemma 1, students of sufficiently high ability (those with $a \geq \bar{a}$) invariably exert effort across all lectures regardless of the task timing t^N. Second, for students with lower ability (those with $a < \bar{a}$), the optimal stopping point, denoted t^{s*}, is contingent upon t^N. Specifically, if the task is scheduled either prematurely ($t^N < t_0(a)$) or excessively late ($t^N > \overline{t^N}$), t^N fails to constitute an effective optimal stopping point. Conversely, if the task is timed intermediately ($t_0(a) \leq t^N \leq \overline{t^N}$), the students' optimal strategy involves halting their effort at t^N, thereby indicating that the timing of the task plays a critical role in augmenting the students' effort. Furthermore, it is at the juncture $t^N = \overline{t^N}$ that the exerted effort reaches its pinnacle.

Figure 1 presents a simple numerical simulation of Lemma 2. The figure on the left shows how students with different abilities a respond to t^N. For a high

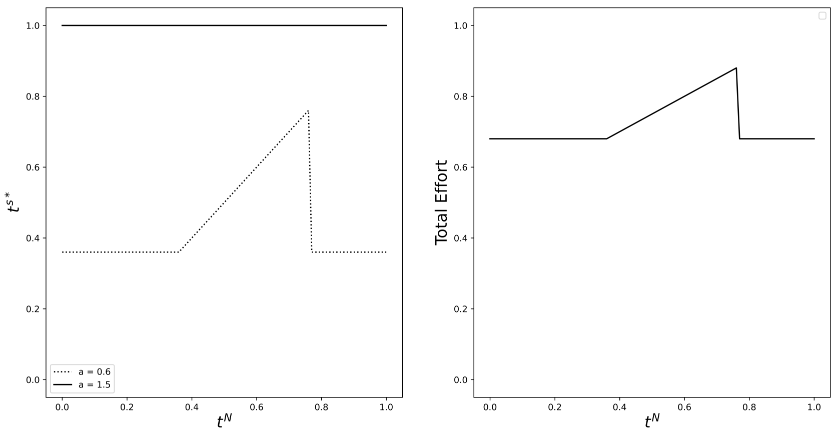


Figure 1: Simulation of Lemma 2. We set $\delta = 0.8$ and $S = 0.1$ and do the numerical simulation. For the figure on the left, we consider how the optimal stopping point of students with respectively two types of ability, $a = 1$ (solid line) and $a = 0.6$ (dotted line), depends on task timing t^N; For the figure on the right, we assume an even split of the two types of students, and examine how the total effort depends on task timing t^N.

ability student with $a = 1$ (solid line), she always chooses $t^{s*} = 1$ regardless of t^N. For a low ability student with $a = 0.6$ (dotted line), the choice of t^{s*} is contingent on t^N. When t^N is appropriately chosen in the middle, e.g., $t^N = 0.6$, they will possibly exert more effort; if t^N is too early or too late, then it is not effective in increasing students' effort. Assuming the probability of being each ability type is $1/2$, we also calculate the expected total effort from a student. Compared to implementing the task in the very early or late of the semester, setting t^N in the middle of the semester, i.e., $t^N = 0.76$, leads to highest total effort.

Additionally, we show in Appendix A that when the discount rate is higher, i.e., $\delta \geq 1$, habit formation may arise, i.e., it is optimal [** **in some cases** **] for low ability students to continue exerting effort even after t^N, suggesting a long-term effect of a one-time intervention (Ferraro and Price, 2013; Allcott and Rogers, 2014; Kesternich et al., 2016; Ito et al., 2018; Brandon et al., 2019; Jessoe et al., 2021).

3 Experimental Design

We implemented our experiment at a large Chinese university in the Fall semester of 2020, when universities in China resumed regular in-person teaching.

Courses. The three courses we conduct our experiment are all compulsory courses for undergraduate students. Table 1 summarizes the main features for each course. It is worthwhile to note that Course C has two parallel sections offered by the same instructor. Students could chose the one that best fit their schedule.[5] In total, 571 students participated in our experiment, including 57 from A, 170 from B, and 344 from C. 477 (84%) students remained until the end of the semester. The dropout rate does not significantly differ between students who are assigned to different lectures ($p = 0.273$, Fisher's Exact Test).

The Notes-Taking Task. The in-class activity we chose is notes-taking. In both the education and psychology literature, studies have shown that assigning students to take notes is an effective learning tool via both the notes-taking production process and the notes review period (Bohay et al., 2011; Piolat et al., 2005; Kobayashi, 2005). In particular, a good note summary should cover the course material well, make effective connections between concepts, and apply the gained knowledge to new contexts (Peper and Mayer, 1978). Therefore, it is considered to be a cognitive demanding task. Besides, taking notes is simple enough, so it would not discourage low-ability students from participating.[6]

Table 1: Course Description

Course	% of Female Students	# of Students	# of Students Per Lecture
A	2%	57	2, 3
B	30%	170	6, 7
C	53%	344	6, 7, 8

[5] We do not find significant differences in grades for these two sections, and pool the data together in the results section.

[6] For example, Leuven et al. (2010) finds that low-ability students achieve less when they are assigned to more difficult tasks.

3.1 Experimental Procedure

In our experiment, students are randomly assigned to take notes in a lecture starting from the third week to the end of the semester.[7] Table A1 to Table A4 in Appendix presents the summary statistics of key variables for randomization and the balance test results. Table 1 reports the number of students assigned to the task per lecture. In total, 88% students in our experiment completed the notes-taking task. The likelihood of submitting the notes does not significantly differ across assignment timing.

Additionally, students completed pre- and post-experiment surveys where they answered questions on their learning motivation, attitude, and personal characteristics.[8] The response rate is 70.2% for the pre-experiment survey and 60.8% for the post-experiment survey.

The details of our experiment implementation could be found in Section B.

4 Experimental Results

In this section, we first present the aggregate treatment effect on engagement and performance, followed by heterogeneous treatment effects analysis for high- and low-ability students. Then we examine the spillover effect of our intervention on other classes and report robustness checks. Since we use transcript data to construct the ability measure for the heterogeneous treatment effects analysis, which is only available to the Course C, we restrict all following analysis to that course. Our main results are robust when we include all three courses in the analysis (see Section 4.4).

4.1 Aggregate Treatment Effects

We first study how the task assignment timing affects attendance. Table 2 presents the regression where the dependent variable is whether an individual

[7]We exclude the first two weeks which are the shopping periods. Each week contains two lectures, and specifically, we announce the experiment in lecture 5, and students are assigned from lecture 6 to the end of the semester.

[8]The detailed survey could be found on xxx.osf.io.

attends a lecture or not. The main independent variable "Task Timing" refers to the lecture when the student is assigned to take notes. The square term captures the possible nonlinear relationship between task timing and attendance; "Lecture Index" refers to which lecture it is in the semester.

In all columns, the estimated coefficients on "Task Timing" are positive and significant ($p < 0.05$), while the coeffients on the square term are negative and significant ($p < 0.05$). That is to say, students who are assigned to take notes for one of the middle lectures are more likely to attend than students assigned to take notes in the early or late lectures. In columns 3 and 4, we add "Lecture Index" and verify the natural decline in attendance.

Table 2: Treatment Effects on Attendance

	(1)	(2)	(3)	(4)
Task Timing/100	2.336**	2.077**	2.336**	2.077**
	(1.044)	(0.988)	(1.045)	(0.988)
Task Timing^2/100	-0.061**	-0.055**	-0.061**	-0.055**
	(0.028)	(0.027)	(0.028)	(0.027)
Lecture Index/100			-1.085***	-1.085***
			(0.096)	(0.096)
Controls	No	Yes	No	Yes
Obs	7,366	7,366	7,366	7,366

Notes: This table reports linear probability model estimates of regressing a dummy variable that indicates whether a student attends a lecture or not, on the task timing and its quadratic term. Control variables include gender, international student, and whether postpone taking the course. Standard errors clustered at student level are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Result 1. *Compared to students assigned to other lectures, those assigned to the middle of the semester have the highest attendance rate.*

We use homework submission as another measure of effort and engagement. There are 12 homework assignments in a semester, approximately once a week, which allows us to replicate the analysis in Table 2. As shown in all columns of Table 3, assigning students to the middle of the semester increases the homework submission rate. Additionally, there is a natural decline for homework submission as well.

Table 3: Treatment Effects on Homework Submission

	(1)	(2)	(3)	(4)
Task Timing/100	1.749***	1.676***	1.749***	1.676***
	(0.620)	(0.582)	(0.621)	(0.582)
Task Timing^2/100	-0.046***	-0.044***	-0.046***	-0.044***
	(0.017)	(0.016)	(0.017)	(0.016)
Homework Index/100			-0.564***	-0.564***
			(0.114)	(0.114)
Controls	No	Yes	No	Yes
Obs	3,048	3,048	3,048	3,048

Notes: This table reports linear probability model estimates of regressing a dummy variable that indicates whether a student submit a homework or not, on the task timing and its quadratic term. Control variables include gender, international student, and whether postpone taking the course. Standard errors clustered at student level are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Result 2. *Assigning students to the middle of the semester helps them obtain higher homework grades.*

As our model predicts, students who are assigned to the middle of the semester exerted more effort, and we examine whether this leads to better academic performance as well. Specifically, we look at whether the timing for the notes-taking task has any effect on students' exam grades, including both midterm and final exam grades (Table 4). Though the sign of the coefficient is similar to that for attendance and homework grades, the inverse U-shape relationship is not significant for the midterm (Columns 1 and 2) and final exam grade (Columns 3 and 4). One possibility is that exam grades could both reflect students' effort and ability. Students who already have high ability may not have significantly increase their exam performance from exerting more effort.

Table 4: Treatment Effects on Exam Grades

	Midterm Exam		Final Exam	
	(1)	(2)	(3)	(4)
Task Timing	0.768	0.599	0.577	0.366
	(0.976)	(0.758)	(1.144)	(1.005)
Task Timing^2	-0.020	-0.014	-0.017	-0.010
	(0.026)	(0.020)	(0.030)	(0.027)
Controls	No	Yes	No	Yes
Obs	254	254	254	254

Notes: This table reports OLS estimates of regressing students' midterm exam grade and final exam grade, on the task timing and its quadratic term. Control variables include gender, international student, and whether postpone taking the course. HC3 robust standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

4.2 Heterogeneous Treatment Effects

As shown in our theoretical analysis, we predict that students will all exert more effort when assigned to the middle of the semester, but only students with low ability will have better exam performance under middle assignment. Prior studies such as Leuven et al. (2010) find that large rewards increase the exam pass rate of high-ability students while decreasing the pass rate of low-ability students. They propose that the financial incentive may exert displacement effect that decreases the intrinsic motivation of low ability students. Moreover, Campos-Mercade and Wengström (2020) finds that the removal of incentives leads to worse performances from low-ability students. The intervention proposed in our papers holds promise towards low-ability students as it allows to target the moment when they start to disengage from the class material.[9]

To measure students' ability, we obtain the pre-experiment transcript data for all students in Course C, and calculate the credit-weighted average grades

[9] Other studies suggest that demographic variables matter for the effectiveness of treatment intervention, including income (Gneezy et al., 2019), age (Levitt et al., 2016), and gender (Angrist and Lavy, 2009; Angrist et al., 2006, 2009; Gong et al., 2021). Moreover, the type of class (math versus reading and social sciences) matter, cf. Bettinger (2012). Additionally, Bellés-Obrero (2020) find that rewarding monetary incentives to top-performance students has a positive impact on the exam grade of students with high intrinsic motivation, but a negative impact on the grade of those with low intrinsic motivation.

point (GPA) of each student. Using the medium value of pre-experiment GPA as a cutoff, we separate students into two groups, where those who have above (below)-median pre-experiment GPA are labelled as high (low) ability students in the following analysis. Built upon the analysis in Table 4, we further interact "Task Timing" and its square term with the low ability dummy and report results in Table 5. The coefficients of "Low Ability $\times$ Task Timing" are positive and significant in all specifications ($p < 0.05$) while the coefficients of "Low Ability $\times$ Task Timing^2" are negative and significant ($p < 0.05$). For low ability students, assigning them to the middle of the semester helped them achieve better exam grades, while there is no significant effect for high ability students. Additionally, based on the estimated coefficients, the optimal timing that leads to highest grades would be around lecture 20, the 63th percentile of the semester's temporal span.[10]

Result 3. *Compared to high ability students, the inverse U-shape effect on low ability students' exam grades is more pronounced.*

Moreover, In Tables A5 and A6, we also show that Result 1 and Result 2 are mostly driven by low ability students, which is consistent with Result 3. Based on the estimated coefficients, the optimal timing that leads to highest attendance rate and homework grades would be around lecture 19, the 59th percentile of the semester's temporal span. [11]

4.3 Spillover Effects and Returns to Education

Next we study the possible spillover effects of our intervention on performance in other courses, especially for low ability students whose academic performance in the intervention courses was significantly affected by the task timing. Specifically, we examine the treatment effect on the students' GPA in other courses in

[10]We calculate this using the estimated coefficients in the quadratic form specification. Midterm Grade: -3.445 / (-0.086 $\times$ 2) = 20.029; Final Exam Grade: -4.557 / (-0.109 $\times$ 2) = 20.904.

[11]Similarly, we calculate this using the estimated coefficients in the quadratic form specification. Attendance: 0.0267 / (-0.0007 $\times$ 2) = 19.071; Homework Grade: 0.0330 / (-0.0009 $\times$ 2) = 18.333.

Table 5: Treatment Heterogeneity on Exam Grades

	Midterm Exam Grade		Final Exam Grade	
	(1)	(2)	(3)	(4)
Task Timing	-0.457	-0.529	-1.137	-1.135
	(0.633)	(0.635)	(0.745)	(0.752)
Task Timing^2	0.011	0.013	0.024	0.024
	(0.017)	(0.018)	(0.020)	(0.020)
Low Ability	-54.236***	-45.515***	-69.136***	-66.188***
	(11.098)	(10.413)	(13.297)	(13.899)
Low Ability × Task Timing	3.445**	3.182**	4.557***	4.401***
	(1.354)	(1.233)	(1.620)	(1.655)
Low Ability × Task Timing^2	-0.086**	-0.078**	-0.109**	-0.106**
	(0.037)	(0.033)	(0.044)	(0.045)
Controls	No	Yes	No	Yes
Obs	254	254	254	254

Notes: This table reports OLS estimates of regressing students' midterm exam grade and final exam grade, on the task timing and its quadratic term, as well as their interaction with low ability dummy. We use students' pre-experiment GPA to measure ability, where students whose pre-experiment GPA is lower than the median are labeled as low ability students. Control variables include gender, international student, and whether postpone taking the course. HC3 robust standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Fall 2020 and those in Spring 2021 separately. The spillover effect for the other courses in the same semester could go either way. On one hand, the intervention may have crowded out the students' effort in other courses, exhibiting a substitution effect. On the other hand, it may boost students' overall engagement level, suggesting a complementary effect.

In Columns 1 - 2 of Table 6, the dependent variable is a student's GPA in all courses (excluding the course covered in our experiment) in Fall 2020; In Columns 3 - 4, the dependent variable is a student's GPA in Spring 2021. There is a significant spillover effect - in column 2, the coefficient of "Task Timing" is 1.836 ($p < 0.05$) and the coefficient of "Task Timing^2" is -0.052 ($p < 0.01$), indicating that students who are assigned to the middle of the semester also achieve better academic performance in other courses of Fall 2020. The spillover effect declines but still carries somewhat to the subsequent semester. We summarize these findings in Result 4.

Table 6: Spillover Effect

	2020 Fall GPA		2021 Spring GPA	
	(1)	(2)	(3)	(4)
Task Timing	1.507**	1.836**	0.693	1.569*
	(0.731)	(0.722)	(0.961)	(0.857)
Task Timing^2	-0.042**	-0.052***	-0.023	-0.047*
	(0.020)	(0.020)	(0.027)	(0.024)
Total Credit	-0.340*	0.216	0.035	0.438*
	(0.196)	(0.220)	(0.251)	(0.240)
Controls	No	Yes	No	Yes
Obs	126	126	125	125

Notes: This table reports OLS estimates of regressing students' 2020 Fall semester GPA and 2021 Spring semester GPA, on the task timing and its quadratic term. The calculation of "2020 Fall GPA" excluded courses covered in the experiment. Only low ability students are included. Control variables include gender, international student, and whether postpone taking the course. HC3 robust standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Result 4. *For low ability students, assigning them to the middle of the semester improves their performance in other courses that they took during the current and* *subsequent semesters.*

Moreover, we estimate how our intervention could affect students' monthly wage after graduation. Combining survey and administrative data from an elite Chinese university, Zou et al. (2022) estimate that 1 unit increase in GPA increases economics and business major students' starting monthly wage by 2.96%. Based on this estimation, we do a simple back-of-the-envelope calculation to demonstrate the economic return of our intervention. Combining the effect of our intervention on the targeting course as well as the spillover effect in other courses (Column 2 of Table 6), the monthly wage for those who assigned to the middle is higher than those assigned to early or late. For example, on average, students that are assigned to lecture 18 are 7.06 points higher in GPA than students who are assigned to lecture 6. For a student in a 4-year undergraduate program, given the optimal implementation of task in one course each semester for the first 3 years, this effect roughly increases their overall GPA by 5.30, consequently leading to a 15.67% increase in the monthly wage.

4.4 Robustness Checks

Full Sample with Three Courses

First, our main treatment effect on attendance and homework grades is robust when expanding the analysis to the full sample with three courses. As shown in Tables 7 and 8, Results 1 and 2continue to hold. Similar to Table 4, we do not find significant results on academic performance for the full sample (Table 9).

Table 7: Treatment Effects on Attendance, Robustness Table

	(1)	(2)	(3)	(4)
Task Timing/100	1.875**	1.525**	1.879**	1.523**
	(0.733)	(0.682)	(0.735)	(0.683)
Task Timing^2/100	-0.050**	-0.041**	-0.050**	-0.041**
	(0.020)	(0.018)	(0.020)	(0.018)
Lecture Index/100			-0.826***	-0.837***
			(0.060)	(0.061)
Controls	No	Yes	No	Yes
Obs	13,996	13,996	13,996	13,996

Notes: This table reports OLS estimates of regressing a dummy variable that indicates whether a student attends a lecture or not, on the task timing and its quadratic term. Students in all the three experiment courses are included. Control variables include female, international student, STEM major, postponing taking the course, and course level dummies. Standard errors clustered at student level are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 8: Treatment Effects on Homework Submission, Robustness Table

	(1)	(2)	(3)	(4)
Task Timing/100	1.175**	1.103**	1.178**	1.103**
	(0.512)	(0.469)	(0.512)	(0.469)
Task Timing^2/100	-0.029**	-0.027**	-0.030**	-0.027**
	(0.013)	(0.012)	(0.013)	(0.012)
Homework Index/100			-0.517***	-0.522***
			(0.084)	(0.082)
Controls	No	Yes	No	Yes
Obs	5,943	5,943	5,943	5,943

Notes: This table reports OLS estimates of regressing a dummy variable that indicates whether a student submits a homework or not, on the task timing and its quadratic term. Students in all the three experiment courses are included. Control variables include female, international student, STEM major, postponing taking the course, and course level dummies. Standard errors clustered at student level are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 9: Treatment Effects on Exam Grades, Robustness Table

	Midterm Exam		Final Exam	
	(1)	(2)	(3)	(4)
Task Timing	0.173	-0.082	0.553	0.283
	(0.715)	(0.571)	(0.785)	(0.645)
Task Timing^2	-0.004	0.004	-0.015	-0.007
	(0.019)	(0.015)	(0.021)	(0.017)
Controls	No	Yes	No	Yes
Obs	477	477	477	477

Notes: This table reports OLS estimates of students' midterm exam grades and final exam grades, on the task timing and its quadratic term. Students in all the three experiment courses are included. Control variables include female, international student, STEM major, postponing taking the course, and course level dummies. HC3 robust standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Alternative Ability Measure

As analysis in Section 4.2 and 4.3 are restricted to students who take Course C, we use an alternative ability measure which covers all students in our experiment. This is whether a student postpones taking the course, e.g., a senior student takes the second-year stats course.[12]

To verify that this is indeed a valid measure of ability, we access students' final grades for Course C from 2017 to 2019 and find that postponing students' GPAs are significantly lower than those students who take the course in the suggested timeframe ($p < 0.01$, two-sided t-test).

Moreover, we conduct the same analysis as those in Tables 5 and 6 by replacing the low ability dummy with a dummy for students who postponed the course. Table 10 presents the results for the grades. The sign of all the coefficients remain the same. Table 11 presents the spillover effect results, and the signs on the coefficients remain consistent.

[12]Students may postpone taking the course for different reasons, and one of the most common reason is that they dropped the course in the previous semester. In 2020, we access students' registration data for Course C, where 104 out of 344 students postponed taking students. Among these students, 87 of them are retaking the course.

Table 10: Treatment Heterogeneity on Grades, Robustness Table

	Midterm Exam Grade		Final Exam Grade	
	(1)	(2)	(3)	(4)
Task Timing	-0.708	-0.714	-0.654	-0.601
	(0.578)	(0.531)	(0.658)	(0.605)
Task Timing^2	0.016	0.017	0.013	0.012
	(0.016)	(0.014)	(0.019)	(0.017)
Unmatched	-41.646***	-41.437***	-54.342***	-49.597***
	(14.881)	(14.471)	(15.645)	(15.809)
Unmatched × Task Timing	1.881	2.415	3.165*	3.452*
	(1.832)	(1.707)	(1.883)	(1.816)
Unmatched × Task Timing^2	-0.036	-0.046	-0.064	-0.070
	(0.051)	(0.046)	(0.051)	(0.048)
Controls	No	Yes	No	Yes
Obs	477	477	477	477

Notes: This table reports OLS estimates of regressing students' midterm exam grade and final exam grade, on the task timing and its quadratic term, as well as their interaction with unmatched dummy. Students who take the course later than scheduled are labeled as unmatched. Control variables include female, international student, STEM major, postponing taking the course, and course level dummies. HC3 robust standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 11: Spillover Effect, Robustness Table

	2020 Fall GPA		2021 Spring GPA	
	(1)	(2)	(3)	(4)
Task Timing	1.270*	1.070*	-0.057	-0.377
	(0.663)	(0.601)	(0.818)	(0.812)
Task Timing^2	-0.033*	-0.026*	0.003	0.014
	(0.018)	(0.015)	(0.022)	(0.021)
Total Credit	0.290	0.175	0.172	0.182
	(0.198)	(0.187)	(0.144)	(0.137)
Controls	No	Yes	No	Yes
Obs	93	93	91	91

Notes: This table reports OLS estimates of regressing students' 2020 Fall semester GPA and 2021 Spring semester GPA, on the task timing and its quadratic term. The calculation of "2020 Fall GPA" excluded courses covered in the experiment. Only students who postpone taking the courses are included. Control variables include pre-experiment GPA, female, international student, STEM major, postponing taking the course, and course level dummies. HC3 robust standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

4.4.1 Counterfactual Analysis

In Result 3, we showed that low ability students who are assigned to the middle of the semester are more likely to have better academic performance. Since everyone got assigned the notes-taking task, one natural question is: can the notes-taking task help students achieve better academic performance compared to those who are not assigned the task at all? We address this question by comparing the final grades of Course C in the 2020 Fall semester to the grades in previous semesters when no notes-taking assignment was implemented. We would expect the gap between grades of high-ability and low-ability students to narrow. Figure 2 presents the dynamic pattern of the final grade gap between high-ability and low-ability students. Using the final grade gap in 2017 - 2019 semesters, we estimate an linear OLS model to predict a counterfactual final grade gap for the 2020 semester if there is no assignment. We find that the realized grade gap is smaller than the counterfactual value. Additionally, when we divide the class into five groups according to the task timing, and plot the final grade gap, we find that the final grade gaps for Groups 2-5 are all smaller than the counterfactual value.

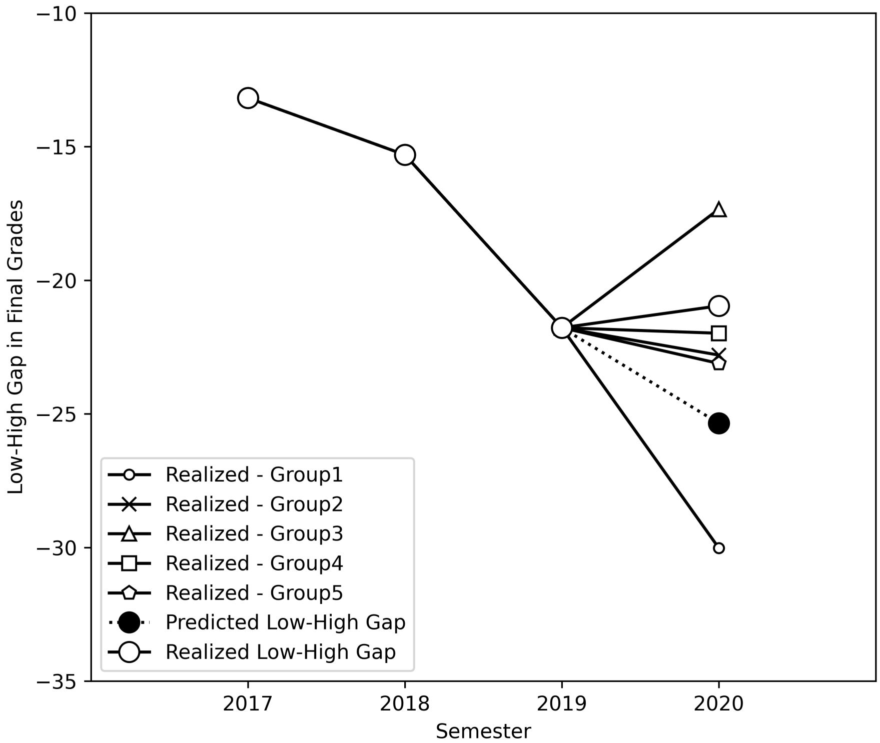


Figure 2: Counterfactual analyses. For each semester, we calculate the GPA of historical semesters for each student, and label them into high/low ability groups according to the median, and calculate the gap between these two groups (the 4 hollow circles). For semester 2020, we evenly split low ability students into 5 groups according to the task timing, where group 1 is the earliest and group 5 is the latest. For each group, we display the difference between its group mean and the average grade of all high ability students (5 different types of markers). Utilizing transcript data in semester 2017, 2018 and 2019, we estimate an OLS model and predict the counterfactual gap in 2020 (the black circle).

5 Conclusion

Students engagement decay is a notoriously challenging problem in education. Intuitively, a student would keep exerting effort until the marginal cost exceeds the marginal benefit, and the increasing cost of exerting effort over time could lead to the engagement decay. One mechanism which is often used in practice is for instructors to assign an additional task accompanying with benefit right after the falloff to increase the overall effort exerted by students. We theoretically analyze that the timing dimension of such intervention, which is rarely examined in the literature, is essential to the effectiveness of intervention, especially for low-ability students.

To test our theoretical predictions, we conducted a field experiment in undergraduate classes. We randomly assign students to different lectures to finish notes-taking tasks and test whether the timing of interventions matters in improving the academic performance of students. The experimental results show that assigning students to the middle of the term help those low ability students achieve higher academic performance. We also document a spillover effect, indicating that those students who are assigned to the middle of the semester also achieve higher grades in other courses. Furthermore, we estimate the long-term welfare effect of the treatment. For example, compared to students who are assigned to lecture 6, those who are assigned to lecture 18 will have 0.7 higher GPA, consequently leading to a 10.101 RMB increase in the monthly wage. We do note one potential general equilibrium effect outside the scope of our current study. If not implemented universally, students who dislike the public nature of posting their notes or other such tasks might decide to avoid these classes altogether.

Despite the high returns to education, students are likely to get disengaged over time, consequently leading to insufficient investment. The toolkit of behavioral economics has been widely applied to improve educational engagement and other outcomes, though the results are mixed (Gneezy et al., 2011; Lavecchia et al., 2016). We contribute by focusing on another dimension of designing the choice architecture: timing. Our results call for attention from both academic researchers as well as education practitioners and policy makers.

Our analyses of students learning can apply to other repetitive behaviors as well. Notice that one extension of the model (Appendix A) shows the possibility of a temporary intervention to have long-term impact, which happens when the time horizon is broad and the cost of exerting effort is not increasing quickly. This is likely to be fulfilled in other scenarios where continuous effort is needed in pursuing long-term goals, e.g., exercise (DellaVigna and Malmendier, 2006), alcohol abstinence (Schilbach, 2019), and skill development. Similar to education, many people fall short of the established objective because they lack persistent investment. We expect that a behavioral intervention with appropriate implementation timing could potentially help people pull through the period of stagnation and stay engaged even after the intervention. All these could be tested in future work.

References

Allcott, Hunt, "Social norms and energy conservation," *Journal of public Economics*, 2011, *95* (9-10), 1082–1095.

_ **and Todd Rogers**, "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," *American Economic Review*, 2014, *104* (10), 3003–3037.

Angrist, Joshua and Victor Lavy, "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 2009, *99* (4), 1384–1414.

_ **, Daniel Lang, and Philip Oreopoulos**, "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics*, 2009, *1* (1), 136–163.

_ **, Eric Bettinger, and Michael Kremer**, "Long-term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia," *American Economic Review*, 2006, *96* (3), 847–862.

Babcock, Philip and Mindy Marks, "The Falling Time Cost of College: Evidence from Half a Century of Time Use Data," *Review of Economics and Statistics*, 2011, *93* (2), 468–478.

Barouni, Mahdi and Stijn Broecke, "The returns to education in Africa: some new estimates," *The Journal of Development Studies*, 2014, *50* (12), 1593–1613.

Barrera-Osorio, Felipe, Marianne Bertrand, Leigh L Linden, and Francisco Perez-Calle, "Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia," *American Economic Journal: Applied Economics*, 2011, *3* (2), 167–195.

Bellemare, Charles and Bruce Shearer, "Gift Giving and Worker Productivity: Evidence from a Firm-level Experiment," *Games and Economic Behavior*, 2009, *67* (1), 233–244.

Bellés-Obrero, Cristina, "Who is Learning? A Field Experiment Comparing Three Different Incentive Schemes in the Same Educational Setting," 2020. Working Paper.

Bettinger, Eric P, "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores," *Review of Economics and Statistics*, 2012, *94* (3), 686–698.

Bohay, Mark, Daniel P Blakely, Andrea K Tamplin, and Gabriel A Radvansky, "Note Taking, Review, Memory, and Comprehension," *American Journal of Psychology*, 2011, *124* (1), 63–73.

Boosey, Luke and Sebastian Goerg, "The Timing of Discretionary Bonuses–Effort, Signals, and Reciprocity," *Games and Economic Behavior*, 2020, *124*, 254–280.

Brandon, Alec, John A List, Robert D Metcalfe, Michael K Price, and Florian Rundhammer, "Testing for Crowd Out in Social Nudges: Evidence from a Natural Field Experiment in the Market for Electricity," *Proceedings of the National Academy of Sciences*, 2019, *116* (12), 5293–5298.

Bryan, Gharad, Dean Karlan, and Scott Nelson, "Commitment devices," *Annual Review of Economics*, 2010, *2* (1), 671–698.

Burgess, Simon, Robert Metcalfe, and Sally Sadoff, "Understanding the response to financial and non-financial incentives in education: Field experimental evidence using high-stakes assessments," *Economics of Education Review*, 2021, *85*, 102195.

Campos-Mercade, Pol and Erik Wengström, "Threshold Incentives and Academic Performance," 2020. Working Paper.

Canadian Council on Learning, "State of Learning in Canada: A Year in Review, 2009-2010," 2007.

Card, David and Alan B Krueger, "Does school quality matter? Returns to education and the characteristics of public schools in the United States," *Journal of political Economy*, 1992, *100* (1), 1–40.

Chen, Yi, Sheng Jiang, and Li-An Zhou, "Estimating returns to education in urban China: Evidence from a natural experiment in schooling reform," *Journal of Comparative Economics*, 2020, *48* (1), 218–233.

Clark, Damon, David Gill, Victoria Prowse, and Mark Rush, "Using Goals to Motivate College Students: Theory and Evidence from Field Experiments," *Review of Economics and Statistics*, 2020, *102* (4), 648–663.

DellaVigna, Stefano and Ulrike Malmendier, "Paying Not to Go to The Gym," *American Economic Review*, 2006, *96* (3), 694–719.

Dishman, Rod K, James F Sallis, and Diane R Orenstein, "The determinants of physical activity and exercise.," *Public Health Reports*, 1985, *100* (2), 158.

Ferraro, Paul J and Michael K Price, "Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-scale Field Experiment," *Review of Economics and Statistics*, 2013, *95* (1), 64–73.

Fredricks, Jennifer A, Phyllis C Blumenfeld, and Alison H Paris, "School engagement: Potential of the concept, state of the evidence," *Review of educational research*, 2004, *74* (1), 59–109.

Fryer, Roland G, "Financial Incentives and Student Achievement: Evidence From Randomized Trials," *The Quarterly Journal of Economics*, 2011, *126* (4), 1755–1798.

_ , "Information, Non-Financial Incentives, and Student Achievement: Evidence From a Text Messaging Experiment," *Journal of Public Economics*, 2016, *144*, 109–121.

— **and Meghan Howard-Noveck**, "High-dosage tutoring and reading achievement: evidence from New York City," *Journal of Labor Economics*, 2020, *38* (2), 421–452.

— **, Tanaya Devi, Richard T Holden et al.**, "Vertical versus horizontal incentives in education: Evidence from randomized trials," Technical Report, National Bureau of Economic Research 2012.

Gilchrist, Duncan S, Michael Luca, and Deepak Malhotra, "When 3+ 1 > 4: Gift Structure and Reciprocity in the Field," *Management Science*, 2016, *62* (9), 2639–2650.

Ginda, Michael, Michael C Richey, Mark Cousino, and Katy Börner, "Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design," *PloS one*, 2019, *14* (5), e0215964.

Gneezy, Uri and Aldo Rustichini, "Pay Enough Or Don't Pay At All," *The Quarterly Journal of Economics*, 2000, *115* (3), 791–810.

— **and John A List**, "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 2006, *74* (5), 1365–1384.

— **, — , Jeffrey A Livingston, Xiangdong Qin, Sally Sadoff, and Yang Xu**, "Measuring Success in Education: The Role of Effort on the Test Itself," *American Economic Review: Insights*, 2019, *1* (3), 291–308.

— **, Stephan Meier, and Pedro Rey-Biel**, "When and Why Incentives (Don't) Work to Modify Behavior," *The Journal of Economic Perspectives*, 2011, *25* (4), 191–209.

Gong, Jie, Tracy Xiao Liu, and Jie Tang, "How Monetary Incentives Improve Outcomes in MOOCs: Evidence from a Field Experiment," *Journal of Economic Behavior & Organization*, 2021, *190*, 905–921.

Goss, Peter and Julie Sonnemann, *Engaging students: Creating classrooms that improve learning*, Grattan Institute, 2017.

Guo, Jinxin, Rongxiu Zhu, Quanqin Zhao, Mingxia Li, and Shuyong Zhang, "Adoption of the Online Platforms Rain Classroom and WeChat for Teaching Organic Chemistry during COVID-19," *Journal of Chemical Education*, 2020, *97* (9), 3246–3250.

Higgins, Stephen T, Sarah H Heil, and Jennifer Plebani Lussier, "Clinical implications of reinforcement as a determinant of substance use disorders," *Annual Review of Psychology*, 2004, *55*, 431–461.

Ito, Koichiro, Takanori Ida, and Makoto Tanaka, "Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand," *American Economic Journal: Economic Policy*, 2018, *10* (1), 240–67.

Jensen, Robert, "The (Perceived) Returns to Education and the Demand for Schooling," *The Quarterly Journal of Economics*, 2010, *125* (2), 515–548.

Jessoe, Katrina, Gabriel E Lade, Frank Loge, and Edward Spang, "Residential Water Conservation During Drought: Experimental Evidence from Three Behavioral Interventions," *Journal of Environmental Economics and Management*, 2021, *110*, 102519.

Kassarnig, Valentin, Andreas Bjerre-Nielsen, Enys Mones, Sune Lehmann, and David Dreyer Lassen, "Class attendance, peer similarity, and academic performance in a large field study," *PloS one*, 2017, *12* (11), e0187078.

Kesternich, Martin, Andreas Löschel, and Daniel Römer, "The Long-term Impact of Matching and Rebate Subsidies When Public Goods Are Impure: Field Experimental Evidence from the Carbon Offsetting Market," *Journal of Public Economics*, 2016, *137*, 70–78.

Kingdon, Geeta Gandhi, "The progress of school education in India," *Oxford Review of Economic Policy*, 2007, *23* (2), 168–195.

Kizilcec, René F, Justin Reich, Michael Yeomans, Christoph Dann, Emma Brunskill, Glenn Lopez, Selen Turkay, Joseph Jay Williams,

and Dustin Tingley, "Scaling up behavioral science interventions in online education," *Proceedings of the National Academy of Sciences*, 2020, *117* (26), 14900–14905.

Kobayashi, Keiichi, "What limits the encoding effect of note-taking? A meta-analytic examination," *Contemporary Educational Psychology*, 2005, *30* (2), 242–262.

Kuijpers, Wilma, Wim G Groen, Neil K Aaronson, Wim H van Harten et al., "A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors," *Journal of Medical Internet Research*, 2013, *15* (2), e2281.

Lamb, Stephen, Jen Jackson, Anne Walstab, and Shuyan Huo, "Educational opportunity in Australia 2015: who succeeds and who misses out," *APO-58167*, 2015.

Lavecchia, Adam M, Heidi Liu, and Philip Oreopoulos, "Behavioral Economics of Education: Progress and Possibilities," in "Handbook of the Economics of Education," Vol. 5, Elsevier, 2016, pp. 1–74.

Leuven, Edwin, Hessel Oosterbeek, and Bas Van der Klaauw, "The Effect of Financial Rewards on Students' Achievement: Evidence From a Randomized Experiment," *Journal of the European Economic Association*, 2010, *8* (6), 1243–1265.

Levitt, Steven D, John A List, Susanne Neckermann, and Sally Sadoff, "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance," *American Economic Journal: Economic Policy*, 2016, *8* (4), 183–219.

Li, Yibing and Richard M Lerner, "Trajectories of school engagement during adolescence: implications for grades, depression, delinquency, and substance use.," *Developmental psychology*, 2011, *47* (1), 233.

Ministry of Education, "2020 Education in Korea," 2020.

Ockenfels, Axel, Dirk Sliwka, and Peter Werner, "Timing of Kindness–Evidence from a Field Experiment," *Journal of Economic Behavior & Organization*, 2015, *111*, 79–87.

Peper, Richard J and Richard E Mayer, "Note Taking as a Generative Activity," *Journal of Educational Psychology*, 1978, *70* (4), 514.

Petry, Nancy M, Bonnie Martin, Judith L Cooney, and Henry R Kranzler, "Give them prizes and they will come: Contingency management for treatment of alcohol dependence.," *Journal of consulting and clinical psychology*, 2000, *68* (2), 250.

Piolat, Annie, Thierry Olive, and Ronald T Kellogg, "Cognitive Effort During Note Taking," *Applied Cognitive Psychology*, 2005, *19* (3), 291–312.

Qiu, Jiezhong, Jie Tang, Tracy Xiao Liu, Jie Gong, Chenhui Zhang, Qian Zhang, and Yufei Xue, "Modeling and Predicting Learning Behavior in Moocs," in "Proceedings of the Ninth ACM International Conference on Web Search and Data Mininge" 2016.

Reschly, Amy L and Sandra L Christenson, "Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct," in "Handbook of research on student engagement," Springer, 2012, pp. 3–19.

Rumberger, Russell W, *Dropping out: Why students drop out of high school and what can be done about it*, Harvard University Press, 2011.

Schilbach, Frank, "Alcohol and self-control: A field experiment in India," *American economic review*, 2019, *109* (4), 1290–1322.

Schultz, T Paul, "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program," *Journal of development Economics*, 2004, *74* (1), 199–250.

Seaton, Daniel T, Yoav Bergner, Isaac Chuang, Piotr Mitros, and David E Pritchard, "Who does what in a massive open online course?," *Communications of the ACM*, 2014, *57* (4), 58–65.

Skinner, Ellen A and Jennifer R Pitzer, "Developmental dynamics of student engagement, coping, and everyday resilience," in "Handbook of research on student engagement," Springer, 2012, pp. 21–44.

Vandelanotte, Corneel, Andre M Müller, Camille E Short, Melanie Hingle, Nicole Nathan, Susan L Williams, Michael L Lopez, Sanjoti Parekh, and Carol A Maher, "Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors," *Journal of nutrition education and behavior*, 2016, *48* (3), 219–228.

Yeo, Sing Chen, Clin KY Lai, Jacinda Tan, Samantha Lim, Yuvan Chandramoghan, Teck Kiang Tan, and Joshua J Gooley, "Early morning university classes are associated with impaired sleep and academic performance," *Nature Human Behaviour*, 2023, *7* (4), 502–514.

Zou, Tao, Yue Zhang, and Bo Zhou, "Does GPA matter for university graduates' wages? New evidence revisited," *Plos one*, 2022, *17* (4), e0266981.

Online Appendices

A Model Extension

In Section 2, we restrict the discount rate $\delta < 1$. In this section, we discuss the case where $\delta \geq 1$ under $c(t) = t$. Given $\delta \geq 1$, now $\frac{\partial^2 V(a,t^s,t^N)}{\partial t^{s2}}$ is indeterminate. Define $D(a, t_x, t_y) = U(a, t_x) - U(a, t_y)$, which compares the difference of utility between stopping at t_x and t_y. The Lemma below shows that even for a student with ability $a < \bar{a}$, **[** i.e., the benefit of exerting effort does not always exceed the cost, it is possible that in some cases **]**, she may keep exerting effort after finishing the assignment, i.e., forming the learning habit.

Lemma 3. *1. When $a \geq a^h = \{a|D(a, 1, t_0(a)) = 0\}$, $t^{s*}(a, t^N) = 1$*

2. When $a^l \leq a < a^h$, where $a^l = \{a|D(a, 1, t_0(a)) = -e^{-\delta \overline{\overline{t^N}}(a)}S\}$,

$$t^{s*}(a, t^N) = \begin{cases} t_0(a), & \text{if } 0 \leq t^N \leq t_0(a) \\ t^N, & \text{if } t_0(a) < t^N \leq \underline{t^N}(a) \\ 1, & \text{if } \underline{t^N}(a) < t^N \leq \overline{\overline{t^N}}(a) \\ t_0(a), & \text{if } \overline{\overline{t^N}}(a) < t^N \end{cases}$$

3. When $a < a^l$,

$$t^{s*}(a, t^N) = \begin{cases} t_0(a), & \text{if } t^N > \overline{t^N}(a) \\ t^N, & \text{if } t_0(a) < t^N \leq \overline{t^N}(a) \\ t_0(a), & \text{if } t^N \leq t_0(a) \end{cases}$$

where, $t_0(a) = -\frac{1}{\delta}W_0(-a\delta e^{-\delta})$, $\overline{\overline{t^N}}(a) = \min\{t \in [t_0(a), 1]|D(a, 1, t_0(a)) \leq -e^{-\delta t}S\}$, $\underline{t^N}(a) = \min\{t \in [t_0(a), 1]|D(a, t, 1) \leq 0\}$, $\overline{t^N}(a) = \min\{t \in [t_0(a), 1]|D(a, t, t_0(a)) \leq -e^{-\delta t}S\}$

Proof. Firstly, we introduce the following remark,

Remark 1. $\frac{\partial D(a,1,t_0(a))}{\partial a} > 0$.

Proof.

$$
\begin{aligned}
D(a,1,t_0(a)) &= \int_{t_0(a)}^{1} \frac{\partial U(a,t^s)}{\partial t^s} dt \\
\frac{\partial [D(a,1,t_0(a))]}{\partial a} &= -\frac{\partial U(a,t^s)}{\partial t^s}|_{t_0(a)} \cdot \frac{\partial t_0(a)}{\partial a} + \int_{t_0(a)}^{1} \frac{\frac{\partial U(a,t^s)}{\partial t^s}}{\partial a} dt \qquad (3) \\
&= \int_{t_0(a)}^{1} \frac{\frac{\partial U(a,t^s)}{\partial t^s}}{\partial a} dt
\end{aligned}
$$

Notice that $\frac{\frac{\partial U(a,t^s)}{\partial t^s}}{\partial a} = \frac{\partial [e^{-\delta} - \frac{1}{a} e^{-\delta t} t]}{\partial a} > 0$ and $t_0(a) < 1$, hence $\frac{\partial D(a,1,t_0(a))}{\partial a} > 0$. □

Next, we examine $t^{s*}(a, t^N)$ for different a.

1. If $a \geq \bar{a} > a^h$, by Lemma 1, $t^{s*} = 1$.

   Next, we move to cases with $a < \bar{a}$. Now $\frac{\partial V(a,t^s,t^N)}{\partial t^s} = e^{-\delta} - \frac{1}{a} e^{-\delta t^s} t^s = 0$ has two roots, the smaller one is $t_0(a) = -\frac{1}{\delta} W_0(-a\delta e^{-\delta})$, while the larger one is $t_{-1}(a) = -\frac{1}{\delta} W_{-1}(-a\delta e^{-\delta})$, and at least the smaller one of the two roots, $t_0(a)$, is on $[0,1]$. $\frac{\partial V(a,t^s,t^N)}{\partial t^s} > 0$ for $t^s \in [0, t_0(a)) \bigcup [t_{-1}(a), \infty)$, $\frac{\partial V(a,t^s,t^N)}{\partial t^s} < 0$ for $t^s \in [t_0(a), t_{-1}(a))$.

2. If $\bar{a} > a \geq a^h$, the definition of a^h guarantees that $t_{-1}(a) \in [0,1]$. Remark 1 implies that for $a \geq a^h$, $D(a,1,t_0(a)) \geq 0$. $V(a,1,t^N) = U(a,1) > U(a,t_0(a)) \geq V(a,t_0(a),t^N)$, hence $t^{s*} \neq t_0(a)$. We further consider the location of t^N,

   - If $t^N < t_0(a)$, Lemma 1 implies that $t^{s*} = 1$.
   - If $t^N \geq t_0(a)$,
     - If $t^N \in [t_0(a), t_{-1}(a))$, $\frac{\partial V(a,t^s,t^N)}{\partial t^s} < 0$ hence $U(a,t_0(a)) > U(a,t^N)$, $t^{s*} \neq t^N$, then $t^{s*} = 1$.
     - If $t^N \in [t_{-1}(a), 1)$, $\frac{\partial V(a,t^s,t^N)}{\partial t^s} > 0$ hence $V(a,1,t^N) = U(a,1) > U(a,t^N) = V(a,t^N,t^N)$, $t^{s*} \neq t^N$, then $t^{s*} = 1$.

Altogether, $\forall\ a \geq a^h$, $t^{s*} = 1$.

3. If $a^l \leq a < a^h$,

   - If $t^N \leq t_0(a)$, Lemma 1 implies that we only need to compare $V(a, t_0(a), t^N)$ and $V(a, 1, t^N)$. By Remark 1, with $a < a^h$, $D(a, 1, t_0(a)) < 0$, hence $V(a, t_0(a), t^N) = U(a, t_0(a)) > U(a, 1) = V(a, 1, t^N)$. Therefore,$t^{s*} = t_0(a)$.
   - If $t_0(a) < t^N$, define $\underline{t^N}(a) = \min\{t \in [t_0(a), 1) | D(a, 1, t) \leq 0\}$ and $\overline{\overline{t^N}}(a) = \max\{t \in [t_0(a), 1] | D(a, 1, t_0(a)) \geq -e^{-\delta t} S\}$. By the definition of a^l, $\underline{t^N}(a) \leq \overline{\overline{t^N}}(a)$. We discuss the following cases,
     - If $t^N > \overline{\overline{t^N}}(a) \geq \underline{t^N}(a)$, $V(a, t^N, t^N) < V(a, t_0(a), t^N)$ and $V(a, 1, t^N) > V(a, t^N, t^N)$, hence $t^{s*} = t_0(a)$.
     - If $\underline{t^N}(a) \leq t^N \leq \overline{\overline{t^N}}(a)$, $V(a, 1, t^N) \geq V(a, t^N, t^N)$ and $V(a, 1, t^N) \geq V(a, t_0(a), t^N)$, hence $t^{s*} = 1$.
     - If $t^N < \underline{t^N}(a) \leq \overline{\overline{t^N}}(a)$, $V(a, 1, t^N) < V(a, t^N, t^N)$ and $V(a, 1, t^N) > V(a, t_0(a), t^N)$, hence $t^{s*} = t^N$.

4. If $a < a^l$, by Remark 1 and the definition of a^l, $V(a, 1, t^N) < V(a, t_0(a), t^N)$. Together with Lemma 1, we only need to compare $V(a, t_0(a), t^N)$ and $V(a, t^N, t^N)$. Following the same logic in Lemma 2, we have:

$$t^{s*}(a, t^N) = \begin{cases} t_0(a), & t^N > t_0(a) \\ t^N, & t_0(a) < t^N \leq \overline{t^N}(a) \\ t_0(a), & t^N \leq t_0(a) \end{cases}$$

□

Figure A1 presents a numerical example of Lemma 3. The $a = 1.5$ (solid line) and $a = 0.6$ (dotted line) are similar to Figure 1, but notice that for a student with middle ability (dashed line, $a = 1.08$), it is possible that when t^N is appropriately chosen, e.g., $t^N = 0.85$, she will continue learning after finishing the task, and keep doing it until the end of the semester ($t^{s*}(a, t^N) = 1$). Again,

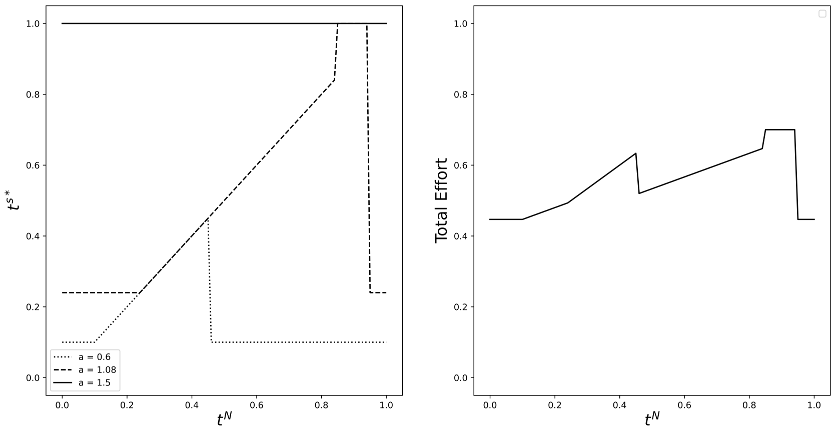


Figure A1: Simulation of Lemma 3. We set $\delta = 2$, $S = 0.1$ and do the numerical simulation. For the figure on the left, we consider how the optimal stopping point of students with respectively three types of ability, $a = 1.5$ (solid line), $a = 1.08$ (dashed line), and $a = 0.6$ (dotted line), depends on task timing t^N; For the figure on the right, we assume an even split of the three types of students, and examine how the total effort depends on task timing t^N.

assuming the probability of each ability type is 1/3, we find that the expected total effort is maximized at $t^N = 0.85$.

B Experiment Implementation Details

Figure A2 illustrates the timeline of the experiment.

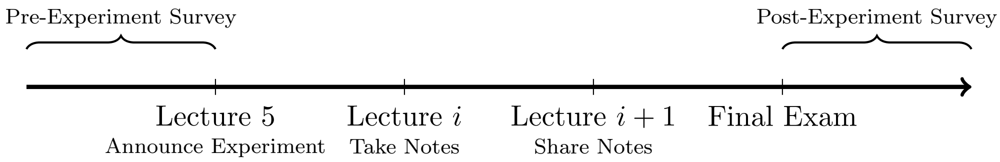
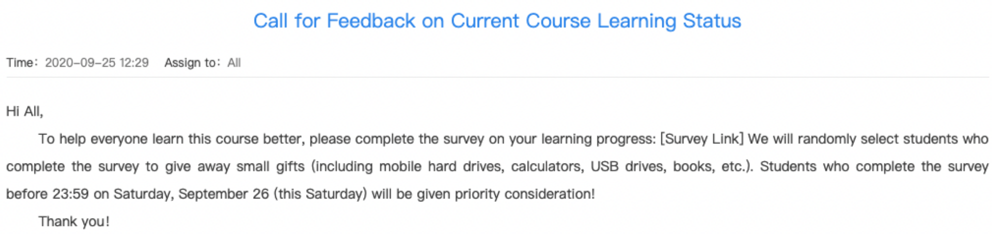


Figure A2: Experiment Timeline

Pre-experiment Survey. We distribute the pre-experiment survey before the beginning of lecture 5, which is the first lecture of week 3. Specifically, we post notice on both the e-learning platform [13] and WeChat group[14] to ask students to complete the survey. Figure A3 provides a sample notice on the e-learning platform, while the content is similar for the post in the WeChat group. The survey is due right before the day of lecture 5.



Call for Feedback on Current Course Learning Status

Time: 2020-09-25 12:29 Assign to: All

Hi All,

To help everyone learn this course better, please complete the survey on your learning progress: [Survey Link] We will randomly select students who complete the survey to give away small gifts (including mobile hard drives, calculators, USB drives, books, etc.). Students who complete the survey before 23:59 on Saturday, September 26 (this Saturday) will be given priority consideration!

Thank you!

Figure A3: Pre-experiment Survey distribution via the e-learning platform notice

Lecture 5: Announce Experiment. In lecture 5, we announce the detailed experiment schedule in the classroom. We also post the notes-taking schedule on both the e-learning platform and WeChat groups before. Figure A4 provides an example notice on the e-learning platform.

Lecture i: Notes-taking. One day before Lecture i, we send a reminder email to students assigned to that particular lecture, and also post notice in

[13]The e-learning platform is an online platform for teaching and learning purpose. On this platform, teachers and teaching sssistants could post notice, arrange assignments, and share slides, while students could access course materials and information.

[14]WeChat groups have been used extensively in China to facilitate teaching, especially after the outbreak of COVID-19 (Guo et al., 2020). Instructors can post course material, grading policy, and answer students' questions in a timely manner. Students can ask both instructors and their classmates questions, sharing course related material and chat.

Notices

The Notes–taking Session of This Course

Time：2020–09–29 16:14 Assign to：All

To help you better grasp the teaching content, there will be a note–taking session this semester. The detailed schedule is as follows:

1. Starting from September 30th, each student will be assigned to one class to be responsible for taking notes for that class. The specific arrangements can be found at the end of this announcement. The students responsible for the first class (September 30th) are A, B, and C.

2. Students responsible for taking notes must send their notes to the email address [TA's Email Address] by 23:59 on the day of the class. The email and attachment should be named in the format "XX Month XX Day Class Notes–Name–Student Number". Electronic notes or photos of paper notes are both acceptable.

3. Before the next class, the teaching assistant will invite one of the students to share their notes in the WeChat group and give a verbal summary of their notes for at least 60 seconds in the form of a voice message and send it to the group. The teaching assistant will notify the students who need to share their notes in advance. Please complete your contact information in the e–learning platform.

4. After the student shares his/her notes, the teaching assistant will send out a questionnaire in the WeChat group and invite everyone to rate the student's notes. The teaching assistant will randomly draw students who fill out the rating questionnaire to give away small gifts. In addition, at the end of this semester, the teaching assistant will also give away small gifts to the students whose notes have been rated highly.

To help everyone understand, here is an example:

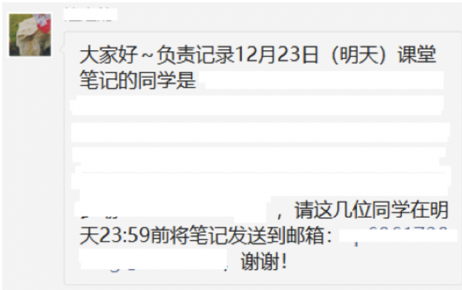
Example: Alice, Bob, Clark and David are assigned to take notes for the October 12th class. After the end of the October 12th class, these four students will send their corresponding notes to the designated email address before 23:59 on the same day. After receiving the notes, the teaching assistant randomly selects Clark to share his notes in the WeChat group before the next class (i.e., October 14th). On the morning of October 14th, Clark shares their notes in the class WeChat group and gives a voice summary of their notes. Then, the teaching assistant sends out a questionnaire in the group and invites the students in the group to rate Clark's notes. After collecting everyone's ratings, the teaching assistant draws lots from the students who filled out the rating questionnaire to give Emma a prize.

For any questions, please feel free to contact [TA's Email Address] by email.

Thank you!

Attachment: Weekly Schedule

Figure A4: Announcement on the e-learning platform the WeChat group. Figure A5 provides an example of notice in the WeChat group. We also specified the time for submitting the notes. For example, if a student is selected to take notes on March 3rd, she is expected to send the notes to our research assistant by the end of that day, though we clarified that the notes-taking task is not compulsory and does not affect the final grade.



TA: Hi all - Please notice that student A, B, C and D are assigned to take notes for the Dec 23 Lecture. Please email your notes to TA by the end of tomorrow.

Figure A5: Reminder of notes-taking in the WeChat group

Lecture $i+1$: Share Notes. Among students who are assigned to the lecture, we randomly select one of them to share the notes and to leave a voice message in the class WeChat group. Almost all students participated in this activity.[15] After students share their notes, all students are asked to rate on the quality of the notes from 0 to 100 points in the WeChat groups. To encourage them to rate, we randomly select one winner for each selected note and distribute small gifts such as USB drivers and calculators.[16] Figure A6 provides an example of notes-sharing and rating in the WeChat group. On average, each note received 12 ratings and the rating frequency declines over the weeks.

Post-experiment Survey. We distribute the post-experiment survey after the final exam via the e-learning platform and the WeChat group.

[15] Only 2 students refused to post their notes.

[16] Specifically, student i's probability of winning the gift P^i_{win} is: $P^i_{win} = \frac{(100-|S_i-\bar{S}|)^2}{\sum_{i=1}^{N}[(100-|S_i-\bar{S}|)^2]}$, where N is the number of ratings, $\bar{S}$ is the median score of ratings, and S_i is the rating score for student i.

Student A: Here is my note and comments are welcome!
TA: The notes for Dec 1 lecture are shared by Student A. Please rate them online. Students who submit ratings before 23:59pm, Dec 2 will get a chance to win a gift!

Figure A6: Notes-sharing and rating in the WeChat group

Notices

Call for Feedback on This Semester's Learning Status

Time: 2021-01-02 16:49 Assign to: All

Dear all,

To further improve teaching quality, we kindly invite you to complete the survey about your learning status this semester: [Survey Link] https://www.wjx.cn/jq/102795856.aspx

We will randomly select students who complete the survey to give away small gifts (including mobile hard drives, calculators, USB drives, books, etc.). Students who complete the survey before 23:59 on January 1st will be given priority consideration!

Thank you!

Figure A7: Post-experiment Survey distribution via the e-learning platform notice

C Randomization Check

Table A1: Course A Randomization

Lecture Index	Female		Whether Matched		STEM Major		International	
	=0	=1	=0	=1	=0	=1	=0	=1
6	3	0	1	2	0	3	3	0
7	2	0	0	2	0	2	2	0
8	2	0	2	0	0	2	2	0
9	3	0	0	3	0	3	3	0
10	2	0	0	2	0	2	2	0
11	2	0	0	2	0	2	2	0
12	2	0	0	2	0	2	2	0
13	3	0	0	3	0	3	3	0
14	2	0	0	2	0	2	2	0
16	2	0	1	1	0	2	2	0
17	3	0	0	3	0	3	3	0
18	2	0	0	2	0	2	2	0
19	2	0	0	2	0	2	2	0
20	2	0	0	2	0	2	2	0
21	3	0	1	2	0	3	3	0
22	2	0	0	2	0	2	2	0
23	2	0	1	1	0	2	2	0
24	3	0	0	3	0	3	3	0
25	2	0	0	2	0	2	2	0
26	2	0	1	1	0	2	2	0
27	1	1	0	2	0	2	2	0
28	3	0	0	3	0	3	3	0
29	2	0	0	2	0	2	2	0
30	2	0	0	2	0	2	2	0
31	2	0	1	1	0	2	2	0
χ^2 **Test p-value**	0.260		0.206		/		/	

Notes: In the experiment, we excluded the first 4 lectures that lied in shopping periods, announced the experiment schedule in lecture 5, and students are randomly assigned from remaining lectures (from lecture 6 to lecture 31). Each cell represents that for each task timing, how many students have the corresponding characteristics.

Table A2: Course B Randomization

Lecture Index	Female		Whether Matched		STEM Major		International	
	=0	=1	=0	=1	=0	=1	=0	=1
6	6	1	0	7	0	7	7	0
7	4	3	0	7	3	4	7	0
8	5	1	0	6	2	4	6	0
9	6	1	0	7	0	7	7	0
10	5	1	0	6	2	4	6	0
11	5	2	0	7	2	5	7	0
12	5	1	0	6	0	6	5	1
13	3	4	0	7	6	1	7	0
14	2	4	1	5	2	4	6	0
15	6	1	0	7	2	5	7	0
16	3	3	0	6	0	6	6	0
17	4	3	0	7	2	5	7	0
18	4	2	0	6	1	5	6	0
19	4	3	0	7	3	4	7	0
20	5	2	0	7	2	5	6	1
21	3	3	0	6	1	5	6	0
22	4	3	0	7	2	5	7	0
23	5	1	1	5	2	4	5	1
24	5	2	0	7	3	4	7	0
25	4	2	0	6	3	3	6	0
26	5	2	0	7	2	5	7	0
27	4	2	1	5	2	4	6	0
28	5	2	0	7	0	7	7	0
29	4	2	0	6	1	5	6	0
30	6	1	0	7	2	5	7	0
31	6	0	0	6	3	3	6	0
χ^2 Test p-value	0.815		0.419		0.189		0.496	

Notes: In the experiment, we excluded the first 4 lectures that lied in shopping periods, announced the experiment schedule in lecture 5, and students are randomly assigned from remaining lectures (from lecture 6 to lecture 31). Each cell represents that for each task timing, how many students have the corresponding characteristics.

Table A3: Course C1 Randomization

Lecture Index	**Female**		**Whether Matched**		**STEM Major**		**International**	
	=0	=1	=0	=1	=0	=1	=0	=1
6	5	2	5	2	7	0	6	1
7	2	5	3	4	7	0	7	0
8	4	3	3	4	7	0	7	0
9	3	3	3	3	6	0	4	2
10	3	4	3	4	7	0	6	1
11	2	5	3	4	7	0	6	1
12	3	3	4	2	6	0	6	0
13	2	5	2	5	7	0	7	0
14	4	3	3	4	7	0	5	2
15	3	3	2	4	6	0	6	0
17	3	4	1	6	7	0	6	1
18	1	6	4	3	7	0	5	2
19	5	1	2	4	6	0	6	0
20	5	2	4	3	7	0	7	0
21	4	3	5	2	7	0	5	2
22	0	6	4	2	6	0	5	1
23	4	3	1	6	7	0	6	1
24	2	5	3	4	7	0	7	0
25	2	4	3	3	6	0	4	2
26	4	3	2	5	7	0	7	0
27	5	2	3	4	7	0	6	1
28	3	3	3	3	6	0	6	0
29	3	4	4	3	7	0	5	2
30	4	3	3	4	7	0	7	0
31	2	4	2	4	6	0	5	1
χ^2 **Test p-value**	0.506		0.901		/		0.526	

Notes: In the experiment, we excluded the first 4 lectures that lied in shopping periods, announced the experiment schedule in lecture 5, and students are randomly assigned from remaining lectures (from lecture 6 to lecture 31). Each cell represents that for each task timing, how many students have the corresponding characteristics.

Table A4: Course C2 Randomization

Lecture Index	**Female**		**Whether Matched**		**STEM Major**		**International**	
	=0	=1	=0	=1	=0	=1	=0	=1
6	3	5	3	5	8	0	8	0
7	3	4	1	6	7	0	5	2
8	3	4	2	5	7	0	5	2
9	1	6	0	7	7	0	5	2
10	2	5	1	6	7	0	6	1
11	4	3	0	7	7	0	5	2
12	4	3	2	5	7	0	6	1
13	1	6	1	6	7	0	6	1
14	5	2	1	6	7	0	5	2
15	2	5	2	5	7	0	5	2
17	1	6	0	7	7	0	6	1
18	2	5	0	7	7	0	6	1
19	6	2	0	8	8	0	7	1
20	3	4	1	6	7	0	5	2
21	1	6	2	5	7	0	5	2
22	3	4	2	5	7	0	7	0
23	2	5	1	6	7	0	7	0
24	3	4	0	7	7	0	6	1
25	4	3	3	4	7	0	3	4
26	2	5	0	7	7	0	7	0
27	6	1	2	5	7	0	6	1
28	5	2	1	6	7	0	4	3
29	3	4	2	5	7	0	6	1
30	0	7	1	6	7	0	6	1
31	3	4	1	6	7	0	5	2
χ^2 Test p-value	0.109		0.609		/		0.685	

Notes: In the experiment, we excluded the first 4 lectures that lied in shopping periods, announced the experiment schedule in lecture 5, and students are randomly assigned from remaining lectures (from lecture 6 to lecture 31). Each cell represents that for each task timing, how many students have the corresponding characteristics.

D Heterogeneity in Treatment Effect on Attendance and Homework Grades

Table A5: Treatment Effects on High vs. Low Ability Students Attendance

	(1)	(2)	(3)	(4)
Task Timing/100	2.668*	2.391*	2.277	2.215
	(1.517)	(1.440)	(1.464)	(1.379)
Task Timing^2/100	-0.072*	-0.067*	-0.057	-0.056
	(0.041)	(0.038)	(0.040)	(0.038)
Lecture Index/100	-0.979***	-0.979***	-1.189***	-1.189***
	(0.142)	(0.142)	(0.131)	(0.131)
Controls	No	Yes	No	Yes
Sample	Low Ability		High Ability	
Obs	3,654	3,654	3,712	3,712

Notes: This table reports OLS estimates of regressing a dummy variable that indicates whether a student attends a lecture or not, on the task timing and its quadratic term. Control variables include female, international student, and postponing taking the course dummies. Standard errors clustered at student level are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A6: Treatment Effects on High vs. Low Ability Students Homework Grades

	(1)	(2)	(3)	(4)
Task Timing/100	3.300***	3.110***	0.394	0.376
	(1.145)	(1.051)	(0.441)	(0.413)
Task Timing^2/100	-0.088***	-0.084***	-0.009	-0.008
	(0.031)	(0.028)	(0.011)	(0.010)
Homework Index/100	-0.966***	-0.966***	-0.169	-0.169
	(0.197)	(0.197)	(0.106)	(0.106)
Controls	No	Yes	No	Yes
Sample	Low Ability		High Ability	
Obs	1,512	1,512	1,536	1,536

Notes: This table reports OLS estimates of regressing a dummy variable that indicates whether a student attends a lecture or not, on the task timing and its quadratic term. Control variables include female, international student, and postponing taking the course dummies. Standard errors clustered at student level are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.