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

Tracy Xiao Liu† Ulrike Malmendier‡ Stephanie W. Wang§ Shuhuai Zhang¶

March 21, 2025

Abstract

A major hurdle in the pursuit of educational attainment and other long-term goals is the difficulty in maintaining the required effort over time. One example is the decline in attendance and engagement among college students over the course of a semester. While prior interventions have focused on increasing engagement at the start and end of courses, we argue, both theoretically and empirically, that the critical time falls in the middle when students are in danger of disengaging due to accumulating effort costs. We conduct a field experiment in three undergraduate STEM classes at a large Chinese university that randomly assigns a note-taking task for different lectures to different students, inducing them to engage with the course content. We find that assigning tasks to low-ability students in the middle of the term, compared to early or late, significantly improves their attendance, homework grades, and exam grades. The data also suggest spillovers to performance in other classes. The improvement in grades predicts 0.5–1.8% higher post-graduation earnings opportunities.

Keyword: self-control, education, field experiment

JEL classification: C93, D91, I23

*We would like to thank participants at the 2024 Economic Science Association Asian-Pacific Regional Conference and 2022 SITE for helpful discussions and comments. We thank Chazel Hakim for excellent research assistance. Our study was approved by University of Pittsburgh Institutional Review Board and was registered in the AEA Registry under AEARCTR-0007467. Liu gratefully acknowledges financial support by the National Natural Science Foundation of China (72222005, 72342032) and Tsinghua University (No. 2022Z04W01032).

†School of Economics and Management, Tsinghua University; liuxiao@sem.tsinghua.edu.cn.

‡Department of Economics and Haas School of Business, University of California, Berkeley; ulrike@berkeley.edu.

§Department of Economics, University of Pittsburgh; swwang@pitt.edu.

¶PBC School of Finance, Tsinghua University; zhangshh.20@pbcsf.tsinghua.edu.cn.

1 Introduction

A key challenge in the pursuit of educational or other long-term goals is the difficulty in maintaining the consistent effort required. In university settings, for example, student attendance and homework submissions tend to fall 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% of participants within the first twelve months (Dishman et al., 1985). In addiction recovery, a known challenge is the decline in the effectiveness of standard drug-abuse treatments over time (Petry et al., 2000; Higgins et al., 2004). Another example is consumer engagement in energy conservation, where initial efforts to reduce energy usage tend 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, strong evidence on the returns to education[2] has motivated researchers to explore many variations in the size and type of monetary and non-monetary incentives to induce higher engagement and performance.(Gneezy and Rustichini, 2000; Fryer, 2011; Gneezy et al., 2011; Fryer, 2016; Lavecchia et al., 2016; Levitt et al., 2016; Clark et al., 2020; Kizilcec et al., 2020). Many of these studies, however, show muted effects, which is surprising given the intense search for effective interventions, as discussed in Angrist and Lavy (2009), Clark et al. (2020), and Burgess et al. (2021), among others.

One aspect that has received less attention in the design of these interventions, whether in the realm of education or in other domains, is the question

[1] Such patterns have been documented in undergraduate classes in the US, Denmark, Singapore, and many other countries, both for in-person (Kassarnig et al., 2017; Yeo et al., 2023) and online courses (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% 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).

of the optimal timing of interventions within longer-term programs. The design of existing studies typically targets students' final achievement or the result of specific tests and milestones throughout the semester. Common educational measures focus on providing incentives or information at the beginning of a course, or on remedial sessions at the end.

In this paper, we argue that the timing of educational interventions is of first-order importance. Rather than relying exclusively on material or teaching at the beginning or end of a semester, interventions designed to sustain long-term effort should account for the time when students start to "fall off." We show that appropriately timed and pre-announced interventions can induce students to stay engaged in the course material.

To see why the timing may be important, consider the factors contributing to the decline in engagement: the increasing difficulty of the course material, accumulating demands on students' time over the course of a semester, their misperceptions of time demands, and procrastination (Fredricks et al., 2004; Reschly and Christenson, 2012; Skinner and Pitzer, 2012). For any combination of these factors, the timing of incentives influences their effectiveness. Toward the beginning of the semester, incentives might not make much of a difference since students are still in the low-effort or low-friction part of the learning experience. Toward the end of the semester, incentives might come too late for those students who have given up on following the course material, and might not affect those students either who are still engaged as they are intrinsically motivated and able to handle the course material. An intermediate timing might therefore hold more promise for an effective intervention, but has not been studied in prior work.[3]

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. If a student starts to miss lectures or does not study the class material, the cost can become prohibitively high as the class content builds on prior material and

[3]Most related are studies comparing student performance with and without an intermediate test or exam (De Paola and Scoppa, 2011; Bangert-Drowns et al., 1991).

becomes too difficult. The model allows this cost to be higher for some students ("low ability") than for others ("high ability"). In this setting, the level of engagement among low-ability students declines over the semester because their costs increase more than the expected benefit from a good performance at the end of the term.

We then consider an intervention that induces students to attend a randomly chosen lecture and engage with the course material. Students incur a utility loss if the task is not completed. The model illustrates the effects of intervention timing: If the randomly chosen lecture is too 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 "get them over the hump." Finally, if the randomly chosen lecture is too late, the cost of staying engaged until 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.

To test these predictions, we implemented a field experi 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 students 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.

We find that students who were assigned the task during the middle of the term, compared to earlier or later weeks, showed better engagement in their education as measured by attendance and homework submissions. "Low-ability" students (defined as those with below-median pre-intervention GPA), who were assigned to the middle of the semester also achieved higher exam performance than those assigned to earlier or later weeks.

The heterogeneity of the treatment effect by ability of the student is impor-

tant for its implementation in practice. While the intervention did not lower high-ability students' performance and one could apply it to all students, it would be preferable to focus on students whose performance actually benefits from the intervention for two reasons. First, from a welfare perspective, educators might not want to impose additional effort costs on students who do not benefit from the intervention. Second, from a practical perspective, it may not be possible to assign all participants to the optimal intervention date. If so, an important question would be how to identify those who would benefit most 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. We found that students who were 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. Our estimates suggest that an assignment in the middle weeks could improve students' wage after graduation by 0.50% to 1.84%.

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 decline in engagement. Frequently, interventions are rolled out concurrently with the announcement or start of the class. Others take place at "crunch time" 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 the typical horizon for declining engagement in a given context, students may change their behavior in the interim in anticipation of the costly 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) and Goss and Sonnemann (2017) for Australia, Kingdon (2007) for India, theCanadian Council on Learning (2007) for Canada, the South Korean Ministry of Education (2020) for South Korea, and Li and Lerner (2011) for China). Moreover, 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 has proposed numerous interventions designed to increase learning and learning outcomes. 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 a light on a dimension less studied so far – the alignment of the timing of incentives with the gradual decline 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). We evaluate the effect of aligning the timing of an educational intervention with the timing of the engagement decline on individual learning dynamics, both theoretically and experimentally. A 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. However, the proposed mechanism that workers increase their first-period effort to signal trustworthiness if the bonus is paid in the middle is orthogonal to the underlying mechanisms we focus on.

Most closely related, De Paola and Scoppa (2011) conduct a randomized field experiment to test the impact of examination frequency and interim feedback provision on achievement. Students undertaking the midterm exam perform better than students who only take the final exam (though high-ability students benefit more). Di Pietro (2013) examines a natural experiment, in which the final exam was delayed while the midterm exam remained at the same time, and estimates a negative effect on students' academic performance. Via a meta-analysis, Bangert-Drowns et al. (1991) show that students who took at least one test during a 15-week term scored about one half of a standard deviation higher on examinations than students who took no tests. Our study helps crystallize the underlying mechanism and proposes a way to optimize intervention timing.

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

2 Model

We introduce a simple theoretical framework to illustrate the role of increasing effort costs over the duration of a learning activity and to analyze how the effect of an educational intervention varies depending on its timing.

Consider a semester that lasts one unit of time, $t \in [0, 1]$. At the beginning of the semester ($t = 0$), a student plans when to exert effort ($e_t = 1$) and when not to ($e_t = 0$) throughout the semester. Academic performance is modeled as a binary outcome, Bad or Good, realized at the end of the semester ($t = 1$), with utilities normalized to 0 for Bad and 1 for Good.

Costly effort increases the probability of obtaining a good academic outcome. We assume that the cost of effort (cost density) $c(t)$ increases over time ($\frac{\partial c(t)}{\partial t} >$

0), reflecting the growing difficulty of the course material. For simplicity, we assume $c(t) \in [0,1]$ and $c(0) = 0$. We model effort as continuous in time to account for studying, completing homework, reviewing, and other activities that happen throughout the term and not just during lectures. (Nevertheless, the lectures can naturally segment the continuous effort.)

We denote periods when the student plans to exert effort by $(t_{2k-1}, t_{2k}]$ and no-effort periods by $(t_{2k}, t_{2k+1}]$, for $k = 1, \ldots, K$, where $0 = t_1 < ...t_{2K} \leq t_{2K+1} = 1$. The parameter $K \geq 1$ represents the number of effort intervals. As we assume $c(0) = 0$, the student always starts with the semester with a period of effort. The student may end with a period of no-effort ($t_{2K} < t_{2K+1}$) or a period with effort ($t_{2K} = t_{2K+1}$).

Students incur a catch-up cost if they want to resume lectures and studying again after a no-effort period. The catch-up cost exceeds the cumulative cost of studying during the time interval they skipped, or else the lecture material cannot be absorbed. As a result, students always catch up on all prior material up to the time when they drop off for good, either at or before the end of the semester. Finally, effort costs depend on students' ability a (with $a > 0$).

We model the total discounted cost of effort over the semester $[0,1]$ as

$$C = \frac{1}{a}\left[\underbrace{\int_0^1 \mathbb{1}_{e_t=1} e^{-\delta t} c(t)dt}_{\text{Cost of Effort}} + \underbrace{\sum_{k=1}^{K-1} \lambda e^{-\delta t_{2k+1}} \int_{t_{2k}}^{t_{2k+1}} c(s)ds}_{\text{Catch-up Cost}}\right], \tag{1}$$

where $\delta \in (0,1]$ is the standard exponential discount rate. The multiplier $\frac{1}{a}$ captures that higher-ability students incur lower effort costs. The first term represents the discounted cost of exerting effort during active periods. The second term captures the cost associated with catching up on missed material from inactive periods. This cost is incurred after an inactive period $(t_{2k}, t_{2k+1}]$ at the restart times t_{2k+1}. We assume the catch-up cost to be proportional to the duration of the missed period $(t_{2k}, t_{2k+1}]$ and the cost density $c(t)$ during that period, but higher as governed by the penalty parameter $\lambda > 1$).

The probability of obtaining a Good outcome depends on the length of intervals during which the student exerts effort, plus the length of non-effort intervals that are later made up, relative to the total length of the semester. Defining t^s as the time at which the student stops exerting effort altogether, the probability equals $t^s/1 = t^s$.

As a simple (discretized) example, consider a student who takes a course with 20 lectures. Let's assume the student exerts effort from lectures 1 to 4, then misses lectures 5 to 8, resumes effort for lectures 9 to 16, and then drops out for the remaining lectures. In this scenario, the total cost of effort comprises the regular effort cost from lectures 1 to 4 and 9 to 16, plus a one-time catch-up cost for catching up on the material missed from lectures 5 to 8. Her academic outcome depends on the total length of effective effort (including both regular and made-up effort), which totals 16 lectures. Given that the course consists of 20 lectures, her probability of achieving a Good outcome is 80%.

For the remaining analysis, we assume for simplicity $c(t) = t$. (A discussion of more general forms of $c(t)$ is in Appendix I.2.) We further focus on the case of catch-up costs being high enough[4], hence the cost of resuming effort after a break is greater than the cost of maintaining continuous effort during the same period. Thus, the student chooses a threshold strategy, namely, to exert effort continuously up to a final stopping time $t^s \in [0, 1]$, which fully characterizes her decision and uniquely determines the total cost of effort. In the absence of an intervention, the total discounted utility of a student with ability a who stops at t^s is

$$U(a, t^s) = \underbrace{e^{-\delta} t^s}_{\text{Benefit of Good Outcome}} \underbrace{- \frac{1}{a} \int_0^{t^s} e^{-\delta t}\, t\, dt}_{\text{Cost of Effort}},$$

and the optimal stopping time is given by

$$t^{s^*}(a) = \arg\max_{t^s} U(a, t^s).$$

[4]The condition being the penalty parameter $\lambda > \frac{2(e^\delta - (1+\delta))}{\delta^2}$, see Appendix I.1 for the detailed derivation.

In particular, denoting

$$t_0(a) = -\frac{1}{\delta} W_0(-a\delta e^{-\delta}),$$

which is the earliest time where the marginal benefit of effort equals the marginal cost, where W_0 is the Lambert W function. We have $t^{s^*}(a) = \min\{t_0(a), 1\}$ in the absence of an intervention.

Experimental Intervention. We now introduce a note-taking activity, reflecting the experimental intervention. The instructor assigns a note-taking task to student i at time t_i^n. If the student fails to exert effort at that time, thus failing to complete the task, she suffers an instantaneous utility loss S.

Dropping subscript i for simplicity (though in the experiment, different students are assigned different note-taking times), the total discounted utility of a student with ability a at time $t = 0$ becomes

$$V(a, t^s, t^n) = U(a, t^s) - \underbrace{\mathbb{1}_{\{t^s < t^n\}} e^{-\delta t^n} S}_{\text{Utility Loss for Missing the Task}}. \tag{2}$$

Given an assignment at time t^n, a student with ability a chooses the largest t^s that maximizes her total discounted utility,

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

Next, we derive an explicit solution for the student's optimal stopping rule as a function of their ability a and the timing of the educational intervention t^n. Define an *Effective Threshold* $t_E(a)$, that is,

$$t_E(a) \coloneqq \max\Big\{t \in [t_0(a), 1] \ : \ \Delta U(a, t, t_0(a)) \geq -e^{-\delta t} S\Big\}.$$

$t_E(a)$ is the latest task time at which a student with ability a prefers to align with the intervention and stop at t^n, rather than stopping at the baseline $t_0(a)$ (and incurring the penalty S). $\Delta U(a, t_x, t_y) = U(a, t_x) - U(a, t_y)$ is the difference in utility between stopping at t_x and t_y. We call students with ability $a \geq 1$ high-ability student, and students lower ability ($a < 1$) low-ability students.

Proposition 1. *(i) High-ability students ($a \geq 1$) always exert effort through the entire semester, $t^{s^*}(a, t^n) = 1$. (ii) For low-ability students ($a < 1$), the optimal stopping point $t^{s^*}(a, t^n)$ depends on the timing of the note-taking task t^n:*

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

Proof. See Appendix I.3. □

Proposition 1 yields two key insights. First, for students with sufficiently high ability ($a \geq 1$) the timing t^n of the note-taking task does not matter. They will always exert effort throughout the entire semester. Second, for students with lower ability ($a < 1$), the timing of the note-taking task does affect whether and when they start to disengage from their class. Specifically, if the task is scheduled either too early ($t^n < t_0(a)$) or too late ($t^n > t_E(a)$), t^n does not affect the optimal stopping point. However, there is a range of times in the middle of the term ($t_0(a) \leq t^n \leq t_E(a)$) for which it becomes optimal for the students to stop their effort at t^n rather than $t_0(a)$. Thus, the timing of the task matters for sustaining the students' effort past when it would cease without the intervention. Indeed, the effort is maximized at $t^n = t_E(a)$.

Figure 1 presents a simple numerical simulation of Proposition 1. The left panel shows how students with different abilities respond to varying t^n. For a high ability student ($a = 1.5$, solid line), the optimal stopping time is always $t^{s^*}(a, t^n) = 1$, regardless of t^n. For a low ability student ($a = 0.6$, dotted line), the choice of $t^{s^*}(a, t^n)$ depends on t^n. When t^n is in the middle of the semester, such as $t^n = 0.6$, the student may exert more effort; if t^n is too early or too late, it is ineffective in increasing the student's effort. In our experiment, we vary the timing of note-taking tasks across students to observe the impact on their effort levels.

The right panel of Figure 1 aggregates the effect of choosing one fixed time t^n for all students in a class, assuming that 50% are high-ability with $a = 1.5$

and the remaining 50% are low-ability with $a = 0.6$. Compared to scheduling the task very early or late in the semester, setting t^n in the middle ($t^n = 0.76$) leads to the highest total effort.

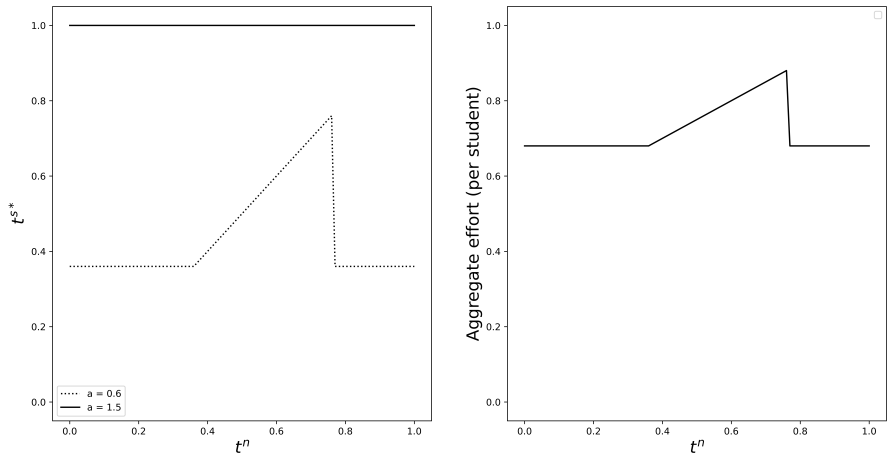


Figure 1: Simulation of Proposition 1. We set $\delta = 0.8$ and $S = 0.1$ for the simulation. The left panel shows the optimal stopping point of students as a function of task timing t^n for two types of students, one with $a = 1.5$ (solid line) and one with $a = 0.6$ (dotted line). The right panel shows the aggregate effort (per student), assuming $a = 1.5$ for 50% of students and $a = 0.6$ for the remaining 50% of students.

As an additional result, we show in Appendix I.4 that when the discount rate is sufficiently high, i.e., $\delta > 1$ (note that the discount factor is $e^{-\delta t}$, hence $\delta > 1$ indicates a steeper instantaneous discount rather than students outweigh future benefits than current benefits), low-ability students continue exerting effort even after the note-taking lecture t^n when they would have stopped exerting effort before lecture t^n without the intervention. This type of habit formation highlights the potential long-term effects 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).

Finally, the theoretical framework allows us to discuss welfare implications of educational interventions. We have derived how timing the intervention at $t_E(a)$ maximizes students' effort, i. e., induces them to absorb as much course

material and attend as many lectures as possible given their ability a. However, the educational intervention uses the threat of lowering student utility by S, which students will incur if the timing is not optimized. Moreover, even without incurring the cost S, the resulting student utility from working longer and stopping only at $t^{s^*}(a, t^n)$ is lower than the utility from stopping at $t_0(a)$ without the educational intervention. Hence, its implementation requires additional justification such as students failing to see the long-term benefits of sustained effort or displaying time inconsistency. This perspective of informational or behavioral constraints is typically assumed in the literature on educational interventions. Our calibration results in Section 4.4, suggesting potentially significant effects on post-graduation wages, is consistent with this perspective.

3 Experimental Design

We designed an experiement to examine whether the timing of participating in an in-class activity has impact on students' engagement and consequently on their learning performances.

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

Courses. We conducted the experiment in three courses which are all compulsory courses for undergraduate students. Table 1 summarizes the main features for each course.[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 ($p = 0.273$, χ^2 Test).

The Note-taking Task. The in-class activity we chose is note-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 note-taking

[5]Course C has two parallel sections offered by the same instructor. Students could chose the one that best fit their schedule. We do not find significant differences in grades for these two sections, and pool the data together in the analysis.

production and processes (Kobayashi, 2005; Piolat et al., 2005; Bohay et al., 2011). 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. Furthermore, taking notes is simple enough, so it should not discourage low-ability students from participating.[6]

Table 1: Course Description

Course	% Female Students	# Students	# Note-takers per Lecture
A	2%	57	2, 3
B	30%	170	6, 7
C	53%	344	6, 7, 8

3.1 Experimental Procedure

In our experiment, students are randomly assigned to take notes for one lecture (the randomization period covers the third week to the end of the semester).[7] Tables A1 to 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 note-taking task. The likelihood of submitting 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. 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 can found in Appendix II.

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

[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.

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 that students take. Since we use transcript data, which is only available for Course C (60% of participants), to examine both heterogeneous and spillover effects, we restrict the following analysis to that course. Our main results are robust when we include all three courses in the analysis (see Section 4.5).

4.1 Aggregate Treatment Effects

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

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

Table 2: Treatment Effects on Attendance

	Attendance			
	(1)	(2)	(3)	(4)
Task Timing	2.336**	2.077**	2.336**	2.077**
	(1.044)	(0.988)	(1.045)	(0.988)
Task Timing^2	-0.061**	-0.055**	-0.061**	-0.055**
	(0.028)	(0.027)	(0.028)	(0.027)
Lecture Index			-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. Independent variables "Task Timing", "Task Timing^2" and "Lecture Index" are scaled by 1/100 for estimation. 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.

submission rate. Additionally, there is a natural decline for homework submission as well.

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 note-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. Our model also predicts that while students who are assigned to complete the task in the middle of

Table 3: Treatment Effects on Homework Submission

	Homework Submission			
	(1)	(2)	(3)	(4)
Task Timing	1.749***	1.676***	1.749***	1.676***
	(0.620)	(0.582)	(0.621)	(0.582)
Task Timing^2	-0.046***	-0.044***	-0.046***	-0.044***
	(0.017)	(0.016)	(0.017)	(0.016)
Homework Index			-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. Independent variables "Task Timing", "Task Timing^2" and "Homework Index" are scaled by 1/100 for estimation. 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.

the semester will exert more effort on average, this increase is essentially driven by low ability students. We explore such heterogeneous effect in the following section.

4.2 Heterogeneous Treatment Effects

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 incentives may crowd out the intrinsic motivation of low ability students. Moreover, Campos-Mercade and Wengström (2020) finds that the removal of incentives leads to worse performance from low-ability students. The intervention proposed in our papers holds promise for the low-ability students as it tries to target the moment they start to disengage from the class material.[8]

[8]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,

Table 4: Treatment Effects on Exam Grades

	Midterm Exam Grade		Final Exam Grade	
	(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.

To measure students' ability, we obtain the pre-experiment transcript data for all students in Course C, and calculate the credit-weighted GPA of each student. We use the median pre-experiment GPA to separate students into two groups, high ability (above median) and low ability (below median) 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, approximately 63% of the semester's progress.[9]

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

Bellés-Obrero (2020) find that monetary incentives for top-performance students has a positive impact on the exam grade of students with high intrinsic motivation, but a negative impact on the grades of those with low intrinsic motivation.

[9]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.

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.

Moreover, in Tables A5 and A6, we also show that Results 1 and 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.[10]

4.3 Spillover Effects

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 course was significantly affected by the task timing. Specifically, we examine the treatment effect on the students' GPA in other courses in 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 over somewhat to the subsequent semester. We summarize these findings in Result 4.

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.*

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

Table 6: Spillover Effect on Other Courses

	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.

4.4 Longterm Effects

We provide some suggestive evidence on the potential impact of our intervention on students' academic performance and, subsequently, their wages after graduation. We use two estimations as benchmarks. The first is from Zou et al. (2022), which uses survey and administrative data from an elite Chinese university, and estimates that a one standard deviation (SD) increase in GPA results in a 10.6% increase in starting wage. The second benchmark is our estimation based on data from the Chinese College Students Survey (CCSS)[11], showing that a one SD increase in GPA is associated with a 2.89% increase in starting wage. we do a simple back-of-the-envelope calculation to approximate the economic return of our intervention. Combining the effect of our intervention on the targeting course (Column 2 of Table 5) and the spillover effect in other courses (Column 2 of Table 6), the GPA for those who assigned to the middle is higher than those

[11]The survey sampled undergraduate students who completed their bachelor's degrees between 2010 and 2014, focusing on a subsample from elite Chinese universities listed under the 985 and 211 projects. This subsample includes a total of 3,581 students.

assigned to early or late. For example, on average, students that are assigned to lecture 18 for the task have, on average, a GPA that is 8.789 points higher than students who are assigned to lecture 6. Over a typical four-year undergraduate program, this translates to an average GPA increase of approximately 1.097, or around 0.174 SD.[12] The resulting wage increase would then roughly range from 0.50% to 1.84%. This is comparable to a similar estimate by Barwick et al. (2024). They find that China's gaming restrictions on college students may increase wages by 0.7%, which is about the return from half of an extra year of work experience in developing countries (Lagakos et al., 2018).

4.5 Robustness Checks

In this subsection, we conduct robustness checks for using the full sample with three courses. To conduct heterogeneous analyses on the full sample, 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 Course C.[13] 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 time frame ($p < 0.01$, two-sided t-test).

First, our main treatment effect on attendance and homework grades is robust when expanding the analysis to the full sample. As shown in Tables A7 and A8, Results 1 and 2 continue to hold. Similar to Table 4, we do not find significant results on academic performance for the full sample (Table A9).

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 A10 presents the results for the grades. The sign on all the coefficients remain the same. Table A11 presents the spillover effect results, and the

[12] In the CCSS sample, the SD of GPA is 6.328, which is comparable to the SD of 6.6 in administrative data used by Barwick et al. (2024) for Chinese undergraduate students.

[13] 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.

signs on the coefficients remain consistent.

Counterfactual Analysis. In Result 3, we show that low ability students who are assigned to the middle of the semester are more likely to have better academic performance. Due to logistical constraints, all students in our experiment were assigned to the note-taking task. This setup raises a natural question: can students who are assigned the task achieve better academic performance compared to those who are not assigned the task at all? Our counterfactual analysis provides supportive evidence.

In Figure 2, our analysis of academic performance from 2017 to 2019 in Course C shows a widening gap between high- and low-ability students. According to our theoretical framework, since the intervention better targets low-ability students, its success should be reflected in a narrowing of this gap in 2020. If the note-taking task indeed benefits low-ability students, we would expect the previously observed trend of an expanding gap to be mitigated. To examine this, we use the 2017–2019 data to estimate a linear OLS model that predicts the counterfactual final grade gap for 2020, assuming no intervention took place. Comparing the realized and counterfactual gaps, we find that the 2020 realized gap is indeed smaller, supporting the hypothesis that the intervention helped improve low-ability students' performance. Furthermore, to examine whether the timing of the task assignment affected outcomes, we divided the low-ability students in 2020 into five groups based on when they were assigned the note-taking task. The results show that, except for the earliest group, all other groups exhibit smaller grade gaps with the high-ability group compared to the counterfactual. Notably, the group assigned to the middle of the semester (Group 3) showed the smallest gap, suggesting that assigning the task mid-semester is particularly effective.

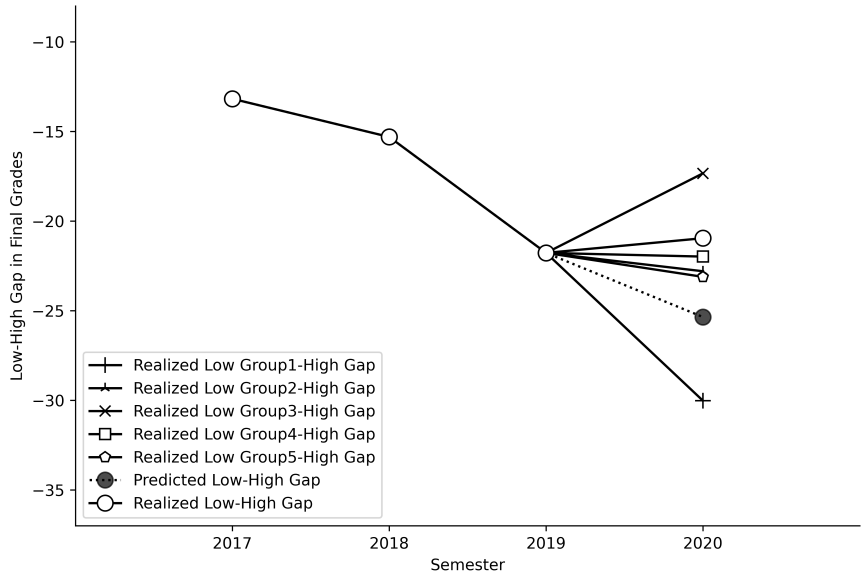


Figure 2: Counterfactual Analysis. For each semester from 2017 to 2020, we calculate the final grade gap between high- and low-ability students in Course C. Students are grouped into high and low ability based on a median split of their GPA before taking this course. The four hollow circles represent the observed grade gap in each semester. Using gaps from 2017 to 2019, we fit a linear OLS model to predict the counterfactual gap for the 2020 semester if no task assignment was implemented (black circle). In 2020, all low-ability students were further evenly split into five groups based on the timing of their task assignment: Group 1 completed the task earliest, and Group 5 latest. For each group, the grade difference between its mean score and the average score of high-ability students is plotted (shown by five distinct markers).

5 Conclusion

Declining student engagement in class over time is a notoriously challenging problem in education. Although many monetary and non-monetary interventions have been tried to stop this decline, the results have been mixed and modest at best. In this paper, we focus the often-ignored timing of the intervention and find that it is of first-order importance both theoretically and experimentally.

Our simple model illustrates the role of increasing effort costs over the duration of a learning activity on effort provision. We find that the optimal timing

for an effort-inducing intervention is the middle of the term so that students who would have disengaged before then are encouraged to "power through" and avoid catch-up costs. To test our predictions, we randomly assigned students to finish a note-taking task for different lectures. The experimental results show that assigning students to the middle of the term help students, especially low-ability ones, achieve better academic performance. We also document a spillover effect, that is, students who are assigned to the middle of the semester also achieve higher grades in other courses.

Despite the high returns to education, students are likely to become disengaged over time, leading to insufficient investment. The toolkit of behavioral economics has been widely applied in trying 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 an often overlooked dimension: timing. Our results suggest that education practitioners and policy makers could improve engagement and outcomes by timing the intervention inline with the pattern of declining engagement.

Our insights can apply to other domains that require sustained engagement as well. Note that one extension of the model (Appendix I.4) demonstrates the possibility of the temporary intervention leading to habit formation. This is likely relevant 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 timing could potentially help people pull through the period of stagnation and stay engaged even after the intervention. Moreover, with the adoption of AI technology in education (Zhai et al., 2021; Zhang and Aslan, 2021), the implementation cost for such interventions would be quite low. For example, we can use simple machine learning methods to identify which students would benefit most and when would be the optimal time for the intervention. This could be explored 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.

Bangert-Drowns, Robert L, James A Kulik, and Chen-Lin C Kulik, "Effects of frequent classroom testing," *The Journal of Educational Research*, 1991, *85* (2), 89–99.

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.

Barwick, Panle Jia, Siyu Chen, Chao Fu, and Teng Li, "Digital Distractions with Peer Influence: The Impact of Mobile App Usage on Academic and Labor Market Outcomes," Technical Report, National Bureau of Economic Research 2024.

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.

Lagakos, David, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman, "Life Cycle Wage Growth across Countries," *Journal of Political Economy*, 2018, *126* (2), 797–849.

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.

Paola, Maria De and Vincenzo Scoppa, "Frequency of examinations and student achievement in a randomized experiment," *Economics of Education Review*, 2011, *30* (6), 1416–1429.

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.

Pietro, Giorgio Di, "Exam scheduling and student performance," *Bulletin of Economic Research*, 2013, *65* (1), 65–81.

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 Morn-

ing University Classes Are Associated with Impaired Sleep and Academic Performance," *Nature Human Behaviour*, 2023, *7* (4), 502–514.

Zhai, Xuesong, Xiaoyan Chu, Ching Sing Chai, Morris Siu Yung Jong, Andreja Istenic, Michael Spector, Jia-Bao Liu, Jing Yuan, and Yan Li, "A Review of Artificial Intelligence (AI) in Education from 2010 to 2020," *Complexity*, 2021, *2021* (1), 8812542.

Zhang, Ke and Ayse Begum Aslan, "AI Technologies for Education: Recent Research & Future Directions," *Computers and Education: Artificial Intelligence*, 2021, *2*, 100025.

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

I Proofs

I.1 Derivation of the lower bound of λ

To determine the lower bound of a penalty factor λ that discourages catch-up behavior, we begin by considering the total discounted cost:

$$C = \frac{1}{a}\left[\int_0^1 \mathbb{1}_{e_t=1} e^{-\delta t} c(t)dt + \sum_{k=1}^{K} e^{-\delta t_{2k-1}} \int_{t_{2k-2}}^{t_{2k-1}} c(s)ds\right],$$

where the first term represents the cost of continuous effort and the second term the cost incurred when catching up after breaks.

When a penalty factor λ is imposed, the catch-up cost is modified so that the total cost becomes

$$C = \frac{1}{a}\left[\int_0^1 \mathbb{1}_{e_t=1} e^{-\delta t} c(t)dt + \sum_{k=1}^{K-1} \lambda e^{-\delta t_{2k+1}} \int_{t_{2k}}^{t_{2k+1}} c(s)ds\right]. \tag{3}$$

To render the extreme strategy—skipping all lectures and catching up only at the end—suboptimal, we require that, when effort is postponed until $t = 1$, the catch-up cost still exceeds the cost of continuous effort. That is,

$$\lambda e^{-\delta} \int_0^1 c(s)ds > \int_0^1 e^{-\delta s} c(s)ds.$$

Assuming a linear cost function $c(t) = t$, the inequality immediately simplifies to

$$\lambda > \frac{2(e^{\delta} - (1+\delta))}{\delta^2}.$$

I.2 Generalized cost density $c(t)$

We now consider a more general cost function $c(t)$. To guarantee a well-behaved solution to the student's maximization problem, we impose the following technical assumption on the curvature of $c(t)$:

Assumption 1. *For all* $t_x, t_y \in [0,1]$,

$$\frac{\partial^2 c(t)}{\partial t^2}\Big|_{t=t_x} \cdot \frac{\partial^2 c(t)}{\partial t^2}\Big|_{t=t_y} \geq 0.$$

This assumption rules out cost functions that are concave on some intervals and convex on others; functions that are uniformly linear, concave, or convex on $[0,1]$ satisfy Assumption 1.

We can now derive the optimal stopping rule of a student with ability a and an educational intervention assigned at time t^n:

Lemma 1. *1. If* $a \geq \min\{\bar{a}, a^h\}$, $t^{s^*}(a, t^n) = 1$.

2. If $a < \min\{\bar{a}, a^h\}$,

$$t^{s^*}(a, t^n) \in \begin{cases} \{t_0(a)\}, & \text{if } t^n \leq t_0(a) \\ \{t_0(a), t^n, 1\}, & \text{if } t^n > t_0(a) \end{cases}$$

where

$$\bar{a} \coloneqq \sup_{t \in [0,1]} \frac{e^{-\delta t} c(t)}{e^{-\delta}},$$

$$a^h \coloneqq \{a \mid \Delta U(a, 1, t_0(a)) = 0\},$$

and

$$t_0(a) \coloneqq \min\Big\{t \in (0,1) \;\Big|\; \frac{\partial V(a, t^s, t^n)}{\partial t^s} = 0\Big\}.$$

Proof. First, note that the derivative of the total discounted utility is given by

$$\frac{\partial V(a, t^s, t^n)}{\partial t^s} = e^{-\delta} - \frac{1}{a} e^{-\delta t^s} c(t^s).$$

Case 1: Suppose that $a > \bar{a}$. Then for every $t^s \in [0,1]$ we have

$$e^{-\delta} - \frac{1}{a} e^{-\delta t^s} c(t^s) > 0.$$

In addition, the intervention penalty term $-\mathbb{1}_{\{t^s < t^n\}} e^{-\delta t^n} S$ is non-decreasing in

t^s. Consequently, $V(a, t^s, t^n)$ is strictly increasing in t^s, and the optimal stopping time is attained at the terminal time, i.e.,

$$t^{s^*}(a, t^n) = 1.$$

Case 2: Now consider when $a \leq \bar{a}$. In this case, there exists some $t' \in (0, 1]$ such that

$$\frac{\partial V(a, t^s, t^n)}{\partial t^s}\Big|_{t^s=t'} \leq 0,$$

while at $t^s = 0$ the derivative satisfies

$$\frac{\partial V(a, t^s, t^n)}{\partial t^s}\Big|_{t^s=0} > 0.$$

By the Intermediate Value Theorem, there exists at least one t in the interval $[0, t']$ for which

$$\frac{\partial V(a, t^s, t^n)}{\partial t^s}\Big|_{t^s=t} = 0.$$

Assumption 1 guarantees that this equation has at most two roots; we denote the smallest such root by

$$t_0(a) = \min\Big\{t \in [0, 1] \ : \ \frac{\partial V(a, t^s, t^n)}{\partial t^s} = 0\Big\},$$

which corresponds to the unique local maximum of $V(a, t^s, t^n)$.

Among the candidate stopping times, note that stopping at $t = 0$ can be ruled out because

$$V(a, t_0(a), t^n) \geq V(a, 0, t^n).$$

It remains to compare the utility from stopping at $t_0(a)$, at t^n, and at the end of the semester ($t^s = 1$). We consider the following subcases:

Case 2a: Suppose $\bar{a} \geq a \geq a^h := \{a \mid \Delta U(a, 1, t_0(a)) = 0\}$, there exist two situations. First, if $\frac{\partial V(a,t^s,t^n)}{\partial t^s} \mid_{t^s=t} = 0$ only has one root $t_0(a)$ on $[0, 1]$, then $\bar{a} = a^h$, $t^{s^*}(a, t^n) = 1$. Second, if $\frac{\partial V(a,t^s,t^n)}{\partial t^s} \mid_{t^s=t} = 0$ has two roots on $[0, 1]$, we denote the larger one as $t_{-1}(a)$.

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

Proof.

$$\Delta U(a,1,t_0(a)) = \int_{t_0(a)}^{1} \frac{\partial U(a,t^s)}{\partial t^s} dt$$

Following the Leibniz integral rule, we have:

$$\begin{aligned} \frac{\partial [\Delta U(a,1,t_0(a))]}{\partial a} &= -\frac{\partial U(a,t^s)}{\partial t^s} \mid_{t^s=t_0(a)} \cdot \frac{\partial t_0(a)}{\partial a} + \int_{t_0(a)}^{1} \frac{\frac{\partial U(a,t^s)}{\partial t^s}}{\partial a} dt \\ &= \int_{t_0(a)}^{1} \frac{\frac{\partial U(a,t^s)}{\partial t^s}}{\partial a} dt \end{aligned} \tag{4}$$

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

Remark 1 implies that for $a \geq a^h$, $\Delta U(a,1,t_0(a)) \geq 0$. $V(a,1,t^n) = U(a,1) > U(a,t_0(a)) \geq V(a,t_0(a),t^n)$, hence $t^{s^*}(a,t^n) \neq t_0(a)$.

We further consider the following subcases,

- If $t^n < t_0(a)$, because $\forall t^s \in [0,t_0(a)]$, $\frac{\partial V(a,t^s,t^n)}{\partial t^s} > 0$, so that

$$V(a,t_0(a),t^n) > V(a,t^n,t^n)$$

  In this situation, $t^{s^*}(a,t^n) = 1$.

- If $t^n > t_0(a)$, we distinguish two further subcases:

  - If $t^n \in [t_0(a), t_{-1}(a))$, then the derivative satisfies $\frac{\partial V(a,t^s,t^n)}{\partial t^s} < 0$ in this interval. This implies that

$$\Delta U(a,1,t^n) = \underbrace{\Delta U(a,1,t_0(a))}_{\geq 0} - \underbrace{\Delta U(a,t_0(a),t^n)}_{< 0} > 0.$$

    Hence

$$V(a,1,t^n) > V(a,t^n,t^n),$$

so that $t^{s^*}(a, t^n)$ cannot equal t^n; therefore, the optimal choice is again $t^{s^*}(a, t^n) = 1$.

– If $t^n \in [t_{-1}(a), 1)$, then the derivative becomes positive for $t^s \geq t_{-1}(a)$, so that

$$V(a, 1, t^n) = U(a, 1) > U(a, t^n) = V(a, t^n, t^n),$$

and once more $t^{s^*}(a, t^n) \neq t^n$, leading to $t^{s^*}(a, t^n) = 1$.

To summarize, when $a \geq a^h$, $t^{s^*}(a, t^n) = 1$.

Case 2b: When $a < a^h$, further consider the location of t^n,

Case 2b-i: If $t^n < t_0(a)$, 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 $t^{s^*}(a, t^n) = t^n$. As $a < a^h$, $V(a, t_0(a), t^n) = U(a, t_0(a) > U(a, 1) = V(a, 1, t^n)$, we can rule out $t^{s^*}(a, t^n) = 1$ either. Therefore, $t^{s^*}(a, t^n) = t_0(a)$.

Case 2b-ii: If $t^n > t_0(a)$, there exist two situations similar in Case 2a.

(1) If $\frac{\partial V(a,t^s,t^n)}{\partial t^s} \mid_{t^s=t} = 0$ only has one root $t_0(a)$ on $[0, 1]$. As $\forall t^s > t_0(a)$, $\frac{\partial V(a,t^s,t^n)}{\partial t^s} < 0$, $V(a, 1, t^n) < V(a, t^n, t^n)$, we can rule out $t^{s^*}(a, t^n) = 1$, hence $t^{s^*}(a, t^n) \in \{t_0(a), t^n\}$.

(2) If $\frac{\partial V(a,t^s,t^n)}{\partial t^s} \mid_{t^s=t} = 0$ has two roots on $[0, 1]$, the comparisons between candidate stopping times are indeterminate, $t^{s^*}(a, t^n) \in \{t_0(a), t^n, 1\}$.

□

Lemma 1 has several implications. First, students with high ability ($a \geq \min\{\bar{a}, a^h\}$) will always exert effort in all lectures regardless of t^n since the marginal benefit from exerting effort always exceeds their marginal cost of effort. Second, for students with low ability ($a < \min\{\bar{a}, a^h\}$), the optimal stopping point $t^{s^*}(a, t^n)$ lies in a limited set. Additionally, whenever the timing of the task is earlier than $t_0(a)$ (i.e., the earliest timing where the marginal benefit of effort equals the marginal cost), t^n will not be effective in increasing students' effort. 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.

I.3 Proof of Proposition 1

Proof. Notice that when $c(t) = t$ and $\delta \leq 1$, $\bar{a} = a^h = 1$.

1. When $a \geq 1$,

   By Lemma 1, we have $t^{s^*}(a, t^n) = 1$. We distinguish following cases based on the student's ability.

2. When $a < 1$, note that $\forall t^s \in [0, 1]$,

$$\frac{\partial^2 V(a, t^s, t^n)}{\partial (t^s)^2} = -\frac{1}{a}(e^{-\delta t^s} - \delta e^{-\delta t^s} t^s) \leq 0.$$

   Thus, by concavity, there exists a unique threshold

$$t_0(a) = -\frac{1}{\delta} W_0(-a\delta e^{-\delta})$$

   satisfying $\frac{\partial V(a,t^s,t^n)}{\partial t^s}\big|_{t^s = t_0(a)} = 0$.

   - If $t^n \leq t_0(a)$, Lemma 1 implies that $t^{s^*}(a, t^n) = t_0(a)$.
   - If $t^n > t_0(a)$, the situation (1) of **Case 2b-ii** in Lemma 1 implies that $t^{s^*}(a, t^n) \in \{t_0(a), t^n\}$. we compare $V(a, t^n, t^n)$ and $V(a, t_0(a), t^n)$. Since

$$V(a, t^n, t^n) - V(a, t_0(a), t^n) = \Delta U(a, t^n, t_0(a)) + e^{-\delta t^n} S$$

     is continuous and strictly decreasing for $t^n > t_0(a)$, we introduce the following auxiliary threshold,

     Effective Threshold:

$$t_E(a) \coloneqq \min\Big\{t \in [t_0(a), 1] \ : \ \Delta U(a, t, t_0(a)) \leq -e^{-\delta t} S\Big\}.$$

     $t_E(a)$ marks the minimal task time beyond which the cost of effort overwhelms any benefit from aligning with the intervention.

– If $t^n > t_E(a)$, $V(a, t^n, t^n) - V(a, t_0(a), t^n) < 0$, $t^{s^*}(a, t^n) = t_0(a)$.

– If $t^n \in (t_0(a), t_E(a)]$, $V(a, t^n, t^n) - V(a, t_0(a), t^n) \geq 0$, $t^{s^*}(a, t^n) = t^n$.

□

I.4 Habit Formation with High Discount Rates

In Section 2 we assumed $\delta \leq 1$ so that the total utility function was globally concave in the stopping time. In this Section, we relax this assumption and consider steeper rates of discounting, $\delta > 1$, with $c(t) = t$. In that case, the loss of global concavity permits an educational intervention to induce not only a temporary increase in effort but, in certain cases, the formation of a learning habit — i.e., sustained effort until $t = 1$.

Notation and Thresholds

Auxiliary Thresholds. To capture the role of the intervention, we further define two additional thresholds:

- *Habit Induction Threshold*:

$$t_H(a) \coloneqq \max\Big\{t \in [t_0(a), 1] \ : \ \Delta U(a, 1, t_0(a)) \geq -e^{-\delta t} S\Big\}.$$

  This threshold characterizes the latest intervention time for which the net utility loss of extending effort from $t_0(a)$ to the end of the semester is smaller than the utility loss from not completing the task.

- *Continuation Threshold*:

$$t_C(a) \coloneqq \min\Big\{t \in [t_0(a), 1] \ : \ \Delta U(a, t, 1) \leq 0\Big\}.$$

  $t_C(a)$ is the minimal intervention time such that the incremental benefit of extending effort from t to the end of the semester becomes non-positive.

Ability Cutoffs. We further classify students by introducing the following cutoff:

$$a^l \coloneqq \{a \mid \Delta U(a, 1, t_0(a)) = -e^{-\delta t_H(a)} S\}.$$

Students with $a < a^l$ are very low-ability for whom the intervention has limited influence.

Proposition 2. *Under $\delta > 1$ with $c(t) = t$, the optimal stopping rule*

$$t^{s^*}(a, t^n) = \arg\max_{t^s \in [0,1]} V(a, t^s, t^n)$$

satisfies:

(i) If $a \geq a^h$, then

$$t^{s^*}(a, t^n) = 1.$$

(ii) If $a^l \leq a < a^h$, then

$$t^{s^*}(a, t^n) = \begin{cases} t_0(a), & \text{if } 0 \leq t^n \leq t_0(a), \\ t^n, & \text{if } t_0(a) < t^n \leq t_C(a), \\ 1, & \text{if } t_C(a) < t^n \leq t_H(a), \\ t_0(a), & \text{if } t^n > t_H(a). \end{cases}$$

(iii) If $a < a^l$, then

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

Proof. We prove the proposition by comparing the total utility obtained by stopping at the candidate times $t_0(a)$, t^n, and 1. We consider $t^{s^*}(a, t^n)$ for different a.

1. If $a \geq \bar{a} > a^h$, or $\bar{a} > a \geq a^h$, by Lemma 1,

$$t^{s^*}(a, t^n) = 1.$$

2. Suppose $a^l \leq a < a^h$,

   - If $t^n \leq t_0(a)$, by Lemma 1,

   $$t^{s^*}(a, t^n) = t_0(a).$$

   - If $t_0(a) < t^n \leq t_C(a)$: By the definition of $t_C(a)$ and $t_H(a)$, for such t^n we have

   $$\Delta U(a, t^n, 1) \geq 0, \quad \Delta U(a, 1, t_0(a)) \geq -e^{-\delta t} S,$$

   Hence, the optimal choice is

   $$t^{s^*}(a, t^n) = t^n.$$

   - If $t_C(a) < t^n \leq t_H(a)$:

   $$\Delta U(a, t^n, 1) < 0, \quad \Delta U(a, 1, t_0(a)) \geq -e^{-\delta t} S,$$

   so that

   $$t^{s^*}(a, t^n) = 1.$$

   - If $t^n > t_H(a)$:

   $$\Delta U(a, t^n, 1) < 0, \quad \Delta U(a, 1, t_0(a)) < -e^{-\delta t} S,$$

   which implies

   $$t^{s^*}(a, t^n) = t_0(a).$$

3. Finally, suppose $a < a^l$. 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 Proposition 1,

- If $t^n \leq t_0(a)$, by Lemma 1,

$$t^{s^*}(a, t^n) = t_0(a).$$

- If $t_0(a) < t^n \leq t_E(a)$, the definition of $t_E(a)$ guarantees that

$$\Delta U(a, t^n, t_0(a)) \geq -e^{-\delta t^n} S,$$

  Hence, $t^{s^*}(a, t^n) = t^n$.

- If $t^n > t_E(a)$,

$$\Delta U(a, t^n, t_0(a)) < -e^{-\delta t^n} S,$$

  so that $V(a, t_0(a), t^n) > V(a, t^n, t^n)$. Therefore, $t^{s^*}(a, t^n) = t_0(a)$.

□

Figure A1 presents a numerical example of Proposition 2. 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, assuming the probability of each ability type is 1/3, we find that the aggregated effort (per student) is maximized at $t^n = 0.85$.

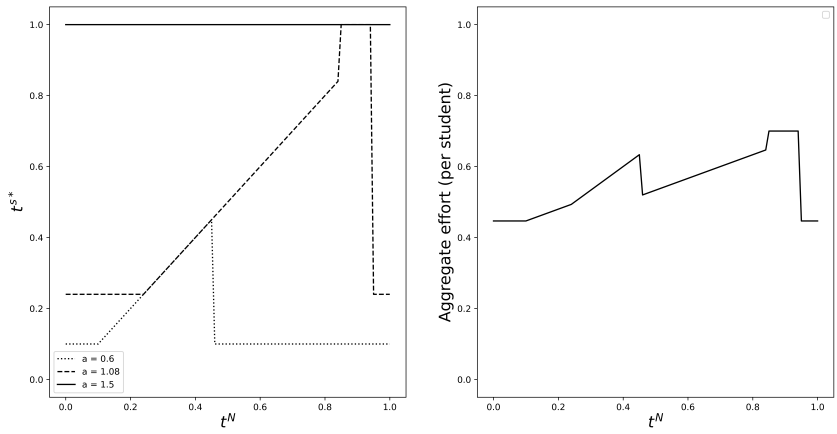


Figure A1: Simulation of Proposition 2. We set $\delta = 2$, $S = 0.1$ and do the numerical simulation. For the figure on the left, we consider how the optimal stopping point of students with respectively three types of ability, $a = 1.5$ (solid line), $a = 1.08$ (dashed line), and $a = 0.6$ (dotted 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 aggregate effort (per student) depends on task timing t^n.

II Experiment Implementation Details

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 Web Learning[14] and WeChat group[15] to ask students to complete the survey.

Lecture 5: Announce Experiment. In lecture 5, we announce the detailed experiment schedule in the classroom. We also post the note-taking schedule on both the Web Learning and WeChat groups before.

Lecture i: Note-taking. One day before Lecture i, we send a reminder email to students assigned to that particular lecture, and also post notice on 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 note-taking task is not compulsory and does not affect the final grade.

Post-experiment Survey. We distribute the post-experiment survey after the final exam via the Web Learning and WeChat group.

[14] Web Learning is a website for teaching and learning purpose. On this website, teachers and teaching assistants could post notice, arrange assignments, and share slides, while students could access course materials and information.

[15] 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.

III Randomization Check

Table A1: Course A Randomization

Lecture Index	% Female	% Postpone	% STEM	% International
6	0.000	0.333	1.000	0.000
7	0.000	0.000	1.000	0.000
8	0.000	1.000	1.000	0.000
9	0.000	0.000	1.000	0.000
10	0.000	0.000	1.000	0.000
11	0.000	0.000	1.000	0.000
12	0.000	0.000	1.000	0.000
13	0.000	0.000	1.000	0.000
14	0.000	0.000	1.000	0.000
16	0.000	0.500	1.000	0.000
17	0.000	0.000	1.000	0.000
18	0.000	0.000	1.000	0.000
19	0.000	0.000	1.000	0.000
20	0.000	0.000	1.000	0.000
21	0.000	0.333	1.000	0.000
22	0.000	0.000	1.000	0.000
23	0.000	0.500	1.000	0.000
24	0.000	0.000	1.000	0.000
25	0.000	0.000	1.000	0.000
26	0.000	0.500	1.000	0.000
27	0.500	0.000	1.000	0.000
28	0.000	0.000	1.000	0.000
29	0.000	0.000	1.000	0.000
30	0.000	0.000	1.000	0.000
31	0.000	0.500	1.000	0.000
χ^2 test p-value	0.260	0.206	/	/

Notes: This table reports, for each lecture (Lecture Index), the proportion of students assigned to complete the note-taking task during that specific lecture who exhibit certain characteristics. These characteristics include gender (Female), whether the student postponed taking the course from the recommended schedule (Postpone), whether the student majors in a STEM field (STEM), and whether they are an international student (International). In the experiment setup, we excluded the initial four lectures as they were part of a shopping period, announced the experiment schedule in lecture 5, and then randomly assigned students from lecture 6 to lecture 31. Note that for this course, lecture 15 was reserved for a midterm exam.

Table A2: Course B Randomization

Lecture Index	% Female	% Postpone	% STEM	% International
6	0.143	0.000	1.000	0.000
7	0.429	0.000	0.571	0.000
8	0.167	0.000	0.667	0.000
9	0.143	0.000	1.000	0.000
10	0.167	0.000	0.667	0.000
11	0.286	0.000	0.714	0.000
12	0.167	0.000	1.000	0.167
13	0.571	0.000	0.143	0.000
14	0.667	0.167	0.667	0.000
15	0.143	0.000	0.714	0.000
16	0.500	0.000	1.000	0.000
17	0.429	0.000	0.714	0.000
18	0.333	0.000	0.833	0.000
19	0.429	0.000	0.571	0.000
20	0.286	0.000	0.714	0.143
21	0.500	0.000	0.833	0.000
22	0.429	0.000	0.714	0.000
23	0.167	0.167	0.667	0.167
24	0.286	0.000	0.571	0.000
25	0.333	0.000	0.500	0.000
26	0.286	0.000	0.714	0.000
27	0.333	0.167	0.667	0.000
28	0.286	0.000	1.000	0.000
29	0.333	0.000	0.833	0.000
30	0.143	0.000	0.714	0.000
31	0.000	0.000	0.500	0.000
χ^2 Test p-value	0.815	0.419	0.189	0.496

Notes: This table reports, for each lecture (Lecture Index), the proportion of students assigned to complete the note-taking task during that specific lecture who exhibit certain characteristics. These characteristics include gender (Female), whether the student postponed taking the course from the recommended schedule (Postpone), whether the student majors in a STEM field (STEM), and whether they are an international student (International). In the experiment setup, we excluded the initial four lectures as they were part of a shopping period, announced the experiment schedule in lecture 5, and then randomly assigned students from lecture 6 to lecture 31.

Table A3: Course C1 Randomization

Lecture Index	% Female	% Postpone	% STEM	% International
6	0.286	0.714	0.000	0.143
7	0.714	0.429	0.000	0.000
8	0.429	0.429	0.000	0.000
9	0.500	0.500	0.000	0.333
10	0.571	0.429	0.000	0.143
11	0.714	0.429	0.000	0.143
12	0.500	0.667	0.000	0.000
13	0.714	0.286	0.000	0.000
14	0.429	0.429	0.000	0.286
16	0.500	0.333	0.000	0.000
17	0.571	0.143	0.000	0.143
18	0.857	0.571	0.000	0.286
19	0.167	0.333	0.000	0.000
20	0.286	0.571	0.000	0.000
21	0.429	0.714	0.000	0.286
22	1.000	0.667	0.000	0.167
23	0.429	0.143	0.000	0.143
24	0.714	0.429	0.000	0.000
25	0.667	0.500	0.000	0.333
26	0.429	0.286	0.000	0.000
27	0.286	0.429	0.000	0.143
28	0.500	0.500	0.000	0.000
29	0.571	0.571	0.000	0.286
30	0.429	0.429	0.000	0.000
31	0.667	0.333	0.000	0.167
χ^2 Test p-value	0.506	0.901	/	0.526

Notes: This table reports, for each lecture (Lecture Index), the proportion of students assigned to complete the note-taking task during that specific lecture who exhibit certain characteristics. These characteristics include gender (Female), whether the student postponed taking the course from the recommended schedule (Postpone), whether the student majors in a STEM field (STEM), and whether they are an international student (International). In the experiment setup, we excluded the initial four lectures as they were part of a shopping period, announced the experiment schedule in lecture 5, and then randomly assigned students from lecture 6 to lecture 31. Note that for this course, lecture 15 was reserved for a midterm exam.

Table A4: Course C2 Randomization

Lecture Index	% Female	% Postpone	% STEM	% International
6	0.625	0.375	0.000	0.000
7	0.571	0.143	0.000	0.286
8	0.571	0.286	0.000	0.286
9	0.857	0.000	0.000	0.286
10	0.714	0.143	0.000	0.143
11	0.429	0.000	0.000	0.286
12	0.429	0.286	0.000	0.143
13	0.857	0.143	0.000	0.143
14	0.286	0.143	0.000	0.286
16	0.714	0.286	0.000	0.286
17	0.857	0.000	0.000	0.143
18	0.714	0.000	0.000	0.143
19	0.250	0.000	0.000	0.125
20	0.571	0.143	0.000	0.286
21	0.857	0.286	0.000	0.286
22	0.571	0.286	0.000	0.000
23	0.714	0.143	0.000	0.000
24	0.571	0.000	0.000	0.143
25	0.429	0.429	0.000	0.571
26	0.714	0.000	0.000	0.000
27	0.143	0.286	0.000	0.143
28	0.286	0.143	0.000	0.429
29	0.571	0.286	0.000	0.143
30	1.000	0.143	0.000	0.143
31	0.571	0.143	0.000	0.286
χ^2 Test p-value	0.109	0.609	/	0.685

Notes: This table reports, for each lecture (Lecture Index), the proportion of students assigned to complete the note-taking task during that specific lecture who exhibit certain characteristics. These characteristics include gender (Female), whether the student postponed taking the course from the recommended schedule (Postpone), whether the student majors in a STEM field (STEM), and whether they are an international student (International). In the experiment setup, we excluded the initial four lectures as they were part of a shopping period, announced the experiment schedule in lecture 5, and then randomly assigned students from lecture 6 to lecture 31. Note that for this course, lecture 15 was reserved for a midterm exam.

IV Additional Tables

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

	Attendance			
	(1)	(2)	(3)	(4)
Task Timing	2.668*	2.391*	2.277	2.215
	(1.517)	(1.440)	(1.464)	(1.379)
Task Timing^2	-0.072*	-0.067*	-0.057	-0.056
	(0.041)	(0.038)	(0.040)	(0.038)
Lecture Index	-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. Independent variables "Task Timing", "Task Timing^2", and "Lecture Index" are scaled by 1/100 for estimation. 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 Submission

	Homework Submission			
	(1)	(2)	(3)	(4)
Task Timing	3.300***	3.110***	0.394	0.376
	(1.145)	(1.051)	(0.441)	(0.413)
Task Timing^2	-0.088***	-0.084***	-0.009	-0.008
	(0.031)	(0.028)	(0.011)	(0.010)
Homework Index	-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 submit a homework or not, on the task timing and its quadratic term. Independent variables "Task Timing", "Task Timing^2", and "Homework Index" are scaled by 1/100 for estimation. 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 A7: Treatment Effects on Attendance, Robustness Table

	Attendance			
	(1)	(2)	(3)	(4)
Task Timing	1.875**	1.525**	1.879**	1.523**
	(0.733)	(0.682)	(0.735)	(0.683)
Task Timing^2	-0.050**	-0.041**	-0.050**	-0.041**
	(0.020)	(0.018)	(0.020)	(0.018)
Lecture Index			-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. Independent variables "Task Timing", "Task Timing^2", and "Lecture Index/100" are scaled by 1/100 for estimation. 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 A8: Treatment Effects on Homework Submission, Robustness Table

	Homework Submission			
	(1)	(2)	(3)	(4)
Task Timing	1.175**	1.103**	1.178**	1.103**
	(0.512)	(0.469)	(0.512)	(0.469)
Task Timing^2	-0.029**	-0.027**	-0.030**	-0.027**
	(0.013)	(0.012)	(0.013)	(0.012)
Homework Index			-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. Independent variables "Task Timing", "Task Timing^2", and "Homework Index" are scaled by 1/100 for estimation. 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 A9: 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.

Table A10: 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 A11: Spillover Effect on Other Courses, 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.