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

Tracy Xiao Liu* Ulrike Malmendier† Stephanie W. Wang‡

Shuhuai Zhang§

September 17, 2022

Preliminary and incomplete.

Please do not cite.

Abstract

Personal plans of longer-term personal improvements or investment in one's human capital often suffer from significant decay over time. In education, a common phenomenon is the decay in attendance and engagement in lectures over the course of a semester. The main response In this field experiment, we study how the timing of a potentially rewarding but cognitively taxing assignment, taking notes and posting them, affects students' academic behavior and performance. We find that assigning such tasks to low-performing students in the middle of the term, compared to early or late in the semester, improves their performance more along numerous dimensions: attendance, homework grades, and exam grades. We argue that, rather than "early intervention," possibly with the goal of habit formation in studying, or "crunch time intervention," engagement interventions are most effective if they target the time when students start to fall off because of accumulating frictions and complications in their semester schedule.

*liuxiao@sem.tsinghua.edu.cn

†ulrike@berkeley.edu

‡swwang@pitt.edu

§zhangshh.20@pbcsf.tsinghua.edu.cn

1 Introduction

Improving students' engagement and performance is a central challenge in education. Babcock and Marks (2011) documents that US college students decreased their time invested in studying from 40 hours per week in 1961 to 27 hours in 2003, while Banerjee and Duflo (2014) shows the significant within-course engagement decay problem in an online setting. While previous studies have varied the size, target, and other features of various monetary and non-monetary incentives for academic behavior and performance, we focus a crucial but less studied dimension: the timing of the incentive.

Students are asked to do a cognitively taxing task, notes-taking, for a randomly selected lecture. We find that assigning low-performing students to the middle of the term, compared to earlier or later weeks, improves their performance more, e.g. higher attendance, homework grades, and exam grades. Using administrative data from previous years to build a counterfactual, we show that the notes-taking task has improves the overall final grades for low-performing students. We also document a spillover effect since students who are assigned to the middle weeks also performed better in their other courses that term. We calculate the welfare effect of the treatment using the Chinese College Students survey data on labor market outcomes — compared to assignment in the early, an assignment in the middle weeks increases students' monthly wage after graduation by 8.08 RMB.

The rest of the paper is organized as follows. Section 3 introduces a theoretical framework. Section 4 presents the experimental design. Section 5 discusses the results, and Section 6 concludes.

2 Literature Review

Our study is closely related to field experiments which examine effect of various instruments in motivating students' learning. Depending on whether the intervention involves money transfer, we can simply categorize them into studies with and without monetary incentives. For monetary incentives, we refer the readers to Gneezy et al. (2011) which presents a nice review of the effectiveness of monetary incentive across domains such as education, prosocial behavior and lifestyle habits. Another survey study by Lavecchia et al. (2016) summarizes main findings for the implications of behavioral economics in the education setting.[1]

Compared to monetary incentives, interventions without monetary incentives is more cost-effective and do not suffer from either crowding out (Gneezy and Rustichini, 2000) or "choking" effect (Ariely et al., 2009). In the experiment of Kizilcec et al. (2020), several non-monetary interventions are implemented in an online education setting. One common intervention is to improve students' perceived education return. By randomly offering the return information to students, Jensen (2010) find that those eighth-grade students who receive educational return information complete 0.2 -

[1]By explicitly manipulating the material gains for subjects, monetary incentives have been widely used in education setting, such as scholarships for outstanding students.[2] Prior studies have examined various factors in the choice sets for monetary incentives. The first one is who receives incentive. The most common target is students (Kremer et al., 2009; Angrist et al., 2002, 2009; Angrist and Lavy, 2009; Braun et al., 2011; Levitt et al., 2016a), and it could be instructors (Fryer et al., 2018), or administrators (Behrman et al., 2015). In particular, Behrman et al. (2015) compares which targeting group responds to the incentive most and find that giving incentives to students, instructors and administrators is the most effective one. The second one is which rewarding criterion should be used. For example, Fryer (2011) finds that incentivizing on education inputs, e.g., attendance, is more effective than rewarding on education outputs such as grades. Bellés-Obrero (2020) studies three different incentive schemes in the online education setting: threshold incentive where students who pass the threshold receives reward, top percentile for students in the top of their class, and improvement for those who improve their expected grade, and find significant interaction effect between incentive schemes and students type, e.g., rewarding top students has positive (negative) impact on those with high (low) intrinsic motivation. Last but not least, how to frame the incentive may also matter, though the result is also mixed (Fryer et al., 2018; Levitt et al., 2016b).

0.35 more years of school over the next 4 years. Similarly, Fryer (2016) find that providing information about human capital and outcomes to students in their sixth and seventh grades could improve their score for college entrance exam in later life. Another common practice is to setup goals for students, and previous studies center on the design of goals, e.g., performance-based goals vs. task-based goals (Clark et al., 2020), whether reward the achieving of goals (Rezaei et al., 2021), and etc. However, the effects are mixed - positive effects on academic performance are shown in Morisano et al. (2010); Clark et al. (2020); Rezaei et al. (2021), while Dobronyi et al. (2019) finds no effect of goal-setting on GPA, course credits, or second year persistence. Besides, the usage of commitment device is also popular in previous experiments, with mixed results as well Ariely and Wertenbroch (2002); Burger et al. (2011); Bisin and Hyndman (2020). In some studies that compare commitment device with other behavioral tools Patterson (2018); Himmler et al. (2019), commitment device is proven to outperform other tools.

Significant treatment heterogeneity is observed across studies with and without monetary incentives and this motivates our study by examining the treatment effect across different sub-samples. First, demographic variables matter, which include income (Gneezy et al., 2019), age (Levitt et al., 2016b), and gender (Angrist and Lavy, 2009; Angrist et al., 2006, 2009; Jie Gong and Tang, 2021). Second, course type matters. For example, Bettinger (2012) find that monetary incentives is only effective for elementary school students' math scores, while the grades for reading and social science are not affected. Third, students' inner motivation matters. Bellés-Obrero (2020) find that rewarding monetary incentives to top-performance students has positive (negative) impact on the exam grade of those with high (low) intrinsic motivation. Last but not least, the ability of students matters. Leuven et al. (2010) finds that large reward increase the exam pass rate of high ability students, while decreasing the pass rate of low ability students. They propose that the financial incentive may exert displacement effect that decreases the intrinsic motivation of low ability students. Our paper is closely related to the last strand of literature.

Most prior studies provide temporary incentives to subjects, and many "well-specified" and "well-targeted" incentives are proven to be effective in the short-term (Gneezy et al., 2011), while the impact of incentive campaigns in the long run are quite mixed. Some studies show that the treatment effects in the short-term are likely to diminish or reverse in the long-term (Ferraro and Price, 2013; Allcott and Rogers, 2014; Kesternich et al., 2016; Ito et al., 2018; Brandon et al., 2019; Jessoe et al., 2021). In contrast, other studies document long-term impacts that are consistent with short-term impacts, especially in the case of cultivating long-term habits, e.g. exercising (Charness and Gneezy, 2009) and saving (Gertler et al., 2018). Therefore, understanding how temporary intervention affects the long-term outcomes even after the removal of incentives is important. One explanation is that temporary interventions could potentially alter underlying inner incentives. On the positive side, the short-term incentive may help develop habit in the long-run (Charness and Gneezy, 2009), while it may crowd out intrinsic motivations (Meier, 2007). These two competing drives may result in heterogeneous long-term effects across settings. Based on prior studies, we are also interested in understanding the effectiveness of a non-monetary instrument on individuals' repeated and long-term behavior, i.e., their regular attendance and homework grades.

A small but growing literature have examined the importance of timing of incentive implementation. Many of them find that a surprising reward given afterwards is more effective than the same amount of monetary reward ahead (Gneezy and List, 2006; Bellemare and Shearer, 2009). Ockenfels et al. (2015) further document that splitting the surprising reward to two-stages produces significantly higher effort from workers. Similar findings are also report in Gilchrist et al. (2016).[3]

3 Model

A semester has T lectures. At the beginning of the semester (t = 0), a student makes a plan and decides whether to exert effort ($e_{it} = 1$) or not ($e_{it} = 0$) in each of the T lectures, and δ is the discount rate. A student's effort cost is $\frac{1}{a}c(t,k)$, where k is the number of lectures learned thus far, and $a \in [\underline{a}, \bar{a}]$ captures the student's ability, $\underline{a} > 0$. Next, we introduce the following assumptions and derive the optimal learning path for students.

Assumption 1 (Student Cost Function)**.** $\forall k < t-1, c(t,k) > c(t,t-1) + \delta^{-1}c(t-1,t-2) + ... + \delta^{t-k-1}c(k+1,k)$

This assumption states that missing any lecture significantly increases the cost of learning subsequent lectures such that it is never optimal for students to skip a lecture and learn the subsequent lectures. This is certainly the case for the math courses of this experiment. Under this assumption, a student's optimal strategy would be a threshold one of "keep exerting effort until t^s, and stop exerting effort afterwards". Therefore, $\forall t \leq t^s$, $k = t-1$ and $c(t,k) = c(t)$.

Assumption 2 (Sigmoid Cost Function)**.** *The cost function is a sigmoid function.*

A sigmoid function, which exhibits a "S" shape, is widely used to model learning curves in psychology(Newell and Rosenbloom, 1981; Leibowitz et al., 2010), economics(Hébert and Woodford, 2021) and artificial intelligence(Gibbs and MacKay, 2000). It fits the following observations for

[3] A lab experiment by Boosey and Goerg (2020) suggests that subjects' output is significantly higher when the bonus is paid in the middle instead of upfront or at the end.

learning costs in the classroom. First, the learning cost is increasing in t. Second, as the course material is relatively easy at the beginning of the semester, the learning cost accumulates slowly at first, followed by larger increments in the middle because of more challenging course content, and then slows down again towards the end of the semester after students have mastered the core course material. For tractability, we use the logistic function, the most frequently used sigmoid function. We have the following functional form for $t > 0$,: $c(t) = \frac{1}{1+e^{-(t-\frac{T}{r})}} - \frac{1}{1+e^{\frac{T}{r}}}$, where r is a constant and $r > 1$.

We model academic performance as a binary outcome, Bad or Good, realized at $t = T$, with the associated utilities B (normalized to 0) and G (normalized to 1). The probability of receiving good outcome is determined by the total effort exerted in the whole semester, $Prob(G) = \frac{1}{T}\int_0^T e_t dt$.

Notes-taking Activity: The instructor assigns a student to take notes at t^N. If a student does not exert effort in that lecture thus failing to complete this assignment, she suffers an instantaneous utility loss S .

A student's total discounted utility at $t = 0$, $V(t^s)$, is defined below,

$$V(t^s) = U(t^s) - \mathbb{1}_{\{t^s<t^N\}} e^{-\delta t^N} S$$
$$U(t^s) = -\int_0^{t^s} \frac{e^{-\delta t}}{a} c(t)dt + \frac{e^{-\delta T}}{T} t^s$$

Student's optimal stopping choice of t^s. A student chooses t^{s*} to maximize her total discounted utility, i.e., $t^{s*} = \arg\max_{t^s} V(t^s)$.

Instructor's optimal choice of t^N. We assume that the instructor only knows the distribution of students' abilities, the cdf of which is F. Furthermore, she would like to maximize the expected total effort (TE) exerted by students,

$$\max_{t^N} TE(t^N) = E[t^{s*}(t^N)|F(a)]$$

We solve the optimal problems by backward induction. First, given t^N, we show that there are three possible optimal stopping times for a student with ability a: $t^B(a)$, t^N or T.

Lemma 1. $t^{s*} \in \{t^B(a), t^N, T\}$,

where $t^B(a) \equiv \min\{t \in [0,T] | g(a) \equiv \frac{e^{-\delta T}}{T} - \frac{e^{-\delta t}}{a} c(t) = 0\}$[4]

Proof. See Appendix C.1. □

Lemma 1 implies that conditional on a student's ability, she only needs to compare $V(t^B(a)) = U(t^B(a)) - e^{-\delta t^N}$, $V(t^N) = U(t^N)$ and $V(T) = U(T)$ when deciding when to stop exerting effort. The next proposition characterizes the optimal stopping time at different ability levels.

Proposition 1.

$$t^{s*} \in \begin{cases} \{T\}, & a^h \leq a \\ \{t^B(a), t^N, T\}, & a^l \leq a < a^h \\ \{t^B(a), t^N\}, & a < a^l \end{cases}$$

where $a^h = \{a | \int_{t^B(a)}^{T} g(a)dt = 0\}$, $a^l = \{a | e^{-\delta t^{HLB}} S + \int_{t^B(a)}^{T} g(a)dt = 0\}$;

Proof. See Appendix C.2. □

Figure 1 provides simulation examples for the optimal stopping time by assuming $\delta = 0.015$, $T = 100$, $r = 2$, $S = 0.8$, $log(a) \in [1, 3]$. Each line represents a student's optimal stopping time with different t^N. Specifically, a student with very low ability would never keep exerting effort to T while those with very high ability would always learn until the end. For those who are in the middle, depending on t^N, they may or may not study until the end.

[4] Intuitively, $t^B(a)$ is the earliest lecture when the marginal benefit of exerting effort is equal to the marginal cost. It is possible that $g(a) > 0$ always holds, and $t^B(a) = \varnothing$. In this case, since learning is always beneficial, $t^{s*} = T$.

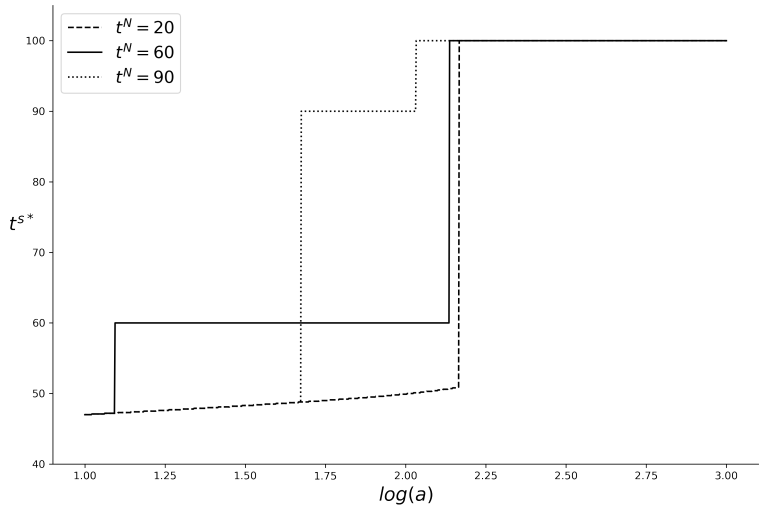


Figure 1: Examples for Optimal Stopping Time

Next, given the students' choices, we consider the timing that maximizes effort exerted. Let N^* be the set of solutions of this optimization problem, the following proposition summarizes the range of the instructor's optimal timing.

Proposition 2. *$\forall t^{N*} \in N^*, t^{N*} \in [\underline{t}^N, \bar{t}^N]$, where $\underline{t}^N = \min\{t^{HLB}(a^h), t^{NH}(\underline{a})\} > 0$ and $\bar{t}^N = t^{HLB}(a^l) < T$.*

Proof. See Appendix C.3. □

Proposition 2 implies that the optimal timing for the notes-taking task is never at the very beginning or the very end of the semester, i.e., $\{0, T\} \notin N^*$. Figure 2 provides a numerical example for the instructor's optimal timing of the assignment. With $\delta = 0.015$, $T = 100$, $r = 2$, $S = 0.8$, $\log(a) \sim N(1.5, 0.1)$, the optimal task timing that maximizes the total effort exerted by students is $t^N = 68$.

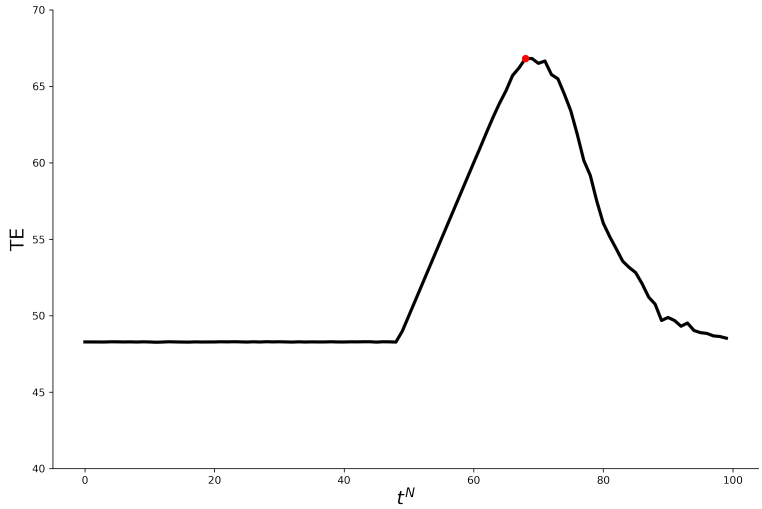


Figure 2: An Example for Instructor's Optimal Task Timing

4 Experimental Design

We are interested in examining whether the timing of participating in an in-class activity has impact on students' engagement and consequently on their learning performances.

Courses. The three courses we conduct our experiment are all compulsory courses for undergraduate students, including (1) Linear Algebra (2) Calculus and (3) Probability and Statistics. Table 1 summarizes the main features for each course such as the proportion of female students. It is worthwhile to note that the statistics course has two parallel sessions offered by the same instructor. Students randomly choose one based on their own course schedule.[5]

The notes-taking task. The in-class activity we chose is notes-taking. In both the education and psychology literature, studies have shown that assigning students to take notes is effective for their learning via both notes-

[5]We do not find significant difference in grades for these two sessions, and pool the data together in the result section.

taking production process and the notes review period (Bohay et al., 2011; Piolat et al., 2005). In particular, a good note summary should cover the course material well, make effective connections between concepts, and apply the gained knowledge to new contexts (Peper and Mayer, 1978). Therefore, it is considered as a cognitive demanding task and Piolat et al. (2005) equivalent its difficulty with chess. Second, taking notes is simple enough, so it would not discourage low-ability students to participate. For example, Leuven et al. (2010) finds that low-ability students achieve less when they are assigned to harder tasks.

Table 1: Course Description

	% of Female Students	# of Students	# of Students Per Lecture
Calculus	2%	57	2, 3
Linear Algebra	30%	170	6, 7
Probability & Statistics	53%	344	6, 7, 8

The notes-sharing activity. Among students who are assigned to a lecture and submit their notes, we randomly select one of them to share her notes and to leave a voice message to illustrate her notes in class WeChat group. Similar to Facebook group, Wechat group has 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, which plays the role of virtual office hour. Students can ask both instructors and their classmates questions, sharing course related material and chit chat there. We expect that students who were asked to share their notes should be more likely to respond to our experiment intervention because of the extra exposure to public. However, in later analyses, we do not find significant difference between the selected student and non-selected ones, and we decide to pool the data together.

4.1 Experimental procedure

We implemented our experiment in the fall semester of 2020 where universities in China have started regular offline teaching. In total, 571 students participated in our experiment, including 170 from linear algebra, 57 from calculus, and 344 from statistics. Since students can dropout in the middle of the class, 477 (84%) students remained until the end of the semester. Furthermore, the dropout rate does not significantly differ between students who are assigned to earlier and later weeks. Additionally, two of our undergraduate research assistants facilitated the experiment and students were told that they are teaching assistants for the course.

First, students are randomly assigned to each lecture from the third week to the end of the semester.[6] Table 1 reports the number of students per lecture. Specifically, we posted notes-taking schedule on both the e-learning platform and WeChat groups before the beginning of the third week. Second, inspired by the limited memory and inattention theory (Mullainathan, 2002; Karlan et al., 2016; Ericson, 2017), we sent a reminder email to students assigned to that particular lecture about the task one day before each lecture, and also specify the time for submitting the notes. For example, if a student is selected to take notes on March 3rd, she is expected to send the notes to our research assistant by the end of that day, though we clarified that the notes-taking task is not compulsory and is not related to the final grading. In total, 88% students in our experiment fulfilled the notes-taking task. Moreover, the likelihood of submitting the notes does not significantly differ across assignment timing.

Next, our research assistants randomly select a student who submitted the notes and encourage her to post her notes in WeChat groups and almost

[6]We exclude the first two weeks because during this time window, students may choose to dropout after experiencing the class.

all students participated in this activity.[7] After students share their notes, the research assistant sends an online survey link to all students in the Wechat group and asks them to rate on the quality of the notes from 0 to 100 points. To encourage them to rate, we randomly select one winner for each selected note and distribute small gifts such as USB drivers and calculators.[8] On average, each note receives 12 ratings and the rating frequency declines over weeks.

Additionally, both before and after the experiment, we distributed survey questionnaires to our subjects, including their learning motivation and attitude, big five personality traits, perceived temptation, time preference (Falk et al., 2018), and so on. Details of our survey could be found in Appendix A and B. Starting from September 25, the second week of the fall semester, we distribute the pre-experiment survey; Starting from December 30, right after the final exam, we distribute the post-experiment survey. The response rate for the pre-experiment survey is 70.2%, while it is 60.8% for the post-experiment survey.

5 Experimental Results

In this section, we present the experimental results. As shown in the model, we predict that students with low and high abilities respond to the intervention differently, i.e., high-ability students would always attend the lecture and take the notes, while low-ability ones would attend more if they are assigned to the middle part of the semester. One variable for approximating high (low) ability is whether a student postpones to take the course, e.g., a senior student takes the first-year calculus. Here we use "matched"

[7]Only two students refused to post their notes.

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

to indicate students who take the course as scheduled, and "unmatched" refer to students who postpone to take the course. Table 2 summarizes the final grades of unmatched and matched student's performance separately in these 3 years, and we see that unmatched students consistently perform worse than matched students. In 2020, we access students' registration data for the Probability & Statistics Course, where 104 out of 344 students are unmatched students. Among unmatched students, 87 of them retake this course. As predicted in the model, we should not expect high ability students to be affected by our treatment, and we indeed didn't find any treatment effect in the subsample of matched students, as shown in Table 3 (None of the coefficients are significant). Hence, in the following analyses we focus on the "unmatched" students. We first show the treatment effect on engagement, i.e., the attendance rate and average homework grades, then we examine how the assignment timing affects the attendance decay. Moreover, we investigate the treatment impact on students' exam grades and the spillover effect on other courses' grades. Finally, we further estimate the education return from good treatment timing.

	Final Grades			
	Year	2017	2018	2019
Unmatched	Mean	72.72	66.18	64.85
	Std. Dev.	18.45	16.93	18.15
	N	22	40	56
Matched	Mean	82.37	77.61	80.64
	Std. Dev.	14.01	15.47	15.79
	N	173	212	179

Table 2: Historical Final Grades, By Whether Matched

5.1 Treatment Effect on Engagement

First, we examine whether the timing for notes-taking assignment has any impact on students' engagement, measured by attendance and homework

Table 3: OLS Regression for Null Results in Matched Subsample

	Attendance		Midterm		Final	
	(1)	(2)	(3)	(4)	(5)	(6)
Task Timing	0.002	0.007	-0.103	-0.708	-0.191	-0.654
	(0.001)	(0.008)	(0.111)	(0.578)	(0.125)	(0.658)
Task Timing^2		-0.000		0.016		0.013
		(0.000)		(0.016)		(0.019)
Controls	No	No	No	No	No	No
Observations	380	380	380	380	380	380

Notes: HC3 Standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

grades. As our model predicted, students who are assigned to the middle of the semester would exert more effort, hence are expected to have higher attendance rate and homework grades. We divide students into 5 groups according to which lectures they are assigned to. On average, students assigned to the middle (Group 3) have the highest attendance rate (Figure 3) and homework grades (Figure 4), which is consistent with the model predictions. In particular, the difference between group 3 and 1/5 are statistically significant (Group 3(0.87) vs. Group 1(0.66), $p = 0.029$, two-sided t-test; Group3(0.87) vs. Group 5(0.58), $p = 0.005$, two-sided t-test).

Moreover, we supplement the above findings with a regression analysis in Table 4. We apply a simple OLS estimation to estimate the treatment effect on attendance rate (Columns 1 and 2) and average homework grades respectively (Columns 3 and 4). The independent variables are "Task Timing", a counting variable for our treatment, and its quadratic term "Task Timing", which is used to capture the inverse-U shape relationship between the treatment timing and the outcome variables. Additionally, in Column 2 and 4 we also control for students' gender, whether they are international

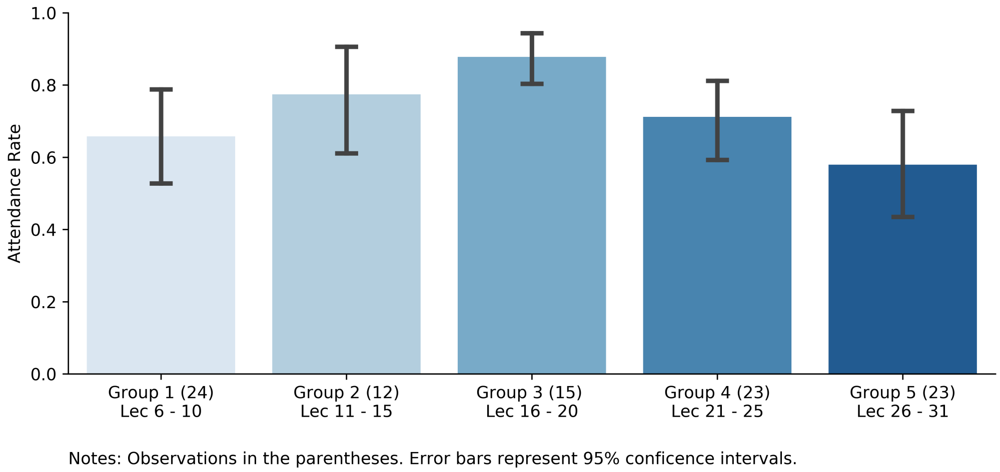


Figure 3: Average Attendance Rate for Unmatched Students, by Groups

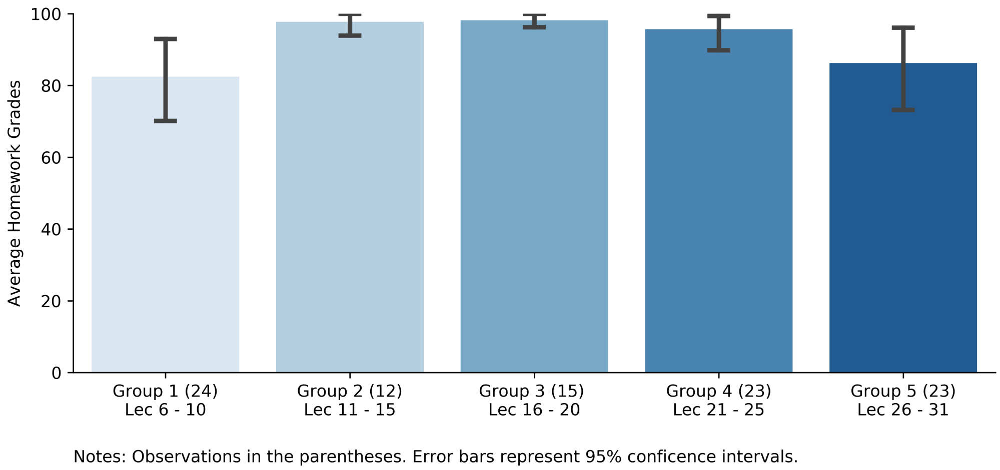


Figure 4: Average Mean Homework Grades for Unmatched Students, by Groups

students, whether they are STEM major, and the course dummies.

Table 4: OLS Regression for Treatment Effects on Engagement

	Attendance Rate		Average HW Grades	
	(1)	(2)	(3)	(4)
Task Timing	0.065***	0.061***	4.466**	4.245**
	(0.021)	(0.020)	(1.760)	(1.776)
Task Timing^2	-0.002***	-0.002***	-0.120**	-0.110**
	(0.001)	(0.001)	(0.047)	(0.047)
Controls	No	Yes	No	Yes
Observations	97	97	97	97

Notes: Control variables include gender, international student, STEM major and course dummies. HC3 Standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

As shown in both Columns 1 and 2, the coefficients of "Task Timing" are positive and significant at the 1% level ($p < 0.01$), while the coefficients of the quadratic term "Task Timing^2" are negative and significant at the 1% level ($p < 0.01$). Furthermore, we conduct a U-test(Lind and Mehlum, 2010) which shows a significant inverse-U shape relationship between the average attendance rate and the timing of notes-taking assignment ($p <$ 0.01). Additionally, a consistent inverse U-shape relationship is also found for the homework grades (Column 3 and 4). We summarize these findings below.

Result 1. *Assigning low ability students to the middle of the semester helps them attend more lectures and obtain higher homework grades.*

Consistent with our model prediction, Result 1 suggests that assigning students to the middle of the semester to take notes is the most effective for improving their overall attendance and homework grades. Since engagement decay is well observed in prior studies (DellaVigna and Malmendier, 2006; Allcott and Rogers, 2014), we are interested in understanding how the task assignment timing affects the decay of attendance. Figure 5 present the time-series data for the average attendance rate by groups. The dots are the actual attendance rate in each group across lectures and the solid lines are the linear fitted line. Overall, students who are assigned to the middle, i.e., Group 3, keep the high attendance rate and does not exhibit much decay over time, followed by group 2. Consistently, Groups 1 and 5 have the strongest attendance decay.

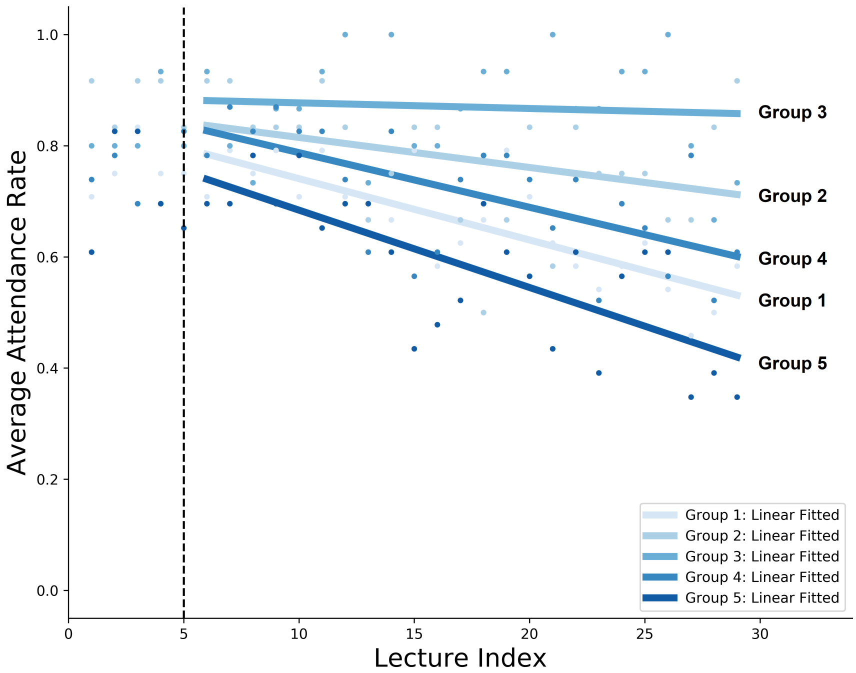


Figure 5: Attendance Rate for Each Lecture, for Unmatched Students, by Groups

Our regression results further confirm this graphical illustration. In Table 5, where the dependent variable is whether an individual attends a lecture or not, the coefficients of "Task Timing" are positive and significant at the 5% level ($p < 0.05$), while the "Task Timing^2" is negative and significant at the 5% level ($p < 0.05$), indicating that students who are assigned to middle are more likely to attend. Furthermore, the interaction term "Task Timing × Lecture Index" is positive and significant at the 1% level ($p < 0.01$), while the interaction term "Task Timing^2 × Lecture Index" are negative and significant at the 1% level ($p < 0.01$). That is to say, students who are assigned to the middle have significantly less decay than other students. We summarize our findings in Result 2.

Table 5: OLS Regression for Treatment Effects on Attendance Dynamics

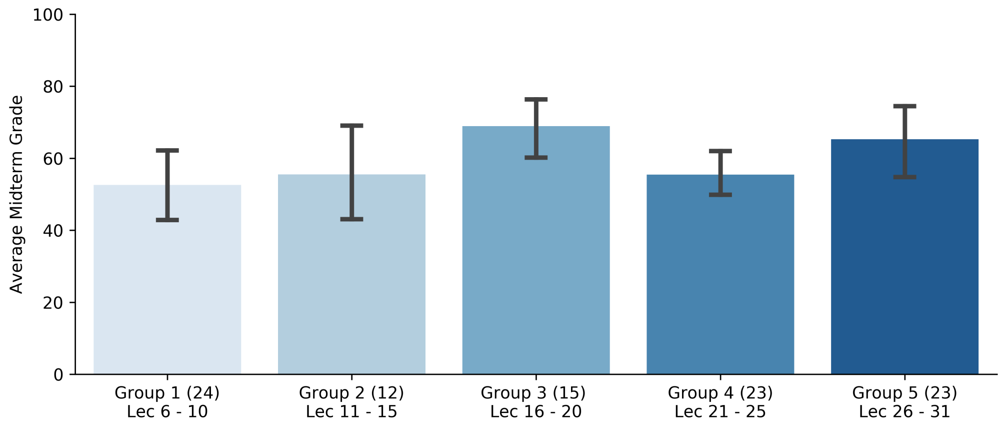
	Whether Attend (0/1)	
	(1)	(2)
Lecture Index	-0.021***	-0.021***
	(0.005)	(0.005)
Task Timing	0.027**	0.024**
	(0.012)	(0.012)
Task Timing^2	-0.001**	-0.001**
	(0.000)	(0.000)
Lecture Index × Task Timing	0.002***	0.002***
	(0.001)	(0.001)
Lecture Index × Task Timing^2	-0.000***	-0.000***
	(0.000)	(0.000)
Controls	No	Yes
Observations	2,815	2,815

Notes: Control variables include gender, international student, STEM major and course dummies. HC3 Standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Result 2. *For low ability students, assigning them to the middle of the semester significantly alleviates their attendance decay problem.*

5.2 Treatment Effect on Exam Performance

Next, we look into the exam performances of students, to examine whether the high engagement of students who are assigned to the middle turns into better exam grades.[9] Figures 6 and 7 present the average midterm and final exam grades by groups. Compared to students who are assigned to the early part of the semester, those who are assigned to the middle (Group 3) do perform better than others. Surprisingly, we also notice that students who are assigned to the late part of the semester (Group 5) also have equally good exam grades as Group 3.



Notes: Observations in the parentheses. Error bars represent 95% conficence intervals.

Figure 6: Average Midterm Grades for Unmatched Students, by Groups

The pattern described above is captured by the regression in Tables 6 and 7 as well. In Columns 1 and 2 of Table 6, the coefficient of "Task Timing" is positive and significant ($p < 0.1$), showing that the timing of notes-taking assignment has a negative impact on midterm grades. In Columns 3 and 4 of Table 6, we consider the quadratic specification. The coefficient

[9]Prior studies have shown that students' attendance rate is significantly positively correlated with their exam grades Dobkin et al. (2010); Arulampalam et al. (2012), especially for high-ability studentsArulampalam et al. (2012), which is also observed in our data - We find that for those matched students, their attendance rate and final exam grade is significantly correlated ($p = 0.002$), while this correlation is not significant for the unmatched students ($p = 0.151$).

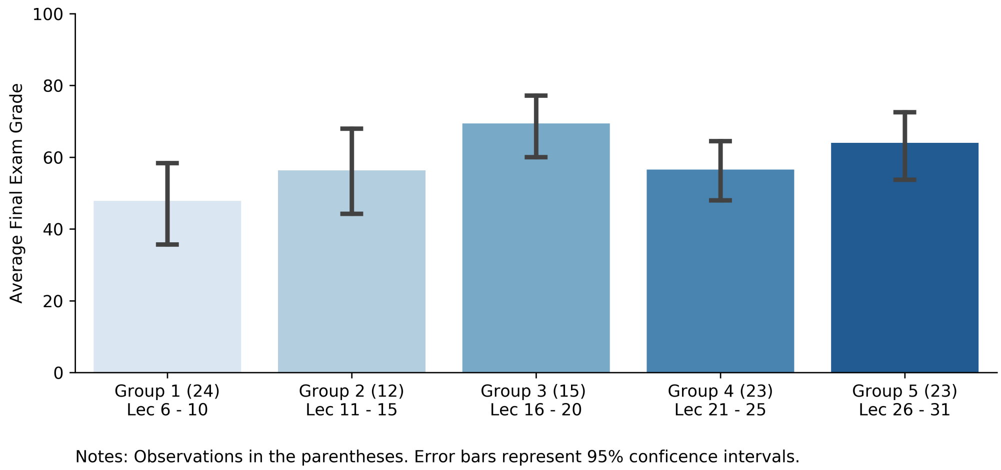


Figure 7: Average Final Exam Grades for Unmatched Students, by Groups

of "Task Timing" is positive but not significant ($p > 0.1$), while the coefficients of "Task Timing^2" are negative and insignificant ($p > 0.1$). Similarly, we find a significant overall effect on the final exam grades, though the non-linear relationship, reflected by the quadratic term is not significant (Table 7). Altogether, these results suggest a more linear impact on the grades. We conjecture that the exam grade is a combination of different characteristics, and the learning engagement could only be the partial motive, hence overall the exam performance may not be perfectly in line with the engagement behavior. We summarize our findings in Result 3.

Result 3. *For low ability students, assigning them to the middle and late part of the semester improves their exam grades.*

Finally, can our notes-taking task help students achieve better final grades? We address this question by comparing the final grades of 2020 semester to previous semesters, where no notes-taking assignment was implemented. Figure 8 presents the dynamic pattern of the final grade gap between unmatched and matched students. Using the final grade gap in 2017 - 2019 semesters, we predict a counterfactual final grade gap in 2020 semester if there is no assignment. We find that the realized grade gap is

Table 6: OLS Regression for Treatment Effects on Midterm Grades

	Midterm Exam Grade			
	(1)	(2)	(3)	(4)
Task Timing	0.478*	0.713**	1.173	1.512
	(0.282)	(0.298)	(1.588)	(1.698)
Task Timing^2			-0.019	-0.022
			(0.044)	(0.045)
Controls	No	Yes	No	Yes
Observations	97	97	97	97

Notes: Control variables include gender, international student, STEM major and course dummies. HC3 Standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7: OLS Regression for Treatment Effects on Final Exam Grades

	Final Exam Grade			
	(1)	(2)	(3)	(4)
Assigned to Which Lecture	0.668**	0.796**	2.510	2.679
	(0.297)	(0.330)	(1.666)	(1.809)
Assigned to Which Lecture^2			-0.051	-0.053
			(0.046)	(0.046)
Controls	No	Yes	No	Yes
Observations	97	97	97	97

Notes: Control variables include gender, international student, STEM major and course dummies. HC3 Standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

smaller than the counterfactual value. Additionally, we plot the final grade gap by groups, and the final grade gaps for Group 2 - 5 are all smaller than the counterfactual value.

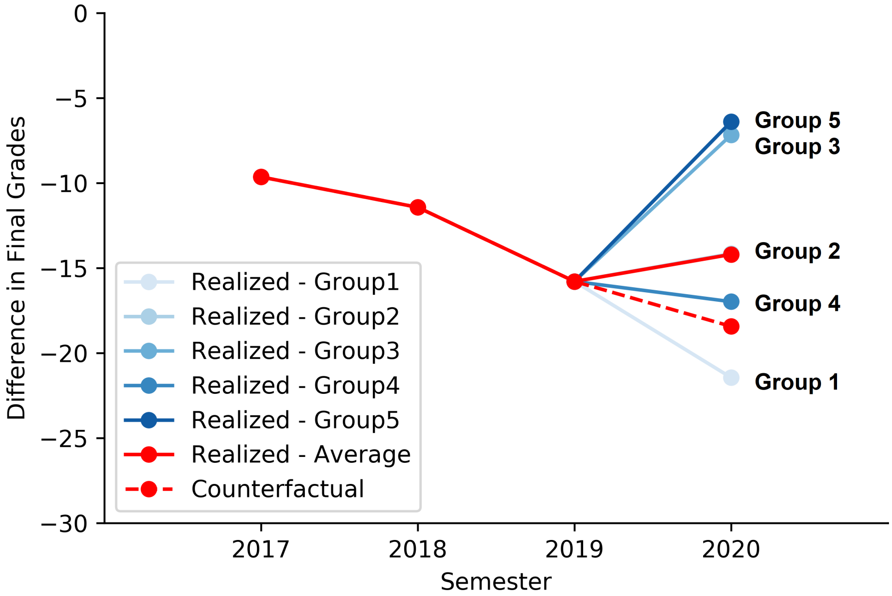


Figure 8: Difference on Final Grades Between Unmatched and Matched Students, 2017 - 2020

Result 4. *For low ability students, the notes-taking task has positive impacts on improving their final grades.*

5.3 Spillover Effect and Education Return

Our results show that students who are assigned to the middle of the semester achieve significantly better academic performance. Based on these results, we are also interested in the spillover effect of our intervention - Do students assigned to the middle also have higher grades in other courses? To answer this question, we examine the treatment effect on unmatched students' GPA in other courses and report results in Table 8.

In column 1 - 2, the dependent variable is a student's credit-weighted average grades (GPA) of all courses (excluding the course used in our experiment) in 2020 fall semester; In column 3 - 4, the dependent variable is a student's credit weighted average grades of all courses in 2021 spring semester. In Columns 1 and 2, we notice a marginally significant spillover effect - for example, in column 2, the coefficient of "Task Timing" is 1.070 ($p < 0.1$) and the coefficient of "Task Timing ^{2}" is -0.026 ($p < 0.1$), indicating that students who are assigned to the middle of the semester also achieve better academic performance in other courses of the 2020 fall semester. However, there is no significant result in Column 3 and 4, suggesting that the spillover effect does not carry to the subsequent semester.

Table 8: OLS Regression for Spillover Effect

	Other Courses in 2020 Fall		Other Courses in 2021 Spring	
	(1)	(2)	(3)	(4)
Task Timing	1.268*	1.070*	-0.116	-0.381
	(0.657)	(0.601)	(0.828)	(0.809)
Task Timing^2	-0.032*	-0.026*	0.005	0.014
	(0.017)	(0.015)	(0.021)	(0.020)
Pre-experiment GPA	0.387	0.351	0.328	0.260
	(0.251)	(0.228)	(0.263)	(0.240)
Total Credits	0.206	0.175	0.223	0.190
	(0.190)	(0.187)	(0.172)	(0.177)
Controls	No	Yes	No	Yes
Observations	93	93	91	91

Notes: Control variables include gender, international student, STEM major and course dummies. HC3 Standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Result 5. *For low ability students, assigning them to the middle of the semester also improve their performances in other courses that they took during the same semester.*

Moreover, we estimate how our intervention here affects students' monthly wage after graduation. Utilizing the Chinese College Students Survey (CCSS) conducted by the China Data Center of Tsinghua University (Li et al., 2012a,b).[10] We find that on average 1 point increase in the GPA yields a 11.544 RMB increase in the monthly wage after graduation.

We conduct a simple back-of-the-envelope calculation to demonstrate the economic return of our intervention. Combining the effect of our intervention on the targeting course as well as the spillover effect in other courses (Column 2 of Table 8), the monthly wage for those who assigned to the middle is higher than those assigned to early or late. For example, on average, students that are assigned to lecture 20 are 5.60 points higher in GPA than students who are assigned to lecture 6. For a student in a 4-year undergraduate program, this effect roughly increases their overall GPA by 0.70,[11] consequently leading to a 8.08 RMB increase in the monthly wage.

5.4 Robustness Check

One criticism of our subsample analyses is that it may suffer from multiple hypotheses testing problem. One alternative "Let data speak" approach for detecting treatment heterogeneity is to apply machine learning method such as Causal forests developed by Athey and Imbens (2016) and Wager and Athey (2018). Specifically, as the probability of a Type I error increases with the number of tests conducted, a large number of tests may lead to spurious heterogeneity estimations. Since causal forests requires no distributional assumptions, it allows greater flexibility in estimating heterogeneous treatment effects and allows us to focus our examination of the

[10]The survey sampled undergraduate students who finished their bachelor degree from 2010 to 2014 and we focus on the subsample from Chinese elite universities, i.e., universities that are listed in the 985 & 211 project. In total, we had 3581 students.

[11]Here we simply assume that the GPA does not change in other semesters and therefore, the overall increase for GPA is 5.60/8=0.70.

data for responsive subgroups (e.g. Davis and Heller (2020), Hermosilla (2018), Knittel and Stolper (2019)). Moreover, such method enables us to construct a more comparable counterfactual for each treated unit.[12] Then we can use predicted individual treatment effects to identify the most responsive quartile of students.

Specifically, we first use predicted individual treatment effects to divide our full sample into two groups, i.e. the largest quartile of predictions and the rest of the sample, and then compare the conditional average treatment effects (CATEs) of the two groups.[13] We find that assigning to later lectures improves final exam grades by 36.97 for subjects in the largest quartile (Table 9), while it presents negative and significant effect on the rest of the sample ($p < 0.01$, two-sided t-test). More importantly, the difference between these two samples is significant, suggesting the existence of significant treatment heterogeneity. Additionally, we also compare the individual characteristics between these two subsamples and the only significant difference is the proportion of matched students (Table 10, 0.733 vs. 0.818, p=0.047, two-sided t-test).

For the ability measure, prior studies such as Cadena and Keys (2015) show that impatient students are more likely to dropout, earn less, and feel more regret afterwards. We use individuals' patience level collected in the pre-experiment survey as another proxy for students' abilities. We construct patience parameter following Falk et al. (2018). Figure 9 presents the distribution of patience level among our students. Compared to Falk et al. (2018) which covers a large variety of subjects across countries (Figure

[12]Conventional approaches such as propensity score matching have poor statistical performance in the presence of irrelevant or many covariates Wager and Athey (2018). By contrast, the causal forests method uses a data-driven approach to adaptively determine weights for each nearby observation and then uses those weights to generate outcome variable predictions. Therefore, it provides an advantage in terms of both an increase in statistical power and a reduction in estimation bias.

[13]We only present the results for final exam grades here, and the analyses for midterm and final grades are similar.

Table 9: Causal Forest for Predicting Treatment Heterogeneity

	Final Exam Grade		Final Grade	
	(1)	(2)	(3)	(4)
ATE	1.24		1.25	
	(1.74)		(1.4)	
Most Responsive Quartile		36.97***		30.23***
		(5.19)		(4.03)
Rest of Sample		-10.2***		-8.27***
		(1.7)		(1.31)
p-value of Difference		0.000		0.000
Observations	477	477	477	477

Notes: Control variables include students' gender, course, international student, major, and matched seniority dummies. The largest quartile indicates the quartile that has the largest predicted responses to the intervention. Rest of Sample indicates subjects other than those in the largest quartile. Robust standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 10: Characteristic Difference in the Largest Quartile vs. Rest of Sample

Proportion of	Largest Quartile	Rest of Sample	p-value of Difference
Matched Seniority	0.733	0.818	0.047**
Female	0.4	0.392	0.879
International	0.1	0.081	0.529
Calculus	0.133	0.112	0.531
Linear Algebra	0.283	0.361	0.12
Probability & Statistics	0.583	0.527	0.282
Science	0.175	0.151	0.538
Engineering	0.2	0.213	0.765
Social Science & Literature	0.625	0.636	0.831

Notes: The largest quartile indicates the one that have the largest predicted responses to the treatment(assigned to the second half of the experiment). Rest of Sample indicates subjects other than those in the largest quartile. significance at the 1%, 5%, and 10% levels, respectively.

A.8.1 in their paper), our under graduate students sample exhibit much higher patience level.

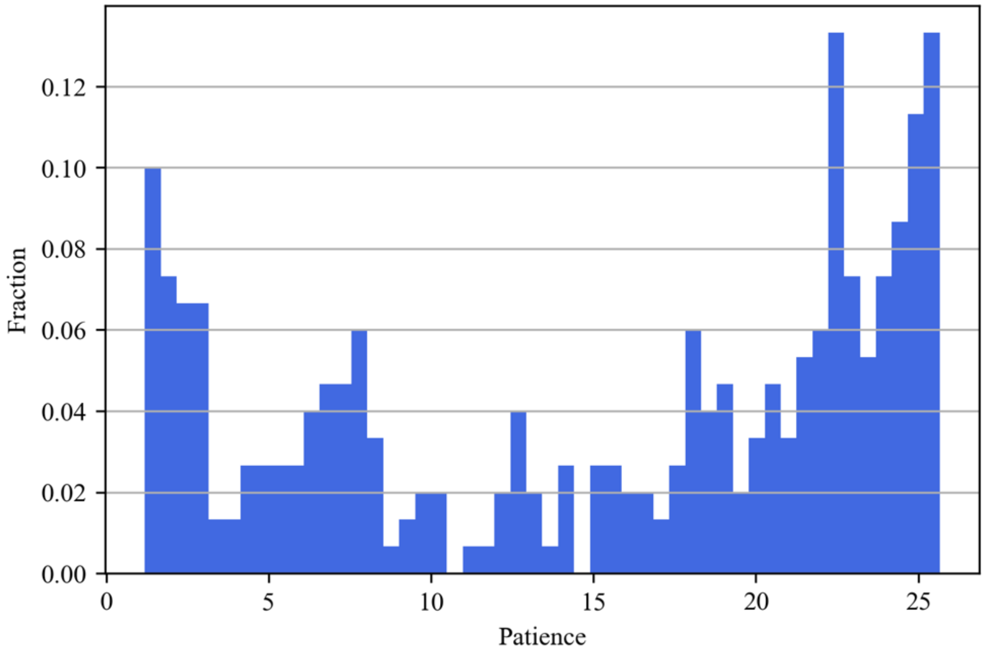


Figure 9: Histogram of Patience Level

Table 11 presents the regression analyses for grades. Now instead of using "matched" to capture ability, we use patience measure in the pre-experiment survey instead and are interested in examining its interaction effect with our treatment variable "Assigning to Which Lecture". Consistent with Result 1, we also find an inverse-U shape relationship between the engagement (measured by attendance rate and average homework grades) and "Task Timing".

6 Conclusion

Improving the academic behavior and performance has long been a central problem in education. We conducted a field experiment at a Chinese elite university. Our study tests whether the timing of interventions matters

Table 11: OLS Regression: Treatment Effects on Attendance Rate and Average Homework Grades in Impatient Subsample

	(1)	(2)	(3)	(4)
	Attendance Rate		Average HW Grades	
Task Timing	0.039***	0.034***	1.290*	1.256*
	(0.014)	(0.013)	(0.674)	(0.672)
Task Timing^2	-0.001***	-0.001***	-0.033*	-0.032
	(0.000)	(0.000)	(0.019)	(0.019)
Controls	No	Yes	No	Yes
Observations	166	166	166	166

Notes: Control variables include gender, whether international student, whether STEM major and course dummies. HC3 Standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

in improving the academic performance of students via assigning students to different lectures to finish notes-taking tasks. The experimental results show that assigning students to the middle of the term help those low ability students achieve higher academic performance (in the sense of higher attendance rate, higher homework grades, and higher final grades. We also document a spillover effect, indicating that those students who are assigned to the middle of the semester also achieve higher grades in other courses. Furthermore, we estimate the long-term welfare effect of the treatment. For example, compared to students who are assigned to lecture 6, those who are assigned to lecture 20 will have 0.7 higher GPA, consequently leading to a 8.08 RMB increase in the monthly wage. Additionally, we propose a simple model to show how the timing of the task affects students' learning behavior, and show that the optimal timing should be assigning students to do the task in the middle of the term.

References

Allcott, Hunt 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.

_ , _ , Erik Bloom, Elizabeth King, and Michael Kremer, "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment," *American Economic Review*, 2002, *92* (5), 1535–1558.

Ariely, Dan and Klaus Wertenbroch, "Procrastination, Deadlines, and Performance: Self-control by Precommitment," *Psychological Science*, 2002, *13* (3), 219–224.

_ , Uri Gneezy, George Loewenstein, and Nina Mazar, "Large Stakes and Big Mistakes," *The Review of Economic Studies*, 2009, *76* (2), 451–469.

Arulampalam, Wiji, Robin A Naylor, and Jeremy Smith, "Am I Missing Something? The Effects of Absence from Class on Student Performance," *Economics of Education Review*, 2012, *31* (4), 363–375.

Athey, Susan and Guido Imbens, "Recursive Partitioning for Heterogeneous Causal Effects," *Proceedings of the National Academy of Sciences*, 2016, *113* (27), 7353–7360.

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.

Banerjee, Abhijit V and Esther Duflo, "(Dis) Organization and Success in an Economics MOOC," *American Economic Review: Papers and Proceedings*, 2014, *104* (5), 514–518.

Behrman, Jere R, Susan W Parker, Petra E Todd, and Kenneth I Wolpin, "Aligning Learning Incentives of Students and Teachers: Results From a Social Experiment in Mexican High Schools," *Journal of Political Economy*, 2015, *123* (2), 325–364.

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.

Bisin, Alberto and Kyle Hyndman, "Present-bias, Procrastination and Deadlines in a Field Experiment," *Games and Economic Behavior*, 2020, *119*, 339–357.

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.

Braun, Henry, Irwin Kirsch, and Kentaro Yamamoto, "An Experimental Study of the Effects of Monetary Incentives on Performance on the 12th-Grade NAEP Reading Assessment.," *Teachers College Record*, 2011, *113* (11), 2309–2344.

Burger, Nicholas, Gary Charness, and John Lynham, "Field and Online Experiments on Self-control," *Journal of Economic Behavior & Organization*, 2011, *77* (3), 393–404.

Cadena, Brian C and Benjamin J Keys, "Human Capital and the Lifetime Costs of Impatience," *American Economic Journal: Economic Policy*, 2015, *7* (3), 126–53.

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

Charness, Gary and Uri Gneezy, "Incentives to Exercise," *Econometrica*, 2009, *77* (3), 909–931.

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.

Davis, Jonathan MV and Sara B Heller, "Rethinking the Benefits of Youth Employment Programs: The Heterfogeneous Effects of Summer Jobs," *Review of Economics and Statistics*, 2020, *102* (4), 664–677.

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

Dobkin, Carlos, Ricard Gil, and Justin Marion, "Skipping Class in College and Exam Performance: Evidence from a Regression Discontinuity Classroom Experiment," *Economics of Education Review*, 2010, *29* (4), 566–575.

Dobronyi, Christopher R, Philip Oreopoulos, and Uros Petronijevic, "Goal Setting, Academic Reminders, and College Success: A Large-scale Field Experiment," *Journal of Research on Educational Effectiveness*, 2019, *12* (1), 38–66.

Ericson, Keith Marzilli, "On the Interaction of Memory and Procrastination: Implications for Reminders, Deadlines, and Empirical Estimation," *Journal of the European Economic Association*, 2017, *15* (3), 692–719.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global Evidence on Economic

Preferences," *The Quarterly Journal of Economics*, 2018, *133* (4), 1645–1692.

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.

Fryer, RG, Steven D Levitt, John List, and Sally Sadoff, "Enhancing the Efficacy of Teacher Incentives Through Framing: A Field Experiment," 2018.

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.

Gertler, Paul, Sean Higgins, Aisling Scott, and Enrique Seira, "The Long-term Effects of Temporary Incentives to Save: Evidence from a Prize-Linked Savings Field Experiment," *Working Paper*, 2018.

Gibbs, Mark N and David JC MacKay, "Variational Gaussian process classifiers," *IEEE Transactions on Neural Networks*, 2000, *11* (6), 1458–1464.

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.

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, Tracy Xiao Liu Jie and Jie Tang, "How Monetary Incentives Improve Outcomes in MOOCs: Evidence from a Field Experiment," 2021. Working Paper.

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.

Hébert, Benjamin and Michael Woodford, "Neighborhood-based Information Costs," *American Economic Review*, 2021, *111* (10), 3225–55.

Hermosilla, Manuel, "Search-and-Match in a Rush: Investigating Reactive Licensing in the Pharmaceutical Industry," *Johns Hopkins Carey Business School Research Paper*, 2018.

Himmler, Oliver, Robert Jäckle, and Philipp Weinschenk, "Soft Commitments, Reminders, and Academic Performance," *American Economic Journal: Applied Economics*, 2019, *11* (2), 114–42.

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.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman, "Getting to the Top of Mind: How Reminders Increase Saving," *Management Science*, 2016, *62* (12), 3393–3411.

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.

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.

Knittel, Christopher R and Samuel Stolper, "Using Machine Learning to Target Treatment: The Case of Household Energy Use," Technical Report, National Bureau of Economic Research 2019.

Kremer, Michael, Edward Miguel, and Rebecca Thornton, "Incentives to Learn," *The Review of Economics and Statistics*, 2009, *91* (3), 437–456.

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.

Leibowitz, Nathaniel, Barak Baum, Giora Enden, and Amir Karniel, "The Exponential Learning Equation as a Function of Successful Trials Results in Sigmoid Performance," *Journal of Mathematical Psychology*, 2010, *54* (3), 338–340.

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, and Sally Sadoff, "The Effect of Performance-Based Incentives on Educational Achievement: Evidence From a Randomized Experiment," Technical Report, National Bureau of Economic Research 2016.

_ , _ , 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, Hongbin, Lingsheng Meng, Xinzheng Shi, and Binzhen Wu, "Does Attending Elite Colleges Pay in China?," *Journal of Comparative Economics*, 2012, *40* (1), 78–88.

_ , _ , _ , and _ , "Does Having a Cadre Parent Pay? Evidence from the first job offers of Chinese college graduates," *Journal of Development Economics*, 2012, *99* (2), 513–520.

Lind, Jo Thori and Halvor Mehlum, "With or without U? The Appropriate Test for a U-shaped Relationship," *Oxford Bulletin of Economics and Statistics*, 2010, *72* (1), 109–118.

Meier, Stephan, "Do Subsidies Increase Charitable Giving in the Long Run? Matching Donations in a Field Experiment," *Journal of the European Economic Association*, 2007, *5* (6), 1203–1222.

Morisano, Dominique, Jacob B Hirsh, Jordan B Peterson, Robert O Pihl, and Bruce M Shore, "Setting, Elaborating, and Reflecting on Personal Goals Improves Academic Performance," *Journal of Applied Psychology*, 2010, *95* (2), 255.

Mullainathan, Sendhil, "A Memory-based Model of Bounded Rationality," *The Quarterly Journal of Economics*, 2002, *117* (3), 735–774.

Newell, Allen and P Rosenbloom, "Mechanisms of Skill Acquisition," *Cognitive Skills and Their Acquisition*, 1981.

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

Patterson, Richard W, "Can Behavioral Tools Improve Online Student Outcomes? Experimental Evidence From a Massive Open Online Course," *Journal of Economic Behavior & Organization*, 2018, *153*, 293–321.

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

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

Rezaei, Sarah, Hendrik Sonnabend, and Bastian Westbrock, "Using Goals to Motivate Students: The Power of Small Financial Rewards," *Working Paper*, 2021.

Wager, Stefan and Susan Athey, "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests," *Journal of the American Statistical Association*, 2018, *113* (523), 1228–1242.

Online Appendices

A Pre-Experiment Survey

1. (7 Likert Scale) How much do you like this class?
2. Are you planning to take more courses in this major?
3. What is the main reason for taking this course?
   (a) Pass the exam.
   (b) This course is useful for future courses that I am going to take.
   (c) The material or methods covered in this course is useful for my future career.
4. (7 Likert Scale) How much do you agree with the following statements?
   (a) I set short-term (daily or weekly) goals as well as long-term goals (monthly or for the semester).
   (b) I keep a high standard for my learning in my online courses.
   (c) I summarize my learning in online courses to examine my understanding of what I have learned.
   (d) I communicate with my classmates to find out how I am doing in my online classes.
   (e) I make sure that I keep up with the weekly readings and assignments for this course.
   (f) I often feel so lazy or bored when I study for this class that I quit before I finish what I planned to do.

(g) I work hard to do well in this class even if I don't like what we are doing.

5. (7 Likert Scale) How much do you describe yourself as

   (a) Extraverted, enthusiastic
   (b) Critical, quarrelsome
   (c) Dependable, self-disciplined
   (d) Anxious, easily upset
   (e) Open to new experiences, complex
   (f) Reserved, quiet
   (g) Sympathetic, warm
   (h) Disorganized, careless
   (i) Calm, emotionally stable
   (j) Conventional, uncreative

6. (7 Likert Scale) How much do you agree the following statements?

   (a) Your intelligence is consistent and cannot be changed in anyways.
   (b) Intelligence is a feature of you that is hard to change dramatically.
   (c) You can learn new things, but you can't really change your basic intelligence

7. The following are possible reasons for studying hard in this course. Please rate how much you agree with the following statement (7 Likert Scale).

   (a) to learn more knowledge

(b) I am afraid of getting low grades

(c) I am afraid of bad performance in the class

(d) It is fun to learn

(e) It is challenging to fully master the knowledge

(f) I want people to think that I am smart

8. (7 Likert Scale) Are you content with your current life?

9. If you want to get help for learning this course from an upperclass student, how many of them would you think of?

10. Do you have a study group for this class? If yes, how many students in this group?

11. If you want to get help for learning this course from your peer, how many of them would you think of?

12. (7 Likert Scale) How much do you agree with the following statements?

    (a) How much are you willing to recommend this course to other students?

    (b) How much do you think this course is helpful for finding a good job?

    (c) How much do you think this course is helpful for your future career?

13. Do you like your current major?

14. Now you need to choose between getting paid today and getting paid 12 months later. Now there are five scenarios and in each case, the amount of the money for today is the same, i.e., RMB 100, but the

amount for 12 month later is different (See Figure 5 in Falk et al. (2018) for detailed parameters).

15. Image now we invite you to participate in a course video watching study. If you watch 100 minutes for two days, you can get a reward of 50 RMB. These two days are the 7th day from today, i.e., one week later, and the 14th day from today, i.e., two weeks later. You can allocate these 100 minutes between these two days. If you watch x minutes in the 7th day, the remaining time for the 14th day is (100-x)/r. The value of r is different in the following four cases. For example, if the conversion ratio is $1 : 1.25$ with r equals to 1.25, then for every 1 more minute you watch on the 14th day, you can watch 1.25 minutes less on the 7th day; if the conversion ratio is $1 : 0.9$, then one more minute on the 14th day implies 0.9 minutes less on the 7th day.

    (a) $1 : 0.9$

    (b) $1 : 1$

    (c) $1 : 1.1$

    (d) $1 : 1.25$

16. (7 Likert Scale) How much do you describe yourself as

    (a) To what extent will you sacrifice immediate interests for long-term interests?

    (b) Knowing it's better to do it right now, you tend to put off tasks.

17. (7 Likert Scale) How much do you describe yourself as

    (a) I am good at resisting temptation

    (b) I have a hard time breaking bad habits

(c) I am lazy

(d) I say inappropriate things

(e) I do certain things that are bad for me, if they are fun

(f) I refuse things that are bad for me

(g) I wish I had more self-discipline

(h) People would say that I have iron self-discipline

(i) Pleasure and fun sometimes keep me from getting work done

(j) I have trouble concentrating

(k) I am able to work effectively toward long-term goals

(l) Sometimes I can't stop myself from doing something, even if I know it is wrong

(m) I often act without thinking through all the alternatives

18. (7 Likert Scale) How much do you agree with the following statements?

    (a) New ideas and projects sometimes distract me from previous ones.

    (b) Setbacks do not discourage me.

    (c) I have been obsessed with a certain idea or project for a short time but later lost interest.

    (d) I am a hard worker.

    (e) I often set a goal but later choose to pursue a different one.

    (f) I have difficulty maintaining my focus on projects that take more than a few months to complete.

    (g) I finish whatever I begin.

(h) I am diligent.

19. Now you need to choose between getting paid in a week and getting paid in 12 months plus a week. Now there are five scenarios and in each case, the amount of the money for one week after is the same, i.e., RMB 100, but the amount for 12 month plus a week is different (See Figure 5 in Falk et al. (2018) for detailed parameters).

20. How many hours did you plan to spend on this course?

21. Until now, how many hours have you spent on this course?

22. Image now we invite you to participate in a course video watching study. If you watch 100 minutes for two days, you can get a reward of 50 RMB. These two days are today and the 7th day from today, i.e., one week later. You can allocate these 100 minutes between these two days. If you watch x minutes today, the remaining time for the 7th day is (100-x)/r. The value of r is different in the following four cases. For example, if the conversion ratio is 1 : 1.25 with r equals to 1.25, then for every 1 more minute you watch on the 7th day, you can watch 1.25 minutes less today; if the conversion ratio is 1 : 0.9, then one more minute on the 7th day implies 0.9 minutes less today.

    (a) 1 : 0.9

    (b) 1 : 1

    (c) 1 : 1.1

    (d) 1 : 1.25

23. Which province did you take your college entry exam?

24. What is your Hukou status?

25. Were you college entry exam waived? For example, a student who wins IMO gold medal is enrolled without taking the colleage entry exam.

26. What is your total score in the college entry exam?

27. What is your math score in the college entry exam?

28. Were you in STEM track when you took college entry exam?

29. What is the name of your high school?

B Post-Experiment Survey

1. What is the main reason for taking this course?

   (a) Pass the exam.

   (b) This course is useful for future courses that I am going to take.

   (c) The material or methods covered in this course is useful for my future career.

2. Do you want to be a student in your current major?

3. How many upper-class students do you know from whom you actually asked for help for this class?

4. How many students in this class do you know from whom you asked for help?

5. Now you need to choose between getting paid today and getting paid 1 months later. Now there are six cases and in each case, the amount for today is the same, which is 200 RMB. But the amount for 1 month later is chosen from {180, 200, 220, 250, 280, 300} RMB.

6. Now you need to choose between getting paid 1 month later and getting paid 2 months later. Now there are six cases and in each case, the amount for 1 month later is the same (200RMB), but the amount for 2 month later is chosen from {180, 200, 220, 250, 280, 300} RMB. We are curious about your choice in each case.

7. Suppose that you win 10 restaurant certificates, each of which can be used (once) to receive a "dream restaurant night" at any restaurant you want. On each such night, you and a companion will get the best table and an unlimited budget for food and drink. There will be no cost to you: all payments including tips are part of the prize.

The certificates can be used immediately, starting tonight, and it is guaranteed that every restaurant will honor them if they are used within two years. However, any certificates that are not used up within this two year period become valueless.

(a) Think about what would be the ideal allocation of these certificates for the first and the second year. From your current perspective, how many of the ten certificates would you ideally like to use in year 1 as opposed to year 2? Choose number 1 - 10.

(b) Some people might be tempted to depart from this ideal allocation. For example, there might be temptation to use up the certificates sooner, and not keep enough for the second year. Or you might be tempted to keep too many for the second year. If you just gave in to your temptation, how many would you use in the first year? Choose number 1 - 10.

(c) Think about both the ideal and temptation. Based on your most accurate forecast of how you would actually behave, how many of the nights would you end up using in year 1 as opposed to year 2? Choose number 1 - 10.

8. (7 Likert Scale) Do you often delay things that should be done in a timely manner?

9. Which of the following emotions did you experience when taking this course? (multiple choices).

   (a) Anger

   (b) Anxiety

   (c) Confusion

(d) Contentment

(e) Fatigue

(f) Happiness

(g) Irritation

(h) Refuse

(i) I do not have strong emotions towards the course

10. (7 Likert scale) How much do you enjoy the class?

11. Which part of the class do you enjoy most?

(a) Lecture

(b) Doing Assignment

(c) Discussion with other students

(d) Discussion with the instructor

(e) Other

12. Which of the following emotions did you experience before finishing this notes-taking task? (multiple choice)?

(a) Anger

(b) Anxiety

(c) Confusion

(d) Contentment

(e) Fatigue

(f) Happiness

(g) Irritation

(h) Refuse

(i) I do not have strong emotions

13. Which of the following emotions did you experience before sharing your notes and leaving a voice message? (multiple choice)

    (a) Anger
    (b) Anxiety
    (c) Confusion
    (d) Contentment
    (e) Fatigue
    (f) Happiness
    (g) Irritation
    (h) Refuse to do so
    (i) Refuse
    (j) I do not have strong emotions

14. Which of the following emotions did you experience after finishing this notes-taking task? (multiple choice)?

    (a) Anger
    (b) Anxiety
    (c) Confusion
    (d) Contentment
    (e) Fatigue
    (f) Happiness
    (g) Irritation
    (h) Refuse

(i) I do not have strong emotions

15. Which of the following emotions did you experience after sharing your notes and leaving a voice message? (multiple choice)

    (a) Anger
    (b) Anxiety
    (c) Confusion
    (d) Contentment
    (e) Fatigue
    (f) Happiness
    (g) Irritation
    (h) Refuse to do so
    (i) Refuse
    (j) I do not have strong emotions

16. (7 Likert scale) Do you agree that writing notes help you spend more time for the class?

17. Before the lecture that you were asked to take notes, did you take notes for this class, or for other class?

18. After finishing the notes-taking task, did you start taking notes for this class, or for other class?

19. In the first half of the semester, on average, how many hours per week did you spend for this class?

20. In the second half of the semester, on average, how many hours per week did you spend for this class?

21. If you did not take notes for the lecture you were assigned to, what are the reason?

22. What is your prediction for your final grades?

23. Now you need to choose between getting paid today and getting paid 12 months later. Now there are five cases and in each case, the amount for today is the same, i.e., RMB100, but the amount for 12 month later is different (See Figure 5 of Falk et al. (2018) for detailed settings).

24. What is the answer for the sum of 35 and 12?

25. Now you need to choose between getting paid in a week and getting paid in 12 months plus a week. Now there are five cases and in each case, the amount for in a week is the same, i.e., RMB100, but the amount for 12 month later is different (See Figure 5 of Falk et al. (2018) for detailed settings).

C Proofs

C.1 Proof of Lemma 1

First, as $V'(t^s) = U'(t^s)$, we have

$$U'(t^s) = -\frac{1}{a}e^{-\delta t^s}c(t^s) + \frac{e^{-\delta T}}{T}$$
$$U''(t^s) = -\frac{1}{a}[-\delta e^{-\delta t^s}c(t^s) + e^{-\delta t^s}c'(t)] = -\frac{1}{a}e^{-\delta t^s}[c'(t^s) - \delta c(t^s)]$$

Now we show the sign of U' and U'' depends on the size of the discount factor.

1. When $\delta \leq \frac{1}{1+e^{\frac{r-1}{r}T}}$, then $U''(t^s) < 0$, $U'(t^s)$ is monotonically decreasing. Since $U'(0) > 0$, $U(t^s)$ is either increasing and $t^{s*} = T$, or achieve maximum at $t^B(a)$. Therefore, $t^{s*} \in \{t^B(a), T\}$.

2. When $\delta \in (\frac{1}{1+e^{\frac{r-1}{r}T}}, \frac{1}{1+e^{-\frac{T}{r}}})$, it is easy to show that $U''(0) < 0$ and $U''(T) > 0$. Combining with the fact where there exists one and only one t^s which satisfies $U''(t^s) = 0$, $U'(t^s)$ is an U-shape function on $[0, T]$. Additionally, $U'(0) > 0$.

   (a) If $\forall t^s$, $U'(t^s) \geq 0$, then $V(t^s)$ is monotonic increasing. Therefore, $t^{s*} = T$. In other words, the student will keep exerting effort until the end of the semester.

   (b) If $\exists t^B(a) \leq t^A$, $U'(t^B(a)) = U'(t^A) = 0$, then $\forall t^s \in [0, t^B(a))$, $U'(t^s) > 0$; $t^s \in (t^B(a), t^A)$, $U'(t^s) < 0$; and $t^s \in [t^A, T]$, $U'(t^s) > 0$. As V(t) is a piecewise function, $t^{s*} \in \{t^B(a), t^N, T\}$.

   Depending on a student's ability, this implies that her optimal stopping timing could be the time where the marginal cost of exerting effort is equal to the marginal benefit, i.e., $t^{s*} = t^B(a)$, or the lecture that she is asked to take the notes, t^N, or the end of the semester, T.

(c) If $\exists t^B(a)$, $U'(t^B(a)) = 0$, and $U'(T) < 0$, similar to the previous case, we have $t^{s*} \in \{t^B(a), t^N\}$.

3. When $\delta \geq \frac{1}{1+e^{-\frac{T}{r}}}$, $U''(t^s) > 0$, $U'(t^s)$ is monotonically increasing. Therefore, $\forall t^s > 0$, $U'(t^s) > 0$, and we have $t^{s*} = T$.

C.2 Proof of Proposition 1

We prove the optimal stopping time for students with different abilities by comparing $V(T)$, $V(t^B(a))$ and $V(t^N)$.

1. $\forall a \geq a^h$, since $\int_{t^B(a)}^T g(a,t)dt > 0$, $V(T) = U(T) = U(t^B(a)) + \int_{t^B(a)}^T g(a,t)dt > U(t^B(a)) \geq V(t^B(a))$.

   As $V(T) - V(t^N) = \int_{t^N}^T g(a,t)dt$, we show that $V(t^N) < V(T)$ always hold. First, if $t^N < t^B(a)$, then $U(t^B(a)) - U(t^N) = \int_{t^N}^{t^B(a)} g(a,t)dt > 0$, hence $V(t^N) < V(t^B(a)) < V(T)$. Second, when $t^N > t^B(a)$, then $V(T) - V(t^N) = \int_{t^N}^T g(a,t)dt > 0$.

   Altogether, we have $\max\{V(t^B(a)), V(t^N), V(T)\} = V(T)$

2. When $a < a^l$, it is easy to show that $V(T) < V(t^B(a))$, hence we only need to compare $V(t^B(a))$ and $V(t^N)$.

   - If $t^N \leq t^B(a)$. Since $\forall t^s < t^B(a), U'(t^s) > 0$, hence $V(t^N) < V(t^B(a))$ and $t^{s*} = t^B(a)$
   - If $t^N \in (t^B(a), t^{NH}]$, where $t^{NH} = \max\{t \in [t^B(a), T] | \int_{t^B(a)}^t g(a,t)dt \geq -e^{-\delta t}S\}$, since $V(t^N) = U(t^N) = U(t^B(a)) + \int_{t^B(a)}^{t^N} g(a,t)dt \geq U(t^B(a)) - e^{-\delta t^N}S = V(t^B(a))$, we have $t^{s*} = t^N$

- If $t^N > t^{NH}$, since $V(t^N) = U(t^N) = U(t^B(a)) + \int_{t^B(a)}^{t^N} g(a,t)dt < U(t^B(a)) - e^{-\delta t^N} S = V(t^B(a))$, $t^{s*} = t^B(a)$

To summarize, the optimal stopping time for these low ability students are below.

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

3. When $a \in [a^l, a^h)$, we compare $V(t^B(a))$, $V(t^N)$ and $V(T)$.

   - If $t^N \leq t^B(a)$,

     First, $\forall t^s < t^B(a), U'(t^s) > 0$, hence $V(t^B(a)) > V(t^N)$. Second, as $a < a^h$ and $\int_{t^B(a)}^{T} g(a,t)dt < 0$, $V(t^B(a)) > V(t^B(a)) + \int_{t^B(a)}^{T} g(a,t)dt = U(T) = V(T)$. Altogether, $t^{s*} = t^B(a)$.
   - If $t^N \in (t^B(a), t^{HLB})$, where $t^{HLB} = \min\{t \in [t^B(a), T] | \int_t^T g(a,t)dt \geq 0\}$.

     First, as $\int_{t^N}^{T} g(a,t)dt < 0$, $V(T) = V(t^N) + \int_{t^N}^{T} g(a,t)dt < V(t^N)$. Second, it is easy to show that $t^{NH} > t^{HLB}$, consequently, we have $V(t^N) = U(t^N) > U(t^B(a)) > V(t^B(a))$. Altogether, $t^{s*} = t^N$.

   - If $t^N \in [t^{HLB}, t^{HUB}]$, where $t^{HUB} = \max\{t \in [t^B(a), T] | \int_{t^B(a)}^{T} g(a,t)dt \geq -e^{-\delta t} S\}$.

     First, as $t^N < t^{HUB}$, $\int_{t^B(a)}^{T} g(a,t)dt \geq -e^{-\delta t} S$, $V(T) = U(T) = U(t^B(a)) + \int_{t^B(a)}^{T} g(a,t)dt > U(t^B(a)) - e^{-\delta t^N} S = V(t^B(a))$. Second, as $t^N \geq t^{HLB}$, $\int_{t^N}^{T} g(a,t)dt \geq 0$, $V(T) = V(t^N) + \int_{t^N}^{T} g(a,t)dt > V(t^N)$. In summary, $t^{s*} = T$.
   - If $t^N > t^{HUB}$, First, as $t^N > t^{HUB}$, $\int_{t^B(a)}^{T} g(a,t)dt < -e^{-\delta t} S$, $V(T) = V(t^B(a)) + e^{-\delta t^N} S + \int_{t^B(a)}^{T} g(a,t)dt < V(t^B(a))$.

Second, as $t^N \geq t^{HLB}$, $\int_{t^N}^{T} g(a,t)dt \geq 0$, $V(T) = V(t^N) + \int_{t^N}^{T} g(a,t)dt > V(t^N)$.

Altogether, $V(t^B(a)) > V(T) > V(t^N)$, and $t^{s*} = t^B(a)$.

To summarize, the optimal stopping time for these middle ability range students are characterized below.

$$t^{s*} = \begin{cases} t^B(a), & t^N < t^B(a) \\ t^N, & t^B(a) \leq t^N < t^{HLB} \\ T, & t^{HLB} \leq t^N \leq t^{HUB} \\ t^B(a), & t^N > t^{HUB} \end{cases}$$

C.3 Proof of Proposition 2

We first prove the following lemma. Lemma 2 implies that a student with higher ability would keep learning longer or at least as long as students with lower ability.

Lemma 2. *If $a_1 > a_2$, then $t^{s*}(a_1) \geq t^{s*}(a_2)$.*

Proof. We prove this by contradiction, i.e., if $a_1 > a_2$, $t^{s*}(a_1) < t^{s*}(a_2)$.

Consider the case where $t^N < t^{s*}(a_1)$. First, since $t^{s*}(a_1) = \arg\max V(t^s|a_1)$, her utility for stopping at any other time stamp, e.g., $t^{s*}(a_2)$, is lower than that stopping at $t^{s*}(a_1)$, i.e., $V(t^{s*}(a_2)|a_1) - V(t^{s*}(a_1)|a_1) = \int_{t^{s*}(a_1)}^{t^{s*}(a_2)} g(a_1)dt < 0$.

Similarly, we have $V(t^{s*}(a_2)|a_2) - V(t^{s*}(a_1)|a_2) = \int_{t^{s*}(a_1)}^{t^{s*}(a_2)} g(a_2)dt = \int_{t^{s*}(a_1)}^{t^{s*}(a_2)} g(a_2)dt > 0$.

Altogether, $\int_{t^{s*}(a_1)}^{t^{s*}(a_2)} g(a_1)dt - \int_{t^{s*}(a_1)}^{t^{s*}(a_2)} g(a_2)dt = \int_{t^{s*}(a_1)}^{t^{s*}(a_2)} [g(a_1) - g(a_2)]dt < 0$. Consequently, $g(a_1) - g(a_2) < 0$

On the other hand, as $g(a,t) = \frac{e^{-\delta T}}{T} - \frac{e^{-\delta t}}{a}c(t)$, we have $\frac{\partial g(a,t)}{\partial a} > 0$, which is a contradiction.

The proof for $t^N \in [t^{s*}(a_1), t^{s*}(a_2)]$ and $t^N > t^{s*}(a_2)$ is similar and we omit it. $\square$

Next, we show that conditional on t^N, the optimal stopping time students t^{s*} for students with different a can be characterized below,

$$t^{s*} = \begin{cases} t^B(a) & \underline{a} \le a < a^{NH}(t^N) \\ \max\{t^B(a), t^N\} & a^{NH}(t^N) \le a < a^H(t^N) \\ T & a^H(t^N) \le a \le \bar{a} \end{cases} \tag{1}$$

, where $a^{NH}(t^N)$ and $a^H(t^N)$ are two ability cutoffs which ties to the size of t^N. Specifically, we have:

$$a^{NH}(t^N) = \begin{cases} \underline{a} & t^N < t^{NH}(\underline{a}) \\ \min\{a| \int_{t^B(a)}^{t^N} g(a,t)dt \ge -e^{-\delta t^N} S\} & t^{NH}(\underline{a}) \le t^N < t^{NH}(a^l) \\ \min\{a| \int_{t^B(a)}^{T} g(a,t)dt \ge -e^{-\delta t^N} S\} & t^{NH}(a^l) \le t^N \end{cases} \tag{2}$$

$$a^H(t^N) = \begin{cases} a^h & t^N < t^{HLB}(a^h) \\ \min\{a| \int_{t^N}^{T} g(a,t)dt \ge 0\} & t^{HLB}(a^h) \le t^N < t^{NH}(a^l) \\ \min\{a| \int_{t^B(a)}^{T} g(a,t)dt \ge -e^{-\delta t^N} S\} & t^{NH}(a^l) \le t^N \end{cases} \tag{3}$$

By Lemma 1, $t^{s*} \in \{t^B(a), t^N, T\}$. Therefore, we only need to compare the value associated with these three time stamps.

1. For students with $a^H(t^N) \le a \le \bar{a}$, we complete the characterization for the optimal stopping time by consider the following three cases.

(a) If $t^N < t^{HLB}(a^h)$, by Equation 3, $a^H(t^N) = a^h$, therefore, we have $a \geq a^h$. By Proposition 1, $t^{s*} = T$;

(b) If $t^{HLB}(a^h) \leq t^N < t^{NH}(a^l)$, then $V(T) - V(t^N) = \int_{t^N}^{T} g(a,t)dt \geq 0$. Second, since $t^N < t^{NH}(a^l)$, $V(t^N) - V(t^B(a)) = \int_{t^B(a)}^{t^N} g(a,t)dt + e^{-\delta t} S > 0$, $V(T) \geq V(t^N) > V(t^B(a))$ and $t^{s*} = T$;

(c) If $t^{NH}(a^l) \leq t^N$, by the definition of a^l and t^{NH}, $t^{NH}(a^l) = t^{HLB}(a^l)$. Therefore, $V(T) - V(t^N) = \int_{t^N}^{T} g(a,t)dt \geq 0$. Moreover, since $a > a^H(t^N)$, $V(T) - V(t^B(a)) = \int_{t^B(a)}^{T} g(a,t)dt + e^{-\delta t^N} S > 0$. Altogether, $V(T) \geq V(t^N)$ and $V(T) > V(t^B(a))$, and $t^{s*} = T$.

Altogether, for students with $a^H(t^N) \leq a \leq \bar{a}$, $t^{s*} = T$.

Next, we show that $\forall a < a^H(t^N)$, $t^{s*} < T$. The proof is contained in the following three cases,

(a) If $t^N < t^{HLB}(a^h)$, then $a < a^h$. By definition of t^{HLB}, $V(T) - V(t^N) = \int_{t^N}^{T} g(a,t)dt < 0$, hence $t^{s*} \neq T$.

(b) If $t^{HLB}(a^h) \leq t^N < t^{NH}(a^l)$, then $\int_{t^N}^{T} g(a,t)dt < 0$. Similarly, $V(T) - V(t^N) < 0$ and $t^{s*} \neq T$.

(c) If $t^{NH}(a^l) \leq t^N$, then $V(T) - V(t^B(a)) = \int_{t^B(a)}^{T} g(a,t)dt + e^{-\delta t^N} S < 0$. Therefore, $t^{s*} \neq T$.

To summarize, $\forall a < a^H(t^N)$, $t^{s*} \neq T$. By Lemma 1, $\forall a < a^H(t^N)$, $t^{s*} \in \{t^B(a), t^N\}$. This implies that for these students, we only need to compare their value for two time stamps: $t^B(a)$ and t^N.

2. For students with $a^{NH}(t^N) \leq a < a^H(t^N)$, we consider the following three cases.

(a) If $t^N < t^{NH}(\underline{a})$

i. If $t^N < t^B(a)$, since $\forall t < t^B(a), g(t,a) > 0$, $V(t^B(a)) - V(t^N) = \int_{t^N}^{t^B(a)} g(t,a) + e^{-\delta t^N} S > 0$, hence $t^{s*} = t^B(a)$;

ii. If $t^B(a) \le t^N$, since $t^N < t^{NH}(\underline{a})$ and $\frac{\partial t^{NH}}{\partial a} > 0$, $\forall a, t^N < t^{NH}$ and $V(t^N) - V(t^B(a)) = \int_{t^B(a)}^{t^N} g(t,a) + e^{-\delta t^N} S > 0$, $t^{s*} = t^N$.

(b) If $t^{NH}(\underline{a}) \le t^N < t^{NH}(a^l)$

i. If $t^N < t^B(a)$, since $\forall t < t^B(a), g(t,a) > 0$, $V(t^B(a)) - V(t^N) = \int_{t^N}^{t^B(a)} g(t,a) + e^{-\delta t^N} S > 0$, hence $t^{s*} = t^B(a)$;

ii. If $t^B(a) \le t^N$, since $a \ge a^{NH}(t^N) = \min\{a | \int_{t^B(a)}^{t^N} g(a,t)dt + e^{-\delta t^N} S \ge 0\}$, $V(t^N) - V(t^B(a)) = \int_{t^B(a)}^{t^N} g(a,t)dt + e^{-\delta t^N} \ge 0$, $t^{s*} = t^N$.

(c) If $t^{NH}(a^l) \le t^N$, $a^{NH} = a^H$ and it is impossible that $a^{NH} \le a < a^H(t^N)$.

In total, for students with $a^{NH}(t^N) \le a < a^H(t^N)$, $t^{s*} = \max\{t^B(a), t^N\}$.

3. For students with $\underline{a} \le a < a^{NH}(t^N)$,

(a) If $t^N < t^{NH}(\underline{a})$, $a^{NH}(t^N) = \underline{a}$ and it is impossible that $\underline{a} \le a < a^{NH}(t^N)$;

(b) If $t^{NH}(\underline{a}) \le t^N < t^{NH}(a^l)$, since $a < a^{NH}(t^N)$, $V(t^N) - V(t^B(a)) = \int_{t^B(a)}^{t^N} g(a,t)dt + e^{-\delta t^N} S < 0$, hence $t^{s*} = t^B(a)$.

(c) If $t^{NH}(a^l) \le t^N$, because $a^{NH}(t^N) = \min\{a | \int_{t^B(a)}^{T} g(a,t)dt \ge -e^{-\delta t^N} S\}$ and $\frac{\partial a^{NH}(t^N)}{\partial t^N} > 0$. Therefore, $a < a^{NH}(t^{NH}(a^l)) = a^l$. Since $\frac{\partial t^{NH}}{\partial a} > 0$, $\forall a < a^l$, $t^N \ge t^{NH}(a^l) > t^{NH}(a)$, $V(t^N) - V(t^B(a)) = \int_{t^B(a)}^{t^N} g(a,t)dt + e^{-\delta t^N} S < 0$, $t^{s*} = t^B(a)$.

Altogether, for students with $\underline{a} \le a < a^{NH}(t^N)$, $t^{s*} = t^B(a)$.

The objective function of the instructor is to choose t^N to maximize students expected total effort. Therefore, based on Equation 1, we have

$$\max_{t^N} TE(t^N) = \int_{\underline{a}}^{a^{NH}(t^N)} t^B(a)dF(a) + \int_{a^{NH}(t^N)}^{a^H(t^N)} \max\{t^N, t^B(a)\}dF(a) + \int_{a^H(t^N)}^{\bar{a}} TdF(a)$$

We define $\underline{t}^N = \min\{t^{HLB}(a^h), t^{NH}(\underline{a})\}$, and first prove that $\forall t^N \in [0, \underline{t}^N)$, $t^N \notin N^*$. In other words, the optimal task timing is never too early.

- Following the characterization in Equations 2 and 3, we have $\forall t^N \in [0, \underline{t}^N)$, $a^{NH}(t^N) = \underline{a}$ and $a^H(t^N) = a^h$. Hence we have

$$\begin{aligned}
TE(t^N) &= \int_{\underline{a}}^{a^{NH}(t^N)} t^B(a)dF(a) + \int_{a^{NH}(t^N)}^{a^H(t^N)} \max\{t^N, t^B(a)\}dF(a) + \int_{a^H(t^N)}^{\bar{a}} TdF(a) \\
&= \int_{\underline{a}}^{\underline{a}} t^B(a)dF(a) + \int_{\underline{a}}^{a^h} \max\{t^N, t^B(a)\}dF(a) + \int_{a^h}^{\bar{a}} TdF(a) \\
&= 0 + \int_0^{a^h} \max\{t^N, t^B(a)\}dF(a) + \int_{a^h}^{\bar{a}} TdF(a)
\end{aligned}$$

  Obviously, $\frac{\partial \int_0^{a^h} \max\{t^N, t^B(a)\}dF(a)}{\partial t^N} \geq 0$. Because $\max\{t^N, t^B(a)\} \not\equiv t^B(a)$, we have $\frac{\partial \int_0^{a^h} \max\{t^N, t^B(a)\}dF(a)}{\partial t^N} \not\equiv 0$. Therefore, $\forall t^N \in [0, \underline{t}^N)$, $TE(t^N) < TE(\underline{t}^N)$, hence $t^N \notin N^*$.

Similarly, we define $\bar{t}^N = t^{NH}(a^l)$. We then show that $\forall t^N \in (\bar{t}^N, T]$, $t^N \notin N^*$. In other words, the optimal task timing is never too late.

- By Equations 2 and 3, $\forall t^N > \bar{t}^N$, $a^H(t^N) = a^{NH}(t^N)$, therefore, we have

$$
\begin{aligned}
TE(t^N) &= \int_{\underline{a}}^{a^{NH}(t^N)} t^B(a)dF(a) + \int_{a^{NH}(t^N)}^{a^H(t^N)} \max\{t^N, t^B(a)\}dF(a) + \int_{a^H(t^N)}^{\bar{a}} TdF(a) \\
&= \int_{\underline{a}}^{a^H(t^N)} t^B(a)dF(a) + 0 + \int_{a^H(t^N)}^{\bar{a}} TdF(a) \\
&= \int_{\underline{a}}^{\bar{a}} t^B(a)dF(a) + \int_{a^H(t^N)}^{\bar{a}} (T - t^B(a))dF(a)
\end{aligned}
$$

Because $a^H(t^N) = \min\{a| \int_{t^B(a)}^{T} g(a,t)dt \geq -e^{-\delta t^N} S\}$, It is easy to show that $\frac{\partial a^H(t^N)}{\partial t^N} > 0$. Besides, since $\frac{\partial \int_{a^H(t^N)}^{\bar{a}} (T - t^B(a))dF(a)}{\partial a^H(t^N)} < 0$, $\frac{\partial TE(t^N)}{\partial a^H(t^N)} < 0$. Altogether, $\frac{\partial TE(t^N)}{\partial t^N} = \frac{\partial TE(t^N)}{\partial a^H(t^N)} \frac{\partial a^H(t^N)}{\partial t^N} < 0$. Therefore, $\forall t^N \in (\bar{t}^N, T]$, $TE(t^N) < TE(\bar{t}^N)$, hence $t^N \notin N^*$.

Altogether, we have $t^{N*} \in [\underline{t}^N, \bar{t}^N]$.

D Sample Notes

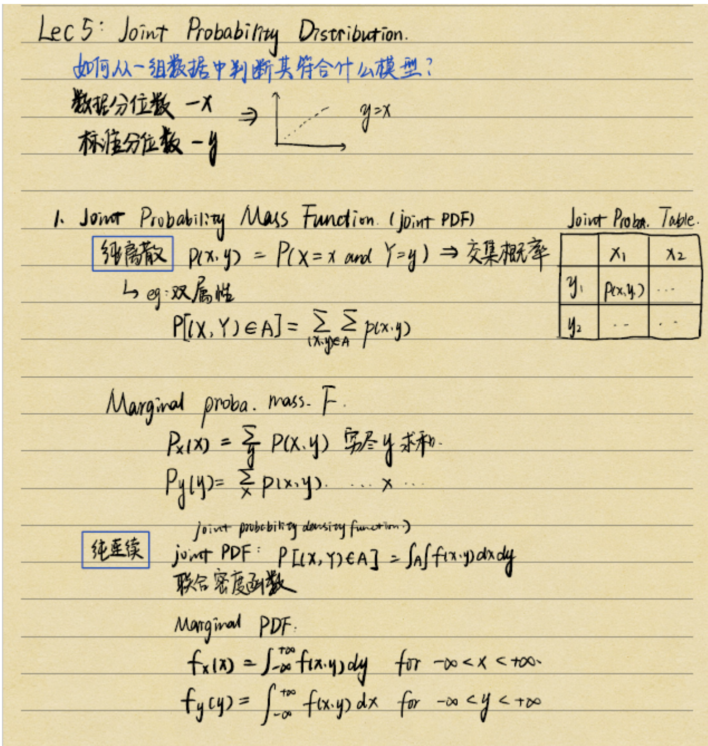


Figure A1: Sample Notes from A Student