Incentive Design on MOOC: A Field Experiment

Jie Gong   Tracy Xiao Liu   Jie Tang *

May 13, 2019

Abstract

In this study, we examine the impact of monetary incentives on learner engagement and learning outcomes in massive open online courses (MOOCs). While MOOCs have innovated education by offering high-quality interactive educational resources to users worldwide, maintaining student learning enthusiasm in these courses is a challenge. To address this issue, we conduct a field experiment in which users are given monetary incentives to engage in online learning. Our results show that those given a monetary incentive are more likely to submit homework and to gain higher grades. The effect largely reflects the continued involvement of regular users whose activity would otherwise decrease over time. We further find that the effect persists even after we remove the monetary incentives and that it spills over into learning behavior in other courses in the same and subsequent semester. Lastly, we find that females and users from regions with fewer educational institutions show stronger treatment effects. Overall, our findings suggest that monetary incentives counteract the decay of learning engagement and may help online education users form persistent learning habits.

Keywords: Incentive, MOOCs, Field Experiment

*Gong: Department of Strategy and Policy, NUS Business School, National University of Singapore. Email: gong@nus.edu.sg. Liu: Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China. Email: liuxiao@sem.tsinghua.edu.cn. Tang: Department of Computer Science and Technology, Tsinghua University, Beijing, China. Email: jietang@tsinghua.edu.cn. We thank Yan Chen, Jonathan Guryan, Jennifer Hunt, Peter Kuhn, Erica Li, Sherry Xin Li, Fangwen Lu, Ben Roth and seminar participants at National University of Singapore, Renmin University, Beijing Normal University, Queensland University of Technology, Tsinghua University, 6th Annual Xiamen University International Workshop on Experimental Economics, The First International Workshop on AI and Big Data Analytics in MOOCs (AI-MOOC), 2017 International Symposium on Contemporary Labor Economics Institute for Economic and Social Research, Jinan University for helpful discussions and comments, and Jiezhong Qiu, Han Zhang, Fang Zhang, and Shuhuai Zhang for excellent research assistance. We gratefully acknowledge the support from XuetangX and the Online Education Office, Tsinghua University. The financial support from Ministry of Education of the People's Republic of China (MOE) Research Center for Online Education is gratefully acknowledged.

1 Introduction

Over the past decade, online learning has transformed into an established component of higher education. Massive Open Online Courses (MOOCs) have attracted millions of learners from various backgrounds and regions by providing high quality educational resources. In addition, traditional universities have launched their own online endeavors, providing entire courses through platforms such as Coursera, EdX, and Udacity as well as engaging their on-campus students through the use of online learning tools. Some of these universities, including MIT and Georgia Tech, allow their students to earn credit from MOOC courses. Outside of the university context, the government and public sector have developed online learning programs that provide education and training to a broader population of students and working professionals. For example, China's "Internet Plus" plan promotes online platforms as an innovative higher education option. Similarly, the respective governments in Singapore and France fund MOOCs for formal education and job training with the aim of targeting those who lack access to traditional education resources due to financial, geographic, or opportunity cost constraints.[1]

However, despite the growth of online learning options, whether these courses instill the same quality of learning as traditional classrooms is unclear. Indeed, it is an open question as to whether online learning, by its very nature, can fully utilize the educational resources invested by teachers and universities. In practice, MOOC courses experience low completion rates and a significant decay in learner activity across the semester. Kizilcec et al. (2013) and Seaton et al. (2014) report that only 5% of MOOC users have completed a course. In a sample of UPenn Coursera courses, Banerjee and Duflo (2014) find that only 2 to 14% of users who started a course showed any activity by the end of the course. Their study further shows that this decrease in activity reflects the challenge of maintaining user self-control or persistence in an online learning environment. While a traditional offline classroom can mitigate self-control challenges through scheduled class meetings, instructor monitoring, peer pressure, or other incentives built into the class structure, the online class environment may need other measures to help users discipline themselves and develop sustained learning habits. While platforms such as Coursera address this issue by requiring advanced payment for their online courses, this solution is less feasible for government or public sector online education programs.

The aim of this study is to examine alternative methods of incentivizing online learners.

[1]Sources: China: http://www.gov.cn/zhengce/content/2015-07/04/content_10002.htm.
Singapore: https://www.imda.gov.sg/about/newsroom/archived/ida/media-releases/2014/ida-first-massive-open-online-course-training-for-data-science-and-analytics-goes-live.
France: http://www.sup-numerique.gouv.fr/.

In particular, we examine whether monetary incentives can effectively mitigate the observed decay in activity across an online course. To do so, we conduct a field experiment on the digital education platform XuetangX, the third worldwide largest MOOC platform. For our study, we select two courses offered in Spring 2015, *Cultural Treasure and Chinese Culture (Chinese Culture)* and *Data Structure and Algorithm (Data Structure)*, and reward their students for completing homework assignments across a four-week period in the middle of the course. For each course, we randomly assign subjects to either the control group or one of six treatment groups. In three of the treatment groups, subjects are rewarded 1, 10, and 100 rmb, respectively, for each completed assignment that exceeds our pre-specified grade threshold. In the other three treatment groups, subjects initially receive a deposit and then lose 1, 10, and 100 rmb, respectively, for each assignment short of the grade threshold. We implement the incentive for three assignments around the middle of the semester and collect learners assignment submission activity and grades before, during, and after the intervention.

Overall, our experimental results show that providing a monetary incentive improves both engagement and performance in online courses. Specifically, we find that a large incentive (100 rmb reward or loss) significantly improves both the assignment completion rate and student grades. By contrast, a medium incentive (10 rmb reward or loss) shows significant effects for students taking *Chinese Culture*, but not for those taking *Data Structure*; a small incentive (1 rmb reward or loss) shows no effects on student engagement in either course. The improved engagement reflects sustained activity by regular users rather than an uptake in activity by inactive users. In addition, we find that the effects persist even after we remove the monetary incentive, and that they spill over to engagement and performance in other courses in the same semester and course completion in the subsequent semester. Lastly, we find that female subjects as well as those from regions with fewer higher education institutions are more responsive to monetary incentives.

Our findings suggest that offering monetary incentive could be a scalable solution to sustain online learners' engagement and performance. In our experiment, 100 rmb (around 15 USD) is universally effective and more than sufficient for one course; the scheme covers only three sets of homework over four weeks. In comparison, the benefits are large in magnitude and persistent over time; we do not find evidence of monetary incentive crowding out intrinsic motivation either. Another point worth noting is that, offering incentive may affect the decision to enroll in the courses (e.g., attract more users); while offline classrooms have limited seats, the online environment does not have such crowding costs and therefore can accommodate the increased users who are attracted by the incentives.

Our findings have implications on the policies that promote online learning. There has been debates about whether low engagement and completion rate is really a failure of MOOCs and should be concerned about. It is plausible that users shop around because the cost of trying courses are negligible. In that case, low engagement and high drop out rates could be the result of a matching process. However, if the specific nature of online environment—e.g., lack of monitoring and peer groups—lead to severe self-control problems, there may be amendments to the structure of the platforms, such as offering monetary incentive, that would alleviate the problem. Indeed, Banerjee and Duflo (2014) find that less organized students are less likely to succeed in a MOOC due to a lack of completion of assignments rather than poor performance on completed assignments. Among the participants in our study, 23% claim that they enroll the course in order to earn a certificate and 22% indicate that they enroll because the content would be helpful to their jobs. Our findings on persistence of the incentive effects also provide some evidence that the baseline low engagement may be sub-optimal. The evidence of spillovers across time and courses further imply that increased engagement is more likely to be driven by changes in learning habits rather than selection of courses.

Our study provides key contributions to three areas of existing research. First, our findings contribute to the literature on incentives in education by adding a new piece of evidence strengthening the effect of financial incentives in motivating learning. Prior studies mostly focus on traditional offline classrooms and find mixed results. On the one hand, several studies find a short-term, positive effect of incentives on students learning performance (Angrist et al., 2002, 2009; Angrist and Lavy, 2009; Braun et al., 2011; Levitt et al., 2016a); a few find a long-term post-incentive effect as well (Angrist et al., 2006, 2009). On the other hand, using a large-scale field experiment in three U.S. cities, Fryer (2011) finds no significant effect of a financial incentive on students state scores.

We make a first attempt to experimentally investigate the effect of financial incentives on online learning. Examining previous evidence in offline classrooms, it is ex ante unclear whether these results will also hold in online settings. It is possible that online learners may have different and/or more diverse motivations in pursuing their education, including intrinsic goals, personal interest in the topic, or career advancement reasons. Therefore, they may be less responsive to monetary incentives than offline students who learn in order to obtain credits and grades toward a degree. Another difference between offline and online learners is that the online learning environment does not provide monitoring and peer group control mechanisms to motivate learning. Thus, the self-control problem becomes a significant hurdle that may be mitigated by the use of a strengthened incentive. Indeed, our findings suggest that offering incentives helps to maintain learning enthusiasm in the online setting.

Another innovation that we make is examining learning activities such as homework completion and lecture video viewing, thus extending from traditionally focused learning outcomes (grades). We also examine the effect on student engagement and performance in other (non-targeted) courses both in the current and subsequent semesters. Our results therefore provide a richer understanding of the education production function and development of learning habits in the presence of incentives. Lastly, from an empirical perspective, a controlled experiment on an online platform allows us to observe various user activities and alleviate potentially confounding effects that may be present in offline classrooms, such as peer interactions and teacher influence.

In the literature of incentives in education, there is variation in what input tasks (e.g., reading a book) or final outputs (e.g., test scores) are rewarded. One findings in Fryer (2011) is that incentives for inputs, such as attendance, tend to work better than incentives that reward outcomes, such as better grades. Clark et al. (2017) conduct two field experiments among college students and find that setting task-based goals has larger positive effects on course performance than setting performance-based goals. Our experiment rewards students for completing homework, therefore the positive treatment effects echo the efficacy of rewarding concrete tasks.

The second area to which our study contributes is the stream of putting behavioral economics to practice. In designing the structures and mechanisms for our study, we leverage a number of behavioral economics theories. We systematically evaluate the efficacy of two design features for monetary incentives: incentive size and framing. In determining the amount of incentive to offer, we draw on Gneezy and Rustichini (2000), who show that small incentives may crowd out intrinsic motivations and lead to inferior performance, as well as Ariely et al. (2009), who find that excessively high incentives may also have a detrimental effect on individual productivity. In determining how to frame our incentives, we draw on Hossain and List (2012) and Hong et al. (2015), who find that framing bonus as loss significantly increases factory worker productivity. We also draw on Andreoni (1995), who finds that positive framing of an incentive significantly increases participants contributions in public goods. Both Fryer et al. (2012) and Levitt et al. (2016b) find that framing incentives in the loss-domain is more effective for enhancing students' performances. Finally, Karlan et al. (2016) finds no significant effect of incentive framing on individuals saving behavior while Chen et al. (2018) similarly finds a lack of a framing effect on students arrival time in experimental sessions.

Our study further contributes to behavioral economics by documenting persistent and spillover effects of incentives. Within this area, Charness and Gneezy (2009) find that paying people to visit a gym helps develop exercise habits and improves health outcomes in the long run. Hussam et al. (2017) find that providing monitoring and incentives can boost hand washing rates even when those manipulations are removed, suggesting that participants internalize the habit of hand washing in the long-run. However, other studies show that treatment effects do not persist over time (Gneezy and List 2006; Meier 2007). We suggest that understanding the long-term and spillover effects of incentives is useful in designing and implementing monetary incentives at both the academic and policy level.

Thirdly, our study contributes to the literature on the effectiveness of online learning. Within this literature, Deming et al. (2015) find evidence that colleges charge lower prices for online coursework, suggesting that online learning technologies make higher education more economically feasible for students. In another study, Acemoglu et al. (2014) argue that web-based technology has the power to democratize education by distributing resources more equally among students and by complementing non web-based inputs of low-skilled local teachers. However, Hansen and Reich (2015) find that MOOC participants from the U.S. tend to live in better neighborhoods than the general population and students with better socio-economic background are more likely to succeed in MOOC courses. Regarding learning effectiveness, Banerjee and Duflo (2014) document a significant engagement decay in online courses and find evidence for a self-control problem. To combat this issue, Zhang et al. (2017) find that promoting social interaction on course discussion boards significantly improves both students completion rates and their course grades. Our findings add yet another possible mechanism to encourage student engagement and performance, i.e., offering a monetary incentive.

The rest of the paper is organized as follows. Section 2 introduces the MOOC platform on which we conduct the experiment. Section 3 describes the experimental design. Section 4 reports our experimental results. Section 5 concludes and discusses implications and future research.

2 Field Setting: XuetangX

XuetangX was launched in China in 2013 as a start-up MOOC platform affiliated with Tsinghua University and the Minister of Education (MOE) of China. By 2015, when we conducted the experiment, it had offered 670 courses to more than 1,700,000 registered users.[2] In addition to providing its own course content, XuetangX also partners with EdX and collaborates with top universities, providing users with access to courses offered by US universities including MIT, Stanford, and UC Berkeley. Compared to other MOOCs, XuetangX is more public in nature, providing a greater number of free courses and accounts

[2] As of May 2018, XuetangX had attracted more than 10 million users and offered more than 1000 courses.

to alleviate education inequality and to promote life-long learning in China.[3]

XuetangX courses can be roughly divided into two fields: art and literature, and science and engineering. Courses in the two fields typically differ in their style, workload, learning objectives and student composition (Qiu et al., 2016).[4] In our study, we draw on one course from each of these fields to conduct our experiment. Most of the courses on XuetangX follow a semester system. At the beginning of each semester, courses are listed for users to browse through and decide whether to enroll. There is no restriction on the number of courses any user can register for in a single semester. Enrollment for the courses remains open throughout the semester so that users can enroll or drop any time before the course ends. Dropping a course does not trigger any penalty. Compared to other MOOCs such as Coursera that require prepayment as an incentive for course completion, XuetangX employs very few structures to foster learning incentives. As such, it provides an open canvas for us to implement a learning incentive treatment.

Courses on XuetangX are structured by chapters, with corresponding lecture videos and assignments posted frequently. Science and engineering courses are generally perceived as more demanding than their art and literature counterparts in that they require more academically challenging assignments. Students taking a course on XuetangX receive a final grade for the course determined by some combination of assignments, exams, and projects. Once enrolled in a course, a user can access the posted course materials from her account in order to view lecture videos, complete assignments, post a thread in course forums, or respond to other students posts. Qiu et al. (2016) describe these activities and summarizes observed patterns in student activities on the platform.

Like other MOOCs, XuetangX faces a participation issue reflected in low user engagement and a decrease in engagement over a course. For example, Qiu et al. (2016) show that both lecture video viewing and assignment submission decrease significantly over time. Similarly, Feng et al. (2018) find that the likelihood of a user dropping a course is positively correlated with him or her dropping another course, suggesting an overall engagement decrease rather than effort reallocation among courses. Figure 1 presents the average homework submission rate and grades of XuetangX users in our two selected courses over time, using 2014 fall data from *Chinese Culture* and *Data Structure*, i.e., one semester before we conducted the experiment. For both courses, homework submission rates and grades dropped quickly after the first two weeks, with a reduction of more than half by midterm. This finding is consistent

[3]For instance, XuetangX and Tsinghua University provide half a million free accounts to more than 0.5 million delivery staff working at Meituan-Dianping, the world's largest online on-demand delivery platform. Source: `http://news.sina.com.cn/o/2018-04-02/doc-ifysvmhv5582478.shtml`.

[4]For instance, art and literature courses on average attract more female users than do science and engineering courses.

with engagement patterns found in other MOOC studies (e.g., Banerjee and Duflo 2014). Another observation in our initial assessment of the two courses is that both submission activity and grades are higher among students in the *Chinese Culture* course than for those in the *Data Structure* course, reflecting differences in the course features and assignment difficulty.

3 Experimental Design

To investigate the effect of monetary incentives on learner engagement and performance, we use a 3×2 factorial design, with our control group receiving no incentive. In treatment groups, we vary the incentive size and framing. On the size dimension, we offer three levels of incentives—small, medium and large—to investigate the degree to which incentive size affects engagement and performance. On the framing dimension, we investigate whether positive versus negative framing (i.e, gain versus loss) leads to different effects for a given size of incentive.

As documented in Qiu et al. (2016), courses in art and literature differ from those in science and engineering in both their requirements, the difficulty level and the composition of students enrolled. Consequently, we select one course from each category to capture any possible heterogeneous treatment effects across courses and disciplines. *Chinese Culture* and *Data Structure* are chosen because both have been offered on XuetangX for three semesters. Their relatively mature course design and materials provide greater confidence that our results are not affected by idiosyncratic shocks from the courses per se. More importantly, as shown in Figure 1, *Data Structure* appears to have more challenging homework assignments that require greater effort from the students to complete. Our premise is that the more challenging course might require a larger amount of incentive to sustain student engagement.

The first course in our experiment, *Chinese Culture*, is offered by the Department of History, Tsinghua University and lasts from March 2, 2015 to June 22, 2015. The course consists of 16 lectures, with one assignment posted to correspond to each lecture.[5] Students can complete the assignments any time before the course ends. These homework assignments collectively account for 40% of a student's final grade. In addition, a midterm exam accounts for 20% and a final exam 40% of the grade. For our experiment, three sets of assignments, i.e., assignments 8, 9, and 10, are subject to our incentive scheme; we collect data on user activities throughout the whole semester. We choose three sets of assignment around the middle of the semester for two reasons: First, the enrolment is mostly stabilized by then and we don't risk losing a large amount of participants during the experiment; second, the first

[5]The only exception is the last lecture, which has two assignments.

few assignments are not subject to any intervention, and provide useful information about participants' baseline activity and performance; we use the data to test randomization and as further controls in the regression.

The second course in our experiment, *Data Structure*, is offered by the Department of Computer Science and Technology, Tsinghua University and lasts from March 3, 2015 to June 23, 2015. There are 12 lectures, and, similar to *Chinese Culture*, an assignment is posted at the end of each lecture. However, each assignment is due one month after it is posted. Each assignment accounts for 5% of a student's final grade (60% in total). In addition, 4 programming projects accounts for 40%. Our intervention targets three assignments, i.e., assignments 6, 7, and 8; again, we collect data on user activities throughout the whole semester.

3.1 Treatments

Each of our treatment groups is provided with a monetary incentive for successful completion of homework. We define successful completion as a homework submission that correctly answers at least 80% of the questions. This threshold is determined using benchmark data from homework records for each course in its previous offerings. Specifically, we summarize student performance for each of the two courses in the 2014 fall semester, i.e., one semester before our experiment and find that, conditional on submission (non-zero grades), both the mean and median scores are 80% for each courses. We therefore consider this to be a feasible target that students can meet with a reasonable amount of effort.[6]

The subjects in our experiment are randomly assigned to either the control or one of six treatment groups. Users in the control group receive no incentive. They receive only one email at the beginning of the experiment, encouraging them to complete their assignments. The same email is also sent to the treatment groups. For instance, the control group in the *Data Structure* course receives the following message:

> *Data Structure* has updated to the 6th homework. If you want to get your certificate, you should finish your homework on time and try your best to get high grades!

In addition to the above message, treatment groups receive a paragraph in their email that outlines their monetary incentive. Depending on which treatment they are assigned to, students may be provided with one of three different levels of payment size, 1 rmb, 10

[6]Interviews with TAs suggest that an average student should be able to complete an assignment within 30 minutes for *Data Structure* and 10 minutes for *Chinese Culture*.

rmb and 100 rmb. 1 rmb is considered a small amount and is used to test whether a small monetary incentive may crowd out the intrinsic motivation of learning. 10 rmb represents a medium amount, an acceptable amount as a reward. 100 rmb is the largest amount and is considered an generous reward.[7]

For each level of incentive size, we also vary how the incentives are framed using a similar implementation as in Hossain and List (2012). For example, our positive framing introduces the incentive as a gain for each homework that receives a score of at least 80%. Positive-framing subjects receive the following message:

> For the next 3 assignments, you will receive an X rmb reward for every assignment that your grade is ≥80% of the total score.[8]

By contrast, the negative framing introduces the incentive as a loss for each homework that fails to meet the 80% threshold. Negative-framing subjects receive the following message:

> For the next 3 assignments, you will receive a one-time bonus of 3×X rmb. However, for every assignment with grade <80% of the total score, the bonus will be reduced by X.

To minimize potential collusion between students, we impose a deadline such that, to claim the monetary reward, students must submit their homework within two weeks of the assignment posting date. User activity from the past semester shows that most homework submissions are made within two weeks of the assignment posting date.[9] Online Appendix A includes a sample of the experimental emails sent to subjects in the treatment groups.

3.2 Experimental Procedure

Table 1 summarizes our experiment timeline as well as our data. We conducted our experiment in the Spring of 2015. On April 6th, 2015, we sent recruiting emails to enrolled students in the two courses (5,714 users in *Chinese Culture* and 9,720 users in *Data Structure*). We also posted a recruiting message on the announcement board for each course. Online Appendix B includes a sample of the recruiting email/message. By sending recruiting emails explicitly, we target active learners who at least may respond to messages and manipulation

[7] For comparison, student TAs at Tsinghua University are paid 24 rmb per hour.

[8] $X \in \{1, 10, 100\}$.

[9] In the pre-experiment survey, only 7.8% of the participants report that they have friends taking the same course. Users' IP addresses are also geographically scattered.

(Chen and Konstan 2015).[10] By April 14, 337 users from *Chinese Culture* and 455 users from *Data Structure* had signed up for our study and filled in a survey on their demographic characteristics. Online Appendix C includes the pre-experiment survey questionnaires. Those participants understand that they signed up for a study about online learning, but were not told the details and the purpose of the experiment. Our sample group excludes users who signed up for XuetangX after we posted the recruiting message and individuals who signed up for our study but did not enroll in either course. There are 9 users who registered for both courses and they are assigned as subjects for only the *Data Structure* course.

Altogether, our subject group consists of 328 users enrolled in the *Chinese Culture* course and 432 enrolled in the *Data Structure* course. Table 2 reports the summary statistics of our participant demographic characteristics as well as the pre-experimental course performance data. The statistics in Table 2 show that our participants are generally young (mean age is 27 years for *Chinese Culture* and 24 years for *Data Structure*), educated (the majority have a high school or college degree), and experienced with MOOC platforms (on average they have taken two courses at XuetangX). A notable difference between the two courses is the gender composition. There are more female than male participants in the *Chinese Culture* course and more male than female participants in the *Data Structure* course. Comparing their activity and performance in the first six weeks of the course (before they sign up for the experiment), we also see that the *Chinese Culture* in general has a higher participation rate and student performance profile, possibly due to less challenging content and requirement.

For each course, we randomly assign each subject to either the control or one of the six treatment groups (i.e., complete randomization). For each of the user demographics and learning experience variables, we conduct Pairwise Kolmogorov-Smirnov tests between our treatment groups and the control group. All comparisons yield $p > 0.10$ for both courses, suggesting that subjects are well balanced across our treatment groups.

Within each course, we send our incentive (control) email to participants on April 20th, right before the 8th (6th) assignment posting for *Chinese Culture (Data Structure).* As mentioned in section 3.1, our control group receives an email that encourages them to complete their homework, while our treatment groups receive an additional paragraph in the email, describing how their homework activity is linked to a monetary incentive. Those offered an incentive have two weeks to complete their homework to qualify for the incentive scheme. For each of the three homework assignments selected for our intervention, we collect participants submission records and grades at the end of the two week period. After each collection,

[10] We are aware of the potential sample selection biased towards frequent or attentive users. Users who volunteered to participate in the experiment are possibly different than students who did not. We are therefore cautious in extending the estimated effects to more representative samples.

participants are immediately informed how much they have earned from the previous homework. On average, participants in *Chinese Culture* earn 61.21 rmb from the intervention stage and those in *Data Structure* earn 38.91 rmb.

Our intervention covers a span of three assignments for each course and ends on June 12th. After the completion of our intervention, we send participants a survey that asks for their responses to the experiment and the previous homework assignments. After we remove the incentive, there remain seven homework assignments for *Chinese Culture* and four homework assignments for *Data Structure*. After both courses close on July 1st, we send participants a final survey to collect their long-term response towards our experiment manipulation. To encourage participation in the post-experiment surveys, we pay 5 rmb for filling out each survey and also draw a 100 rmb prize among the respondents. Online Appendices C and D include our two post-experiment survey questionnaires.

4 Results

In this section, we first examine treatment effects on homework submission rates, homework grades, and lecture video viewing time during our intervention, and then examine long-term effects on the same set of outcome variables after incentives are removed. We exclude one participant from the *Data Structure* course and four from the *Chinese Culture* course as these participants dropped their respective courses before any monetary incentive occurred.[11] Altogether, in the following analyses, we have 324 subjects in the *Chinese Culture* course and 431 in the *Data Structure* course.

4.1 Treatment Effect on Assignment Submission

Figure 2 presents the rate of homework submission before and during the intervention for each course. This figure shows that the control group exhibits a significant decay in its submission rate over the course.[12] For example, the average submission rate for the first seven *Chinese Culture* homework assignments is 52%; this rate drops to 32% for assignments 8-10 ($p < 0.01$, 2-sided test of proportions). We find a similar significant decrease in assignment submission rates for our 1 rmb treatment group (from 44% to 30%, $p < 0.01$, 2-sided test of proportions). Interestingly, we find that the decrease in the submission rate for our 10

[11] The dropouts occurred after the randomization procedure. We test whether dropout correlates with treatment and find no significant effect.

[12] We do not find any significant difference between framing the incentive as gain versus loss and therefore pool the two framing treatments together. The only exception is that in *Chinese Culture*, a 1 rmb loss induces fewer submissions than a 1 rmb reward ($p < 0.01$, 2-sided test of proportions).

rmb group (42% to 38%) is insignificant ($p = 0.11$, 2-sided test of proportions), and that our 100 rmb group exhibits a submission rate increase from 49% to 50% ($p = 0.79$, 2-sided test of proportions). Thus, we conclude that both 10 and 100 rmb incentives are effective in maintaining users' motivation to complete homework assignments in the *Chinese Culture* course. There is no evidence that 1 rmb incentive maintains users engagement or crowds out their intrinsic motivation.

Examining the results for the *Data Structure* course, we find that the average assignment submission rate for first five assignments for the control group is 19%; this rate drops to 8% for assignments 6-8 and the change is statistically significant ($p < 0.01$, 2-sided test of proportions). For the 1 rmb group, we find that the submission rate decreases from 17% to 6%, and for the 10 rmb group it decreases from 19% to 8% (both changes are statistically significant, $p < 0.01$, 2-sided test of proportions). Only our 100 rmb treatment group shows an incentive effect, i.e., counteracting the decay, with an insignificant decrease from 14% to 13% ($p = 0.58$, 2-sided test of proportions).

For both courses, we find that a large incentive successfully maintains user engagement as measured by homework submission rates. Our results also unfold the difference in engagement across the two courses. *Data Structure* has a much lower baseline submission rate, and the 10 rmb incentive is sufficient to keep users engaged in *Chinese Culture* but not in *Data Structure*. One possible reason could be the higher cost required to complete assignments in the latter course. We observe that assignment questions for *Chinese Culture* mostly cover facts delivered in course lecture videos, while those for *Data Structure* require the user to master a method and algorithm.

We next examine the effect of a monetary incentive at the individual level. Examining within-user changes before and after our intervention, we classify our subjects into three types: users who decrease submissions, users who increase submissions, and users who do not submit homework either before or during the intervention. Figure 3 presents the share of each type of users. In the *Chinese Culture* course, we find that 52% of users in the control group reduce their submission rate for assignments 8-10 (intervention period), compared to 44% in the 1 rmb group, 32% in the 10 rmb group, and 28% in the 100 rmb group, with a significant difference for the control and 10 (100) groups ($p = 0.02$ and $p < 0.01$, respectively, 2-sided test of proportions). We further find that 26% of users in the control group increase their submission rate for assignments 8-10, compared to 37% in the 10 rmb group and 46% in the 100 rmb group, with a significant difference for the control and 100 rmb group ($p = 0.02$, 2-sided test of proportions). Of those who increase their submissions for assignments 8-10, only 2% (1%, 5%) of the control (10 rmb, 100rmb) group had not previously submitted assignments. Finally, we find no incentive effect for those users who never submit homework assignments for assignments 1-10.

We next examine individual user engagement behavior in the *Data Structure* course. In this course, we find a significant effect on assignment submission behavior for only the 100 rmb group. We find no effect of incentives on our non-submitter group.[13] Overall, our individual-level analysis suggests that the treatment effect (higher submission rate) reflects the maintenance of existing engagement levels rather than the motivation to begin submitting homework assignments.

Lastly, we supplement the above findings with a regression analysis. We apply a simple OLS estimation on the sample during the intervention period to estimate the treatment effect on homework submission. The econometric specification is as follows:

$$Y_{ijt} = \alpha + \sum_j \beta_j \times Treatment_j + X_i + \varepsilon_{ij} \tag{1}$$

where i indexes individual participants, j indexes treatment groups and t indexes time periods (weeks). Y_{ijt} is the submission record (1 if submitted, 0 otherwise) of participant i in treatment group j at week t. $Treatment_j$ are dummy variables for treatment groups. X_i includes participant characteristics, such as gender, age, education, job status and experience with online learning. Lastly, ε_{ijt} is the error term and is clustered at the individual user level. The coefficients of interests are β_j, which capture the difference in homework submission between the respective treatment group and the control. We fit equation (1) with linear probability models and present the results in Table 3.[14] The sample comprises of homework records *during* the intervention periods.

Overall, regression results confirm our graphical evidence. First, 100 rmb incentive significantly increases homework submission for both courses. As shown in Columns 1 and 4, a 100 rmb incentive raises the probability of submission by 20 percentage points in *Chinese Culture* and 6.9 percentage points in *Data Structure*. Meanwhile, a 10 rmb incentive has an effect only in *Chinese Culture*.[15] Small amount of incentive (1 rmb) does not show adverse effects, i.e., no evidence of crowding out intrinsic motivation. Second, we include the interaction between incentive amount and frame and present the results in Columns 2 and 4. These results show that the coefficients of the interaction terms are small and statistically insignificant, suggesting that framing the incentive as rewards or punishments does not have

[13]More specifically, 38% of the control group show engagement decay, a similar share as in the 1 and 10 rmb groups, while our 100 rmb shows only a 20% decay. The ratio of users who increase their homework submission is 5% in the control, 6% in the 1 rmb and 10 rmb treatments, and 12% in the 100 rmb treatment. The ratio of "start-up" submissions is 2% in the control, 1% in the 10 rmb and 5% in the 100 rmb group.

[14]Probit or logit models yield similar results.

[15]We cannot reject the hypothesis that the effects for the 10 rmb and 100 rmb groups are the same. However, the effect for the 1 rmb group is significantly different from that for the 100 rmb group ($p = 0.003$).

differential effects. Lastly, we include the interaction between treatment and a dummy variable indicating whether a participant's pre-intervention submission rate is above or below the sample mean. As shown in Columns 3 and 6, the coefficients are positive and significant, implying that our observed treatment effects are driven largely by sustained activity by regular users rather than an uptake in activity by inactive users.

4.2 Treatment Effects on Grades

In this section, we examine whether a monetary incentive impacts assignment performance, as measured by the grade received on an assignment. Figure 4 presents the unconditional mean of homework grades for our courses. A non-submission is recorded as receiving a grade of zero. Examining the effectiveness of incentives on assignment performance, we find similar results as in our homework assignment submission analysis. For example, in the *Chinese Culture* course, only the 10 rmb and 100 rmb groups exhibit a significant continuance of performance on assignments 8-10. In the *Data Structure* course, only the 100 rmb group exhibits a significant continuance of performance on assignments 6-8.

Figure 5 presents the conditional mean of homework grades for the courses. Here, a non-submission is excluded from the sample. Conditional on assignment completion, we find that the average grades during the intervention period are higher for our treatment groups than for the control group, although this difference is statistically insignificant. However, it is possible that our incentive biases the sub-set of students who are motivated to submit their homework assignment. If so, then it is possible that a higher grade reflects a bias due to the self-selection of higher-performing students. To address this possibility, we first test if our treatments motivate different types of users to submit homework assignments. Appendix Tables A1 and A2 report the estimations for our treatment effects on user demographic characteristics and baseline performance. These statistics show no strong evidence of differential user composition. Second, we formally address any potential bias using Lee Bounds in a regression analysis below.

As a baseline, we fit equation (1) to estimate the treatment effects on assignment grades during intervention period. Table 4 reports the OLS estimates for the *Chinese Culture* (panel A) and *Data Structure* (panel B). The dependent variable is the grade in terms of the correction rate (0 to 1), and the sample includes participants' homework grades *during* the weeks when incentive is offered. Therefore, coefficients on the treatment indicators can be interpreted as the differences in grades between the respective treatment group and the control during our intervention. Panel A, Column 1 show that, for students in *Chinese Culture*, a 1 rmb incentive has a marginally significant effect on the unconditional grade, while 10 rmb and 100 rmb incentives both significantly improve homework grades relative to the control group. Panel A, Column 2 show that the treatment has positive and significant effect on conditional grades, though the magnitude is smaller than in Column 1.

Continuing with Table 4, Columns 3 and 4 report the upper and lower bounds, respectively, of the treatment effects on conditional grades, using the method developed by Lee (2009).[16] The results show that all coefficients are positive, although the estimated lower bound for the 100 rmb group is marginally insignificant.

The results in Panel B, Columns 1 and 2 for the *Data Structure* course show a significant positive effect on unconditional and conditional grades for only the 100 rmb group. The results in Columns 3 and 4 show that the estimated upper and lower bounds for the 100 rmb treatment are always positive and very precisely estimated. Taken together, we find positive effects on assignment grades, which are unlikely to be entirely driven by sample selection.

In our final set of analyses, we examine the effect of a monetary incentive on the amount of time a user spends watching the course lecture videos. Here, we conjecture that spending more time on the videos could be a mechanism through which treated participants gain higher grades. For both courses, lectures are delivered in the form of videos. We collect data on participants daily video activity (e.g., when they start a video, pause or resume the video, or spend idle time with the video open), and apply a machine-learning algorithm to capture their effective viewing time. This measure allows us to obtain specific information about learning activity and effort, which is difficult to observe or quantify in a traditional offline classroom environment.

The results in Table 5 show that the 100 rmb treatment increases weekly video time by 5 to 6 percent for both courses (Columns 1 and 4); the results further show that these effects do not differ with incentive framing and that they reflect continued activity by active users. Interestingly, in the *Chinese Culture* course, we find that that neither the 1 rmb nor 10 rmb treatment motivates participants to increase their course video viewing time. One explanation for this finding is that learners are using a non-linear navigation strategy. As documented in Guo and Reinecke (2014), successful users (certificate earners) skip 22% of the content in a course and frequently jump backward to earlier lecture sequences to gain specific information. This non-linear navigation implies that better performance does not necessarily come from longer hours spent viewing course materials.

[16] Specifically, in our context, we consider the case in which the 20% increase in assignment submissions in the 100 rmb group arises from the least capable students, i.e., the largest downward bias. Then, to construct a balanced sample, we drop these bottom 20% grades in the 100 rmb group so that the resulting estimates constitute the upper bound of the true effect. Similarly, if we assume that the increased submissions are made by the most capable students, then we exclude the top 20% grades of the 100 rmb group so that the estimate from the refined and balanced sample represents the lower bound of the true effect.

4.3 Post-intervention and Spillover Effects

Our results indicate that providing a monetary incentive can improve both engagement and performance in an online learning environment. We next examine whether this effect persists after the incentive is removed. A number of studies have shown that short-term incentives may fail in the long run (Gneezy and List 2006; Meier 2007).[17] By contrast, Charness and Gneezy (2009) find more promising results that a monetary incentive can instill long-term exercise habits. Moreover, there are concerns that offering students financial incentives may weaken or crowd out their intrinsic motivation on similar, subsequent tasks (Gneezy et al. 2011). If so, the removal of incentives may decrease student effort and performance.

We examine learning behavior and course performance on the remaining assignments in each course after we stop the intervention. There remains about 2 months until the courses end, during which participants in *Chinese Culture* are assigned seven more homework assignments, and those in *Data Structure* 4 more assignments. Table 6 presents the results for our treatment effects on the submission rate, unconditional grades, and conditional grades for these post-intervention assignments. For the *Chinese Culture* course, we find that the 10 rmb and 100 rmb groups continue to increase their homework submission rates and assignment performance. For instance, participants who were offered a 10 rmb incentive are still 12.5 percentage points more likely to submit their homework, and their grades are 0.146 points higher. While the magnitude is lower than during intervention, we cannot reject the hypothesis that the treatment effects are equal between during- and post-intervention phases, i.e., no significant decay.

For the *Data Structure* course, the long-term effects of the 100 rmb treatment remains positive, although statistically insignificant. We also find no significant difference between the during- and post-intervention effects. Overall, our post-intervention results suggest that the positive effect of a monetary incentive does not decay once the incentive is over. This finding may indicate that our incentive makes students more aware of the marginal return of submitting assignments and thus they are more likely to continue doing so.

The finding of a persistent treatment effect after the incentives are removed is important for several reasons. First, we demonstrate a long-run effect of monetary incentives on learning, which is promising for practitioners and policy makers because most incentive programs in education are only temporary and are restricted to certain tasks or tests. Second, treated students continue to engage more and perform better, indicating habit formation or certain

[17] In particular, Gneezy and List (2006) shows that the positive reciprocity that occurs from the receipt of a gift persists for only a few hours. Meier (2007) finds that the success of a matching mechanism in the area of charity donations does not carry to the post-experiment periods, generating a negative net effect on the participation rate.

type of learning about the online course experience. Third, there is no evidence of decrease in performance, suggesting that offering incentives, either small or large amount, does not necessarily crowd out students' intrinsic motivation to learn. Also, the fact that removal of incentives does not reverse student engagement towards the pre-intervention level suggests that the baseline low engagement is sub-optimal for the students.

Lastly, we investigate the spillover effect of a monetary incentive to other courses during the same semester, as well as those during the subsequent semester. Since 89% of the subjects in our experiment are enrolled in multiple courses, we are interested in whether our observed increase in engagement extends to other courses (a positive spillover) or is achieved at the expense of time and effort spent in other courses (a negative spillover). Using data on our subjects' video viewing time and assignment grades in other courses, we find an overall positive spillover effect: treated participants, especially those in the 100 rmb group, spend more time watching course videos and achieve higher homework grades in their other non-rewarded courses (Table 7). Furthermore, we find that treated participants in the *Chinese Culture* course still outperform the control group in their subsequent semester courses, as measured by completing a course and obtaining a certificate (Table 8).[18]

The findings on the spillover effects lend further support to a long-run improvement on student engagement. If offering monetary incentive in one course improves student engagement at the expense of lowering effort in other courses, the overall effect is ambiguous and such a policy is not a scalable solution. However, we find that the participants do not compromise their engagement and performance in other courses, nor stop exerting effort once the incentive is removed. Our intervention appears to have helped them learn about online learning process, or about disciplining themselves, which, in turn, shift their learning behavior towards a more persistent and sustainable pattern.

5 Discussion

To the best of our knowledge, ours is one of the pioneering attempts to evaluate the effectiveness of a monetary incentive in online learning. Overall, our findings suggest that providing a monetary incentive can help improve user engagement and raise the return of MOOCs. Our findings have both practical and academic implications. On a practical level, the platform where we conduct our experiment, XuetangX, has adopted several initiatives to encourage learning based on our findings. For instance, they plan to launch a certificate discount for those students who exhibit good performance in their courses, and to develop a

[18] The criteria of earning a certificate vary by courses, but usually involve students' homework, project and exam performance.

scholarship program to motivate learning.[19]

On a broader level, our findings can also be used by public programs that promote online learning in helping them to design a platform to better utilize the resources invested by teachers, universities, and public sectors in online courses. In designing online public courses, it is useful to understand what groups may be more responsive to a monetary incentive. Examining our results by gender as well as by access to offline education resources, we find that females show a greater effect of incentive on learning behaviors and outcomes, as do participants with limited geographic access to offline classrooms (see Tables 9 and 10, respectively).[20] For our geographic data, we use subjects' IP addresses and control for the GDP per capita of the region to ensure our differences are not driven by local economic conditions. These heterogeneous effects imply that offering a monetary incentive may help reduce educational disparity.

In addition to suggesting how incentives may be used to broaden educational access, our findings can also help course designers in determining the appropriate type of incentive for a particular course. For example, we find that a medium-level incentive works for the *Chinese Culture* course but not for the *Data Structure* course, possibly due to differences in difficulty level and effort required. It is also possible that course designers may use complementary—possibly non-financial—incentives to engage users.[21] In fact, the *Data Structure* course includes a multi-stage programming tournament throughout the course, where winners can access materials (programming projects) that are available exclusively for the Tsinghua computer science department.

On the academic side, our findings can be used as the basis for future research. For example, we timed our intervention to take place in the middle of the respective courses for both empirical and logistic reasons. However, it would be interesting to see what effect would occur if the intervention were provided at the beginning of each course instead. With big data on user activity, we could even possibly predict the "hazard rates" for users at any given moment, and design customized instruments to keep them engaged and improve learning outcomes. Online education platforms are valuable testbeds for putting behavioral economics principles into practice. The variety of course settings, scale, student backgrounds, and available rich activity logs allow researchers to conduct a number of experiments and test the generalizability of their results.

[19] An interview with the CTO of XuetangX, Mr. Jian Guan was conducted on May 19, 2017.

[20] For heterogeneity by gender and offline educational resources, we conduct the analysis with the *Chinese Culture* course because of its more diverse student composition. *Data Structure*, for instance, has too few female students to test the gender difference.

[21] For example, Jalava et al. (2015) examine the effect of non-financial incentives among primary school students and find improved test performance when employing rank-based grading or offering students a symbolic reward.

Lastly, our findings have economic implications, especially in addressing the issue of the return from online learning courses in the labor market. Deming et al. (2016) show that job applicants with degrees from online institutions are less likely to receive a callback in the job recruitment process than those with degrees from traditional offline institutions. This finding may reflect current issues with the efficacy and quality of online learning, which could be improved by strengthening online students' incentives. If online education were perceived as being as effective as in-person instruction from a traditional school, this could have implications for traditional schools in how they shape their future offerings, leading to a profound change in how higher education is delivered. Overall, our study provides insight into how to increase student engagement and performance in online courses. The implications of our findings raise a number of intriguing avenues for future research.

References

Acemoglu, Daron, David Laibson, and John A List, "Equalizing Superstars: The Internet and the Democratization of Education," *American Economic Reviews Papers and Proceedings*, 2014, *104.*

Andreoni, James, "Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments," *The Quarterly Journal of Economics*, 1995, *110* (1), 1–21.

Angrist, Joshua and Victor Lavy, "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *The 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," *The 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," *The American Economic Review*, 2002, *92* (5), 1535–1558.

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

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

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.

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

Chen, Jingnan, Miguel A. Fonseca, and Shaun B. Grimshaw, "Using Norms and Monetary Incentives to Change Behavior: A Field Experiment," 2018. Working Paper.

Chen, Yan and Joseph Konstan, "Online field experiments: a selective survey of methods," *Journal of the Economic Science Association*, 2015, *1* (1), 29–42.

Clark, Damon, David Gill, Victoria Prowse, and Mark Rush, "Using goals to motivate college students: Theory and evidence from field experiments," Technical Report, National Bureau of Economic Research 2017.

Deming, David J, Claudia Goldin, Lawrence F Katz, and Noam Yuchtman, "Can Online Learning Bend the Higher Education Cost Curve?," *The American Economic Review Papers and Proceedings*, 2015, *105* (5), 496–501.

_ , Noam Yuchtman, Amira Abulafi, Claudia Goldin, and Lawrence F Katz, "The Value of Postsecondary Credentials in the Labor Market: An Experimental Study," *American Economic Review*, 2016, *106* (3), 778–806.

Feng, Wenzheng, Jie Tang, and Tracy Xiao Liu, "Dropout Analysis and Prediction for Large Scale Users in Moocs," 2018. Working Paper.

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

_ , Steven D Levitt, John List, and Sally Sadoff, "Enhancing the Efficacy of Teacher Incentives Through Loss Aversion: A Field Experiment," 2012. Working Paper.

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.

— **, 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.

Guo, Philip J. and Katharina Reinecke, "Demographic Differences in How Students Navigate Through MOOCs," in "Proceedings of the First ACM Conference on Learning @ Scale Conference" L@S '14 ACM New York, NY, USA 2014, pp. 21–30.

Hansen, John D and Justin Reich, "Democratizing Education? Examining Access and Usage Patterns in Massive Open Online Courses," *Science*, 2015, *350* (6265), 1245–1248.

Hong, Fuhai, Tanjim Hossain, and John A List, "Framing Manipulations in Contests: A Natural Field Experiment," *Journal of Economic Behavior & Organization*, 2015, *118*, 372–382.

Hossain, Tanjim and John A List, "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations," *Management Science*, 2012, *58* (12), 2151–2167.

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol, "Habit formation and rational addiction: A field experiment in handwashing," 2017.

Jalava, Nina, Juanna Schrøter Joensen, and Elin Pellas, "Grades and rank: Impacts of non-financial incentives on test performance," *Journal of Economic Behavior & Organization*, 2015, *115*, 161–196.

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.

Kizilcec, René, Chris Piech, and Emily Schneider, "Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses," in "Proceedings of the Third International Conference on Learning Analytics and Knowledge" ACM 2013, pp. 170–179.

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.

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.

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.

Seaton, Daniel, Isaach Chuang, Piotr Mitros, David Pritchard et al., "Who Does What in a Massive Open Online Course?," *Communications of the ACM*, 2014, *57* (4), 58–65.

Zhang, Dennis J, Gad Allon, and Jan A Van Mieghem, "Does Social Interaction Improve Learning Outcomes? Evidence From Field Experiments On Massive Open Online Courses," *Manufacturing & Service Operations Management*, 2017, *19* (3), 347–367.

Online Appendices

A Sample Experiment Email: Large Incentive Treatments for Data Structure

Dear MOOCer

Thanks for participating in our study. We will give you 5 rmb to collect your responses in two more surveys which will be distributed later this semester. Additionally, you will have a chance to win a 100 rmb big reward.

Data structure has updated to the 6th homework. If you want to get your certificate, you should finish your homework on time and try your best to get high grades!

For the next 3 assignments, you will receive an 100 rmb reward if you answer 8 out of 10 questions correctly in an assignment (negative framing: you will receive a one-time bonus of rmb 300. However, for every assignment you fail to answer 8 out of 10 questions correctly, we will deduct you 100 rmb).

The reward is accumulative.

For example, if you answer correctly in one assignment, you gain 100 rmb. If you answer correctly in two assignments, you gain 200 rmb. If you answer correctly in all assignments, you gain 300 rmb (negative framing: if you fail one assignment, you lose 100 rmb. If you fail two assignments, you lose 200 rmb. If you fail all three assignments, you lose 300 rmb.).

Additionally, for you to get the reward, you also need to submit all assignments within two weeks after we post them.

This payment will be made in the end of the course in the format of top-up card.

B Sample Recruitment Email

Dear fellow students in Data Structure,

MOOC is an education innovation which has drawn attention around the globe, and MOOCers' learning behaviors are of great value for research.

Our research team of Tsinghua, along with XuetangX.com, is conducting a study on MOOCers' learning behaviors, which is the first in China. We want to invite you to join us, making history with us! Once participating in this study, you can earn money while taking the MOOC course. Slots are limited. First come, first served!

About our study

This is a study on MOOCers' learning habits, especially assignment completion. Participants are required to complete a few questionnaires and follow some other instructions of our study.

Rules for participants

Participants must:

a) finish assignments of MOOC courses independently;

b) not reveal any information related to this study to others.

Payment

Participants are expected to get 5 to 300 yuan. Each participants exact amount of reward depends on his/her behavior in the study. Participants will be paid the equivalent value of top-up cards for mobile phones after the end of the course.

Research results

After the end of our project, we will send the written research results to all participants via email.

Please read the Project Agreement and sign up.

C Pre-Experiment Survey

We want you to join the most cutting-edge study on MOOCers' learning behaviors!

To sign up for this study, you may need one or two minutes to finish this questionnaire.

Seats are limited. First come, first served! Applicants admitted to our study will be informed via email or private message in MOOC in two weeks.

–XuetangX.com & MOOC Research Team of Tsinghua University

Introduce myself in 20 seconds:

I am a ( ) girl ( ) boy.

I was born in (which year).

I live in (which city) (which province), China.

My education level is (), and current employment status is ().

I am (description of me)

(Very curious/ a born genius/ Never focus on trivial things/ A plan maker/ Always in the limelight/Always active)

About the Data Structure Course (Chinese Culture)

1. I take this course because:

   (a) I am interested in the content of this course;

   (b) It is helpful for my current work;

   (c) My friend(s) recommended it;

   (d) I want to get a MOOC certificate.

   (e) Others

2. In terms of the content of this course:

   (a) I have never learnt it before and have no idea at all;

   (b) I am currently learning related courses offline;

   (c) I have learnt a little bit before;

   (d) I am an expert.

About XuetangX.com

1. I think learning MOOC in XuetangX is:

(a) Very unpleasant
(b) Rather unpleasant
(c) Rather pleasant
(d) Very pleasant

2. I am studying together with some other people. Among them, the three people that have the most intimate relationship with me are (you can leave it blank if you are learning alone):

   (His/her username in XuetangX) We are Classmates/Friends/Boyfriend & Girlfriend/-Couple/Relatives

3. Sometimes, I cannot persist in learning MOOC courses, because:

   (a) It never happens to me! I am a persistent learner!
   (b) I have a heavy load of schoolwork;
   (c) I need to date very often;
   (d) I don't want to bother to turn on the computer;
   (e) No one reminds me;
   (f) Others

4. In XuetangX.com, I hope to:

   (a) Spend more time learning in XuetangX.com;
   (b) Maintain the status quo;
   (c) Spend less time learning in XuetangX.com.

5. When I am pursuing my goals, I am:

   (a) Very persistent;
   (b) Rather persistent;
   (c) likely to quit;
   (d) Very likely to quit.

6. In terms of competitiveness (my eagerness to win),

   (a) I am highly competitive;

(b) I am rather competitive;

(c) I am not very competitive;

(d) I am not competitive at all.

D Post-Experiment Survey I

1. Does the assignment grade of this class matter to you? 1-4 likert scale
2. Does getting the certificate of this class matter to you? 1-4 likert scale
3. In your opinion, to get the certificate, how much should you get for your assignment grades? (0 to 100)

   The next questions are for treatments only.
4. In the past three assignments with additional reward, if there is a week you did not receive this reward, what will you do next?
   (a) This does not apply to me. I receive rewards every time.
   (b) Increase my effort on assignments and try to get the reward next time.
   (c) Keep the current effort level.
   (d) I do not care much about the reward.
   (e) Not sure.
5. In the past three assignments with additional reward, if there is a week you receive this reward, what will you do next?
   (a) This does not apply to me. I never receive rewards.
   (b) Increase my effort on assignments and try to get the reward next time.
   (c) Keep the current effort level.
   (d) I do not care much about the reward.
   (e) Not sure.
6. How important is the assignment reward to you? 1-4 likert scale
7. When the goal is to get 80% questions in assignment correct and additional incentive is provided for achieving this goal, how important is it to you? 1-4 likert scale
8. When the goal is to get 80% questions in assignment correct and no additional incentive is provided for achieving this goal, how important is it to you?
   (a) Very important. I always want to achieve the goal regardless of additional reward.
   (b) Relatively important
   (c) Relatively not important
   (d) Not important at all. I do the assignments for getting the additional reward.

E Post-Experiment Survey II

1. After we stop providing additional reward for assignments (control: After we stop sending you reminder email for assignments), do you think it is necessary to work hard on assignments? (1-4 likert scale)

2. After we stop providing additional reward for assignments (control: After we stop sending you reminder email for assignments), do you think it is necessary to get 80% of questions correctly? (1-4 likert scale)

3. Do you know other people receive different amount of rewards (control: Do you know other people receive additional rewards for doing assignments)?

4. If you know that others receive higher reward than you (control: If you know that others receive additional rewards for doing assignments), what will you do?

   (a) Will work harder

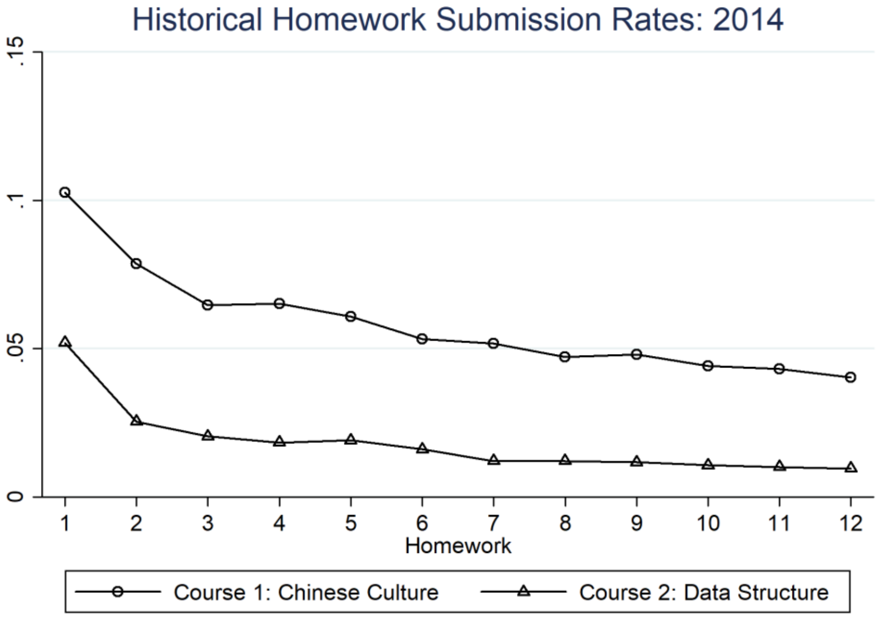
   (b) Will slack off

   (c) I do not care

   (d) Other

5. Do you think participating in this project help you study this class? 1-4 likert scale

Figure 1: Historical Homework Records



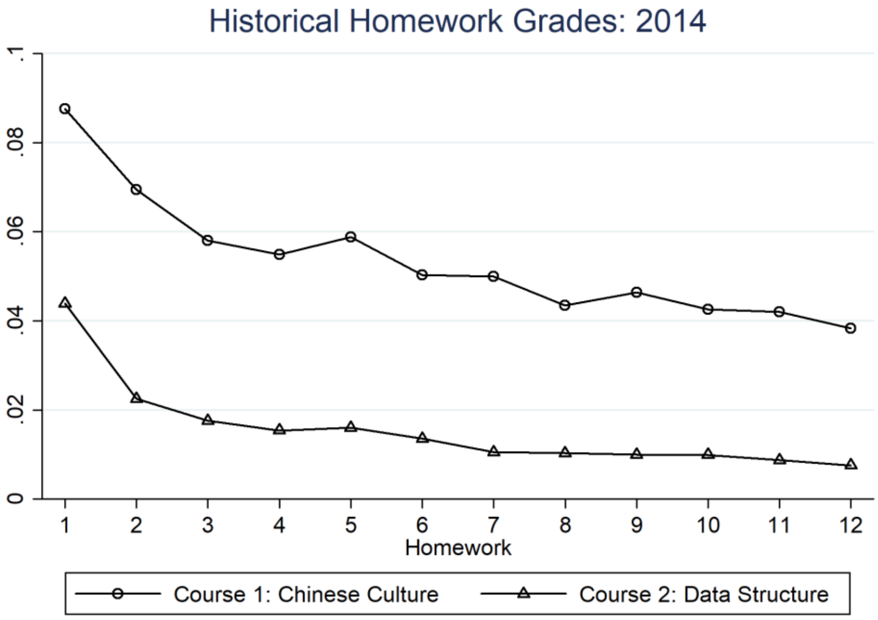
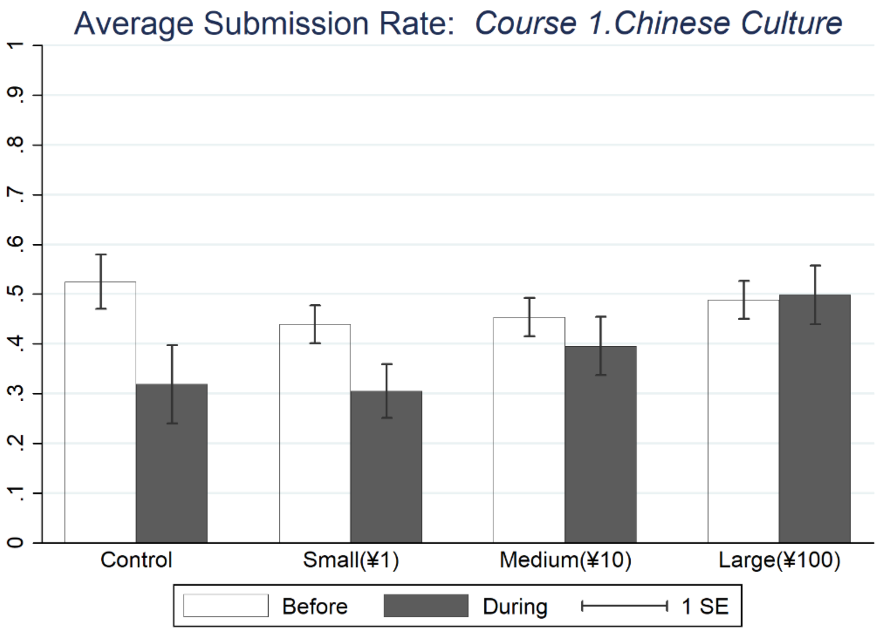


Figure 2: Average Homework Submission Rate before and during Intervention



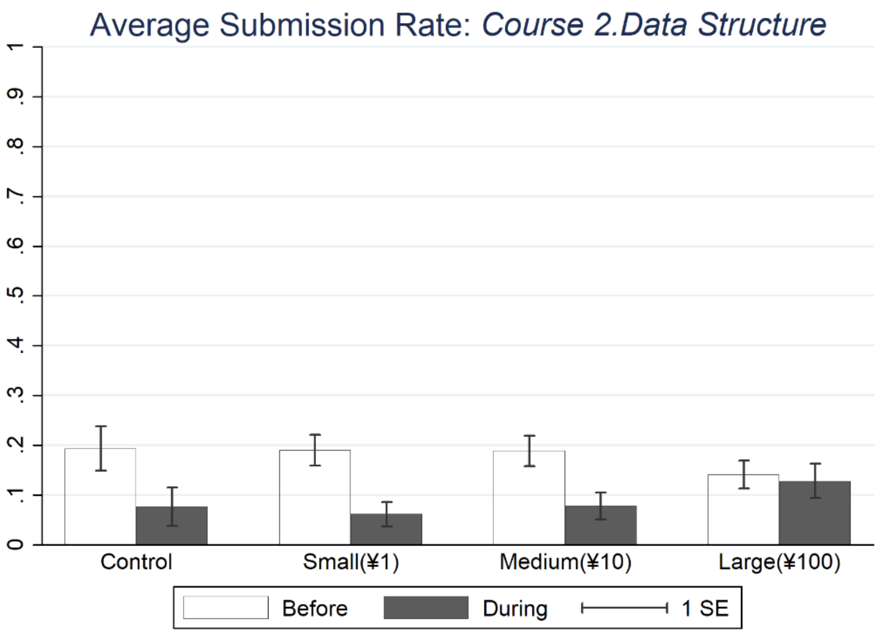
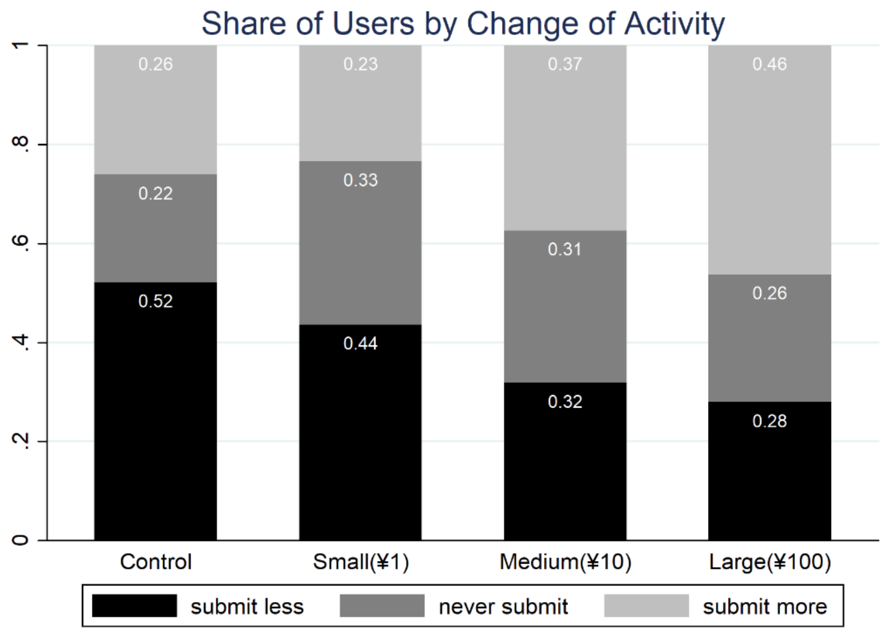
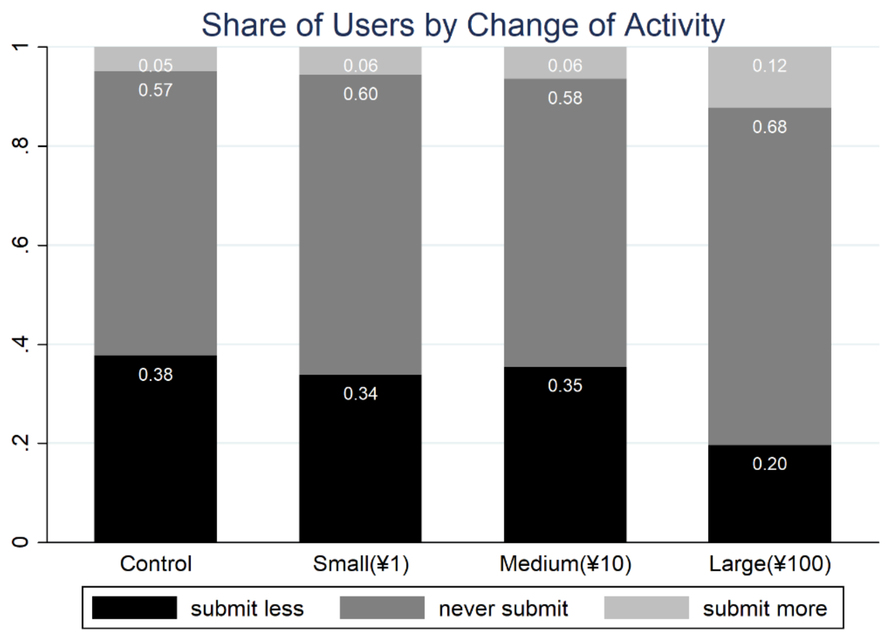


Figure 3: Share of Participants by Changes of Submission Activity

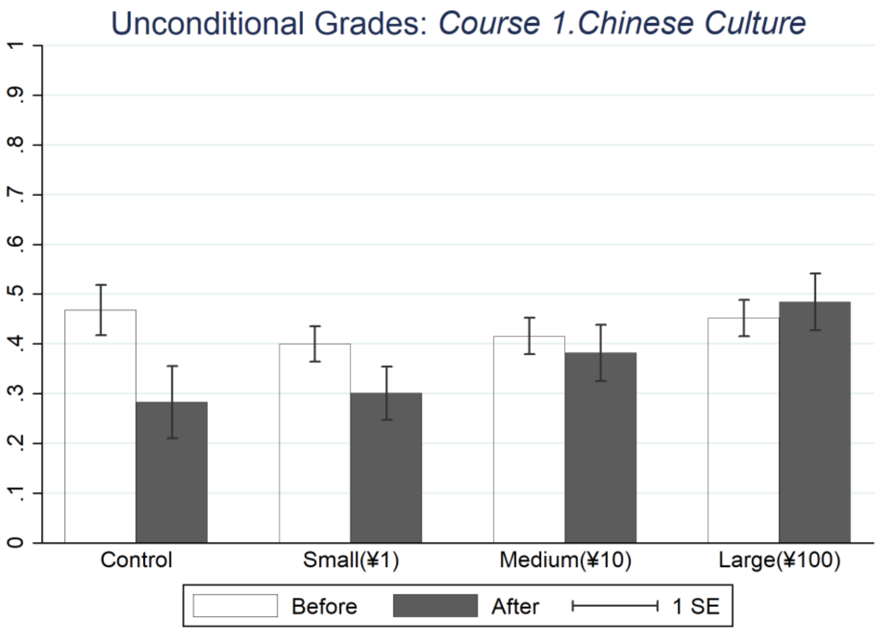


(a) Course 1: Chinese Culture



(b) Course 2: Data Structure

Figure 4: Unconditional Means of Homework Grades before and during Intervention



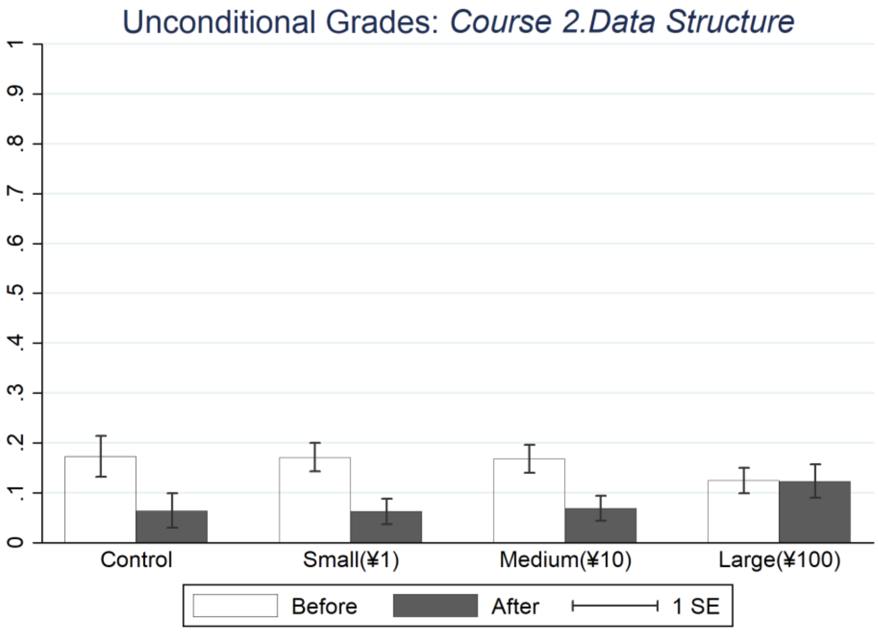
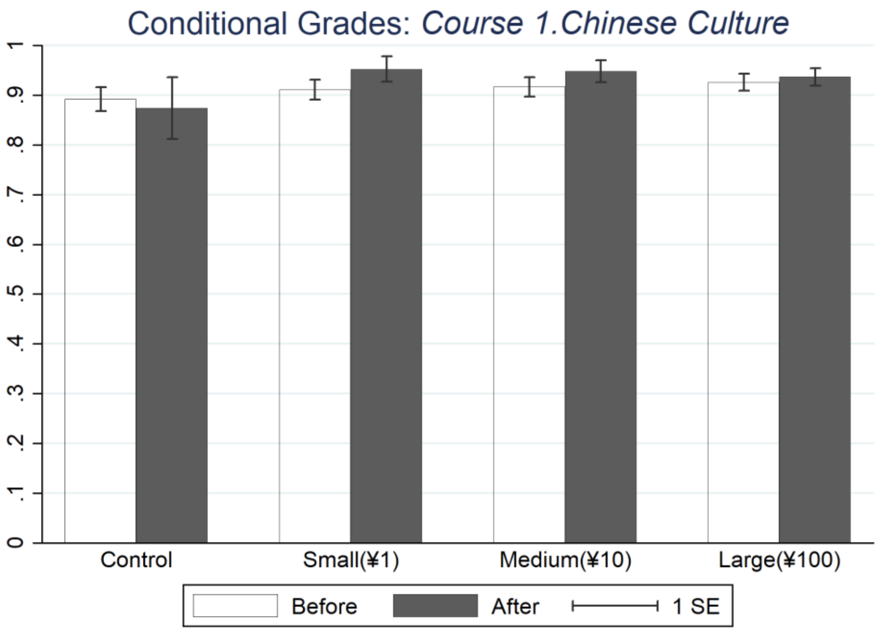


Figure 5: Homework Grades Conditional on Submission, before and during Intervention



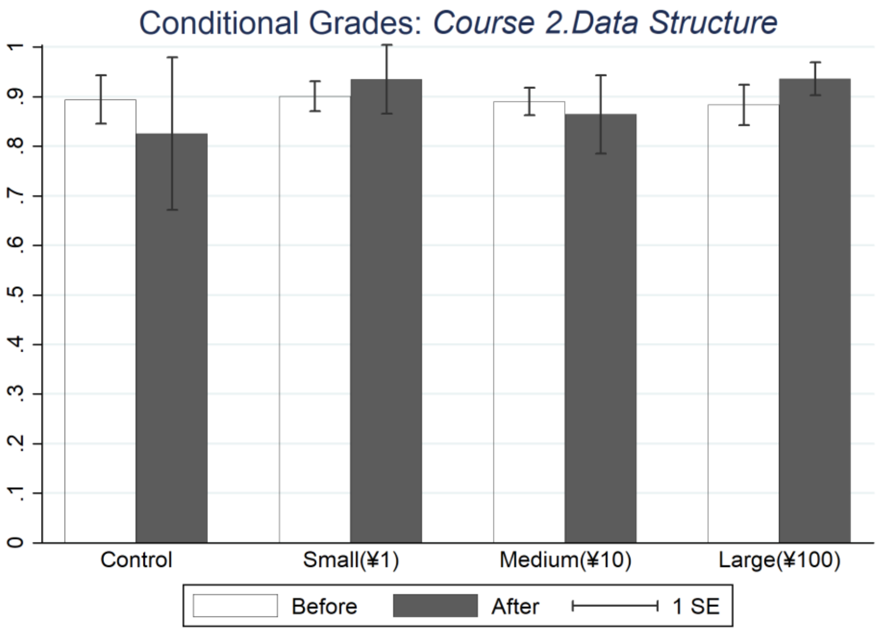


Table 1: Experiment Timeline and Data Collected

Date	Task	Data collected
April 6	Recruit email	
April 6-14	Sign up Pre-experiment survey	Demographics Baseline performance
April 20	Incentive announcement	Homework submission, grades & video logs
June 16	Post-experiment survey 1	Feedback on intervention
July 1	Post-experiment survey 2	Feedback on post-intervention
August 13	Payment	

Table 2: Summary Statistics

Sample	*Chinese Culture* (1)	*Data Structure* (2)	*All* (3)	*All* (4)
	Mean	Mean	min	max
Male	0.377	0.807	0	1
	(0.485)	(0.395)		
Age	27.11	23.72	15	59
	(8.060)	(5.119)		
Education				
Middle school and below	0.140	0.0758	0	1
	(0.347)	(0.265)		
Senior high/technician school	0.648	0.697	0	1
	(0.478)	(0.460)		
College and above	0.213	0.227	0	1
	(0.410)	(0.420)		
Employment Status				
Student	0.506	0.716	0	1
	(0.501)	(0.451)		
Unemployed	0.0528	0.0512	0	1
	(0.224)	(0.221)		
Employed	0.438	0.230	0	1
	(0.497)	(0.421)		
Retired	0.00311	0.00233	0	1
	(0.0557)	(0.0482)		
Subject and MOOC background				
Experience with the subject	1.820	2.248	1	4
	(0.938)	(0.900)		
Friends taking the same course	0.0926	0.0626	0	1
	(0.290)	(0.243)		
Time commitment	0.475	0.566	1	3
	(0.564)	(0.541)		
Retake this course	0.194	0.367	0	1
	(0.396)	(0.482)		
Number of courses taken	1.981	2.248	0	28
	(3.345)	(3.638)		
Number of certificates obtained	0.395	0.146	0	5
	(0.828)	(0.523)		
Activity before experiment				
Homework score	0.598	0.330	0	1
	(0.431)	(0.436)		
Homework submission rate	0.470	0.176	0	1
	(0.412)	(0.286)		
Weekly video hours	2.407	1.523	0	13.27
	(3.132)	(2.622)		
Observations	324	431		

Note. Columns (1) and (2) report the mean and standard deviations of the main variables for participants who enrolled in each of the respective courses and signed up for our experiment. Columns (3) and (4) pool participants from the two courses and report the minimum and maximum of the variable values.

Table 3: Homework Submission Rate during Intervention

	Outcome: Whether homework assignment has been submitted on time					
	Course 1: Chinese Culture			Course 2: Data Structure		
	(1)	(2)	(3)	(4)	(5)	(6)
¥1	0.063	0.098	-0.055	-0.022	-0.045	-0.033
	(0.066)	(0.075)	(0.085)	(0wu.033)	(0.040)	(0.030)
¥10	0.134**	0.142**	-0.028	-0.002	0.009	-0.034
	(0.065)	(0.069)	(0.084)	(0.030)	(0.035)	(0.032)
¥100	0.201***	0.232***	0.056	0.069**	0.062*	-0.010
	(0.063)	(0.069)	(0.089)	(0.033)	(0.036)	(0.035)
¥1 x punish		-0.074			0.047	
		(0.062)			(0.039)	
¥100 x punish		-0.019			-0.021	
		(0.065)			(0.030)	
¥100 x punish		-0.065			0.016	
		(0.060)			(0.043)	
¥1 x active			0.192			0.025
			(0.128)			(0.067)
¥10 x active			0.295**			0.075
			(0.127)			(0.062)
¥100 x active			0.273**			0.255***
			(0.127)			(0.072)
Observations	891	891	891	1,263	1,263	1,263
R-squared	0.53	0.53	0.54	0.36	0.36	0.42

Note. The sample includes homework submission records during the intervention period where participants in treatments are rewarded with a monetary incentive. The unit of observation is participant*homework. All specifications include user controls (as reported in Table 2), including gender, age, education, employment status, course and MOOC background, and baseline activity before the experiment. Robust standard errors are clustered at the participant level and are shown in parentheses. *** significant at the 1%, **5%, and *10% level.

Table 4: Treatment Effects on Homework Grades

	Unconditional grade	Grade conditional on submission	Upper bound	Lower bound
	(1)	(2)	(3)	(4)
Panel A: Chinese Culture				
¥1	0.105*	0.088**	0.090**	0.088**
	(0.063)	(0.041)	(0.039)	(0.041)
¥10	0.162**	0.093**	0.126***	0.089**
	(0.063)	(0.041)	(0.038)	(0.042)
¥100	0.220***	0.078**	0.120***	0.064
	(0.060)	(0.039)	(0.036)	(0.040)
Observations	871	351	315	314
R-squared	0.51	0.06	0.15	0.06
Panel B: Data Structure				
¥1	-0.007	0.101	0.120	0.098
	(0.033)	(0.094)	(0.092)	(0.094)
¥10	0.000	0.127	0.124	0.125
	(0.029)	(0.087)	(0.087)	(0.088)
¥100	0.077**	0.161**	0.183***	0.154**
	(0.032)	(0.064)	(0.064)	(0.066)
Observations	1,214	110	107	107
R-squared	0.35	0.33	0.37	0.32

Note. The sample includes homework grades during the intervention period where participants in treatments are rewarded with a monetary incentive. The unit of observation is participant*homework. Column (1) uses the unconditional grade as the outcome variable, i.e., equals zero in the case of no submission. Columns (2) to (4) use the conditional grade, i.e., grade is missing in the case of no submission. Columns (3) and (4) report the upper and lower bounds of treatment effects using Lee bounds (Lee 2009). All specifications include user controls (as reported in Table 2), including gender, age, education, employment status, course and MOOC background, and baseline activity before the experiment. Robust standard errors are clustered at the participant level and are shown in parentheses. *** significant at the 1%, **5%, and *10% level.

Table 5. Treatment Effects on Video Hours

	Outcome: ln (weekly video hours)					
	Course 1: Chinese Culture			Course 2: Data Structure		
	(1)	(2)	(3)	(4)	(5)	(6)
¥1	-0.005	0.005	-0.052	-0.007	-0.019	-0.007
	(0.047)	(0.052)	(0.058)	(0.035)	(0.044)	(0.044)
¥10	0.001	-0.003	-0.068	0.008	-0.004	-0.009
	(0.046)	(0.051)	(0.056)	(0.035)	(0.044)	(0.045)
¥100	0.050	0.060	-0.045	0.060	0.096**	0.012
	(0.046)	(0.050)	(0.060)	(0.038)	(0.045)	(0.048)
¥1 x punish		-0.021			0.025	
		(0.042)			(0.039)	
¥100 x punish		0.008			0.024	
		(0.042)			(0.035)	
¥100 x punish		-0.021			-0.074	
		(0.040)			(0.049)	
¥1 x active			0.070			0.002
			(0.090)			(0.067)
¥10 x active			0.117			0.044
			(0.088)			(0.067)
¥100 x active			0.160*			0.151*
			(0.087)			(0.084)
Observations	1,188	1,188	1,188	1,263	1,263	1,263
R-squared	0.35	0.35	0.36	0.25	0.26	0.26

Note. The sample includes video viewing records during the intervention period where participants in treatments are rewarded with a monetary incentive. The unit of observation is participant*week. All specifications include user controls (as reported in Table 2), including gender, age, education, employment status, course and MOOC background, and baseline activity before the experiment. Robust standard errors are clustered at the participant level and are shown in parentheses. *** significant at the 1%, **5%, and *10% level.

Table 6. Treatment Effects after Incentives are Removed

	Course 1: Chinese Culture			Course 2: Data Structure		
	Submission	Unconditional grade	Grade conditional on submission	Submission	Unconditional grade	Grade conditional on submission
	(1)	(2)	(3)	(4)	(5)	(6)
¥1	0.062	0.086	0.016	0.014	0.019	0.005
	(0.062)	(0.058)	(0.028)	(0.034)	(0.032)	(0.043)
¥10	0.125**	0.146**	0.047*	0.017	0.024	0.002
	(0.062)	(0.060)	(0.027)	(0.033)	(0.031)	(0.060)
¥100	0.130**	0.136**	0.031	0.046	0.050*	0.032
	(0.061)	(0.059)	(0.025)	(0.032)	(0.030)	(0.051)
Observations	2,079	1,970	685	1,684	1,584	121
R-squared	0.46	0.46	0.05	0.27	0.28	0.20

Note. The sample includes homework submission records and grades after the intervention period where monetary incentives are removed. The unit of observation is participant*homework. All specifications include user controls (as reported in Table 2), including gender, age, education, employment status, course and MOOC background, and baseline activity before the experiment. Robust standard errors are clustered at the participant level and are shown in parentheses. *** significant at the 1%, **5%, and *10% level.

Table 7: Spillover to Other Courses during the Same Semester

	Participants in Chinese Culture			*Participants in Data Structure*		
	Video hour on other courses during intervention	Video hour on other courses after intervention	Grade of other courses	Video hour on other courses during intervention	Video hour on other courses after intervention	Grades of other courses
	(1)	(2)	(3)	(4)	(5)	(6)
¥1	0.045	0.057	0.003	0.056	-0.015	-0.004
	(0.045)	(0.039)	(0.038)	(0.044)	(0.042)	(0.022)
¥10	0.043	0.075	-0.000	0.033	-0.047	0.005
	(0.049)	(0.046)	(0.039)	(0.038)	(0.037)	(0.018)
¥100	0.089*	0.082*	0.045	0.072*	-0.020	0.042*
	(0.052)	(0.045)	(0.041)	(0.041)	(0.036)	(0.022)
Observations	891	2,079	262	1,263	1,684	377
R-squared	0.16	0.13	0.33	0.10	0.06	0.20

Note. The sample includes participants' video activity and performance in other courses they enrolled in during the same semester. All specifications include user controls (as reported in Table 2), including gender, age, education, employment status, course and MOOC background, and baseline activity before the experiment. Robust standard errors are clustered at the participant level and are shown in parentheses. Columns (3) and (6) use average unconditional grades as the dependent variable and further control for the number of courses in which the participant enrolled. *** significant at the 1%, **5%, the *10% level.

Table 8: Spillover to Subsequent Semester

	Participants in Chinese Culture		*Participants in Data Structure*	
	Number of enrolled courses	Certification rate	Number of enrolled courses	Certification rate
	(1)	(2)	(3)	(4)
¥1	-0.470	0.037	0.001	0.023
	(0.401)	(0.030)	(0.331)	(0.014)
¥10	-0.468	0.064*	-0.384	-0.002
	(0.333)	(0.034)	(0.335)	(0.010)
¥100	0.213	0.071**	-0.374	-0.001
	(0.342)	(0.034)	(0.333)	(0.010)
Observations	297	235	421	272
R-squared		0.25		0.11

Note. The sample includes participants' enrolment and performance in the semester after our experiment. Columns (1) and (3) report Poisson estimates of the treatment effects on the number of courses enrolled in the following semester. Columns (2) and (4) report the OLS estimates of the effects on the likelihood of obtaining certificates from enrolled courses. All specifications include user controls (as reported in Table 2), including gender, age, education, employment status, course and MOOC background, and baseline activity before the experiment. Robust standard errors are clustered at the participant level and are shown in parentheses. *** significant at the 1%, **5%, and *10% level.

Table 9. Heterogeneity by Gender

	Submission	Video Hours	Unconditional grade	Grade conditional on submission
	(1)	(2)	(3)	(4)
Panel A: Female				
¥1	0.098	0.032	0.131*	0.107**
	(0.084)	(0.068)	(0.076)	(0.052)
¥10	0.193**	0.015	0.219***	0.139***
	(0.080)	(0.061)	(0.073)	(0.049)
¥100	0.266***	0.099	0.276***	0.116**
	(0.080)	(0.063)	(0.072)	(0.046)
Observations	549	549	536	218
R-squared	0.48	0.37	0.48	0.08
Panel B: Male				
¥1	-0.008	-0.044	0.046	0.097
	(0.098)	(0.077)	(0.105)	(0.061)
¥10	0.007	-0.087	0.037	0.067
	(0.095)	(0.073)	(0.104)	(0.063)
100	0.079	-0.032	0.115	0.060
	(0.090)	(0.072)	(0.096)	(0.062)
Observations	342	342	335	133
R-squared	0.65	0.44	0.62	0.14

Note. Panel A uses the sample of female participants from *Chinese Culture* during the intervention period and Panel B uses male participants from the same class and same time period. The unit of observation is participant*homework for columns (1), (3) and (4), and participant*week for column (2). All specifications include user controls (as reported in Table 2), including gender, age, education, employment status, course and MOOC background, and baseline activity before the experiment. Robust standard errors are clustered at the participant level and are shown in parentheses. *** significant at the 1%, **5%, and *10% level.

Table 10. Heterogeneity by Offline Educational Resources

	Submission (1)	Video Hours (2)	Unconditional grade (3)	Grade conditional on submission (4)
Panel A: Few offline Edu Institutions				
¥1	0.286***	0.156*	0.358***	0.175
	(0.109)	(0.086)	(0.108)	(0.115)
¥10	0.305***	0.053	0.331***	0.141
	(0.097)	(0.075)	(0.100)	(0.110)
¥100	0.299***	0.182**	0.321***	0.135
	(0.088)	(0.074)	(0.090)	(0.109)
Observations	372	372	366	166
R-squared	0.56	0.37	0.56	0.16
Panel B: More offline Edu Institutions				
¥1	0.006	-0.072	0.028	0.055
	(0.084)	(0.069)	(0.079)	(0.042)
¥10	0.084	-0.012	0.109	0.070**
	(0.080)	(0.071)	(0.075)	(0.031)
100	0.168**	-0.048	0.192**	0.076**
	(0.083)	(0.072)	(0.077)	(0.029)
Observations	387	387	381	160
R-squared	0.61	0.43	0.59	0.14

Note. Participants' offline education resources are measured by the number of higher education institutions in their location (traced by IP address). Participants are divided by the sample median into the subsample of fewer (Panel A) or more offline educational resources (Panel B). Both panels use the participants from *Chinese Culture* during the intervention period. The unit of observation is participant*homework for columns (1), (3) and (4), and participant*week for column (2). All specifications include user controls (as reported in Table 2), including gender, age, education, employment status, course and MOOC background, and baseline activity before the experiment. Robust standard errors are clustered at the participant level and are shown in parentheses. *** significant at the 1%, **5%, and *10% level.

Appendix Table A1: Testing Sample Selection (Chinese Culture)

	Male	Age	Education	Employment	Experience	Network	Time Commitment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
¥1	0.041	1.578	-0.123	0.235	0.057	-0.048	0.111
	(0.103)	(1.729)	(0.138)	(0.212)	(0.206)	(0.070)	(0.121)
¥10	-0.081	0.755	-0.119	0.137	-0.074	-0.094	0.230**
	(0.101)	(1.606)	(0.137)	(0.214)	(0.199)	(0.066)	(0.116)
¥100	-0.075	1.320	-0.004	0.142	0.059	-0.018	0.105
	(0.101)	(1.700)	(0.131)	(0.213)	(0.200)	(0.073)	(0.116)
Observations	224	220	212	222	224	224	224
R^2	0.01	0.00	0.01	0.01	0.00	0.01	0.02
Mean dep var, control	0.371	25.971	2.187	1.765	1.800	0.143	2.286

	Retake	# of courses taken	# of certificates obtained	Mean HW score (before experiment)	Submission rate (before experiment)	Video hours (before experiment)
	(8)	(9)	(10)	(11)	(12)	(13)
¥1	0.076	0.152	-0.073	-0.022	-0.034	-0.460
	(0.074)	(0.568)	(0.200)	(0.041)	(0.064)	(0.640)
¥10	-0.018	-0.590	-0.018	0.007	-0.024	-0.949
	(0.066)	(0.546)	(0.202)	(0.037)	(0.064)	(0.630)
¥100	0.058	-0.316	-0.079	0.036	0.020	0.049
	(0.072)	(0.567)	(0.191)	(0.035)	(0.062)	(0.642)
Observations	224	224	224	224	224	224
R^2	0.01	0.01	0.00	0.01	0.00	0.02
Mean dep var, control	0.114	1.800	0.486	0.859	0.690	3.783

Note. The sample includes participants who made at least one homework submission during the intervention period. Each column reports the estimates of regressing treatment dummies on participant predetermined characteristics. Robust standard errors are clustered at the participant level and are shown in parentheses. *** significant at the 1%, **5%, and *10% level.

Appendix Table A2: Test Sample Selection (Data Structure)

	Male (1)	Age (2)	Education (3)	Employment (4)	Experience (5)	Network (6)	Time Commitment (7)
¥1	-0.063	-0.079	0.149	-0.112	0.149	-0.119	0.010
	(0.075)	(1.031)	(0.156)	(0.207)	(0.228)	(0.080)	(0.124)
¥10	-0.055	0.226	0.140	-0.058	0.180	-0.141*	-0.057
	(0.073)	(1.023)	(0.148)	(0.206)	(0.227)	(0.077)	(0.132)
¥100	-0.170*	0.713	0.278*	-0.242	0.013	-0.049	-0.036
	(0.091)	(1.376)	(0.153)	(0.209)	(0.236)	(0.091)	(0.137)
Observations	162	162	157	162	162	162	162
R^2	0.02	0.00	0.03	0.01	0.01	0.04	0.00
Mean dep var, control	0.920	23.120	2.000	1.520	2.320	0.160	2.480

	Retake (8)	# of courses Taken (9)	# of certificates obtained (10)	Mean HW score (before experiment) (11)	Submission rate (before experiment) (12)	Video hour (before experiment) (13)
¥1	0.029	-1.144	-0.018	0.029	0.010	0.293
	(0.124)	(1.101)	(0.102)	(0.043)	(0.071)	(0.826)
¥10	-0.036	-1.325	-0.024	0.042	-0.022	0.156
	(0.122)	(1.079)	(0.097)	(0.040)	(0.072)	(0.788)
¥100	0.004	-1.043	0.213	0.000	0.006	0.107
	(0.131)	(1.128)	(0.181)	(0.045)	(0.078)	(0.806)
Observations	162	162	162	162	162	162
R^2	0.00	0.02	0.03	0.02	0.00	0.00
Mean dep var, control	0.440	2.960	0.120	0.856	0.472	3.325

Note. The sample includes participants who made at least one homework submission during the intervention period. Each column reports the estimates of regressing treatment dummies on participant predetermined characteristics. Robust standard errors are clustered at the participant level and are shown in parentheses. *** significant at the 1%, **5%, and *10% level.