Journal of Economic Behavior and Organization 190 (2021) 905–921

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

How monetary incentives improve outcomes in MOOCs: Evidence from a field experiment[☆]

Jie Gong[a], Tracy Xiao Liu[b,*], Jie Tang[c]

[a] Department of Strategy and Policy, NUS Business School, National University of Singapore, Singapore
[b] Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China
[c] Department of Computer Science and Technology, Tsinghua University, Beijing, China

ARTICLE INFO

Article history:
Received 7 October 2020
Revised 13 June 2021
Accepted 14 June 2021
Available online 5 September 2021

Keywords:
Online platform design
MOOCs
Field experiment

ABSTRACT

In this study, we examine the impact of monetary incentives on user engagement and learning outcomes in massive open online courses (MOOCs). While MOOCs offer high-quality interactive educational resources to users worldwide, maintaining user engagement and enthusiasm on these platforms 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 homework grades. 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. Overall, our findings suggest that monetary incentives counteract engagement decay and may help online users form persistent learning habits.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Over the past decade, online learning has become an established component of higher education and corporate training and a multibillion-dollar market in modern economies.[1] MOOCs offer affordable (often free) educational content from highly reputable sources and have the potential to vastly increase human capital accumulation. However, MOOCs struggle with low completion rates and a significant decay in learner activity across the semester. For instance, Kizilcec et al. (2013) and Seaton et al. (2014) document that only 5% of MOOC users have completed a course. The low completion rates could be a result of a matching process if, for example, the users just shop around or engage in the sections that are most important

☆ We thank Yan Chen, Jonathan Guryan, Jennifer Hunt, Peter Kuhn, Erica Li, Sherry Xin Li, Fangwen Lu, Ben Roth, and seminar participants at the 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 support from XuetangX and the Online Education Office, Tsinghua University. Financial support from Ministry of Education of the People's Republic of China (MOE) Research Center for Online Education, NSFC for Distinguished Young Scholar (61825602) and the National Key Research and Development Program of China (No. 2018YFB1004503) is gratefully acknowledged.

* Corresponding author.
E-mail addresses: gong@nus.edu.sg (J. Gong), liuxiao@sem.tsinghua.edu.cn (T.X. Liu), jietang@tsinghua.edu.cn (J. Tang).

[1] Coursera, for instance, raised $103 million in Series E funding that valued the company around $1 billion.

https://doi.org/10.1016/j.jebo.2021.06.029
0167-2681/© 2021 Elsevier B.V. All rights reserved.

to them. For such users, it is rational and expected not to complete a course, as the opportunity cost of testing interest in a new subject is negligible.

However, the low completion rates can be a significant concern if it is due to either content or individual problems. If the quality of content and instruction discourage user engagement, it is relatively easier to fix. But if the specific nature of an online environment—e.g., lack of monitoring and peer groups—leads to severe self-control problems, it could be more challenging to alter individual learning habit. Indeed, Banerjee and Duflo (2014) find that less organized students are less likely to succeed in a MOOC due to a failure to complete assignments rather than poor performance on completed assignments.

The aim of this study is to examine alternative methods of incentivizing online users and motivating self-discipline. In particular, we examine whether monetary incentives can effectively mitigate the observed decay in activity across an online course. We conduct a field experiment on XuetangX, the third largest MOOC platform worldwide. The setting appears to entail an engagement challenge due to individual learning habits. Past complement rates have been low but a good share of registered users have clear goals (i.e., they are not shopping around). Of the participants in our study, 23% state that they enrolled in the course to earn a certificate and 22% because the content would be helpful in their jobs.

We select two courses offered in Spring 2015, *Cultural Treasure and Chinese Culture (Chinese Culture)* and *Data Structure and Algorithm (Data Structure)*, and reward 760 participants for completing homework assignments across a 4-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 (the official currency of China),[2] respectively, for each completed assignment that exceeds our prespecified 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 that falls short of the grade threshold. We implement the incentive for three assignments around the middle of the semester and collect learners' 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) on average improves the assignment completion rate by 12.8% and assignment grades by 9.2% (conditional on submission). A medium incentive (10 rmb reward or loss) shows significant effects for users 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 and those from regions with fewer higher education institutions are more responsive to monetary incentives.

Our findings suggest that offering a monetary incentive could be a scalable solution to sustain online users' engagement and performance. In our experiment, 100 rmb is universally effective and more than sufficient for one course; the scheme covers only three assignments over 4 weeks. In comparison, the benefits are large in magnitude and persistent over time; we do not find evidence of a monetary incentive crowding out intrinsic motivation. Our findings on persistence of the incentive effects also provide evidence that the baseline low engagement may be suboptimal. The evidence of spillover across time and courses further implies that increased engagement is more likely to be driven by changes in learning habits rather than selection of courses. Another point worth noting is that offering an incentive may affect the decision to enroll in a course (e.g., attract more users); although offline classrooms have limited seats, the online environment does not have such crowding costs, and therefore can accommodate more users who are attracted by the incentives. Along the same line, we are aware of the potential sample selection due to the way we recruit participants. Users who volunteered to participate in the experiment may differ from those who did not. Our findings might be biased towards frequent or attentive users. We are therefore cautious in extending the estimated effects to more representative samples.

Our study contributes to the stream of literature that puts behavioral economics to practice; in particular, regarding the effect of financial incentives in motivating learning. Prior studies mostly focus on traditional offline classrooms and find mixed results.[3] 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., 2016b), though significant treatment heterogeneity exists between courses (Bettinger, 2012) and students (De Paola et al., 2012). A few also find a long-term post-incentive effect (Angrist et al., 2006; 2009; De Paola et al., 2012) while others find negative impact in the long-run, especially for low-ability students (Campos-Mercade and Wengström, 2020; Leuven et al., 2010). Additionally, using a large-scale field experiment in three U.S. cities, Fryer (2011) finds no significant effect of a financial incentive on students' scores. One of the most relevant field experiments is Bellés-Obrero (2020), which reports a significant interaction effect between incentive schemes and students type in the online education setting. For example, rewarding top students has positive (negative) impact on those with high (low) intrinsic motivation.

We also make a first attempt to experimentally investigate the size and framing effect of financial incentives on online learning. Examining previous evidence in offline classrooms, it is ex ante unclear whether these results will hold in online settings. It is possible that online learners may have different and/or more diverse motivations for pursuing their education, including intrinsic goals, personal interest in the topic, or career advancement. Therefore, they may be less responsive to

[2] 1 rmb $\approx$ 0.15 US dollars.

[3] Campos-Mercade and Wengström (2020) provides a comprehensive review for monetary incentives in education.

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.

Our study further contributes to the literature of behavioral economics that documents the persistent and spillover effects of incentives. Charness and Gneezy (2009) find that paying people to visit a gym helps develop exercise habits and improves health outcomes in the long run. Royer et al. (2015) show that although incentive programs that target the use of a gym have limited lasting effects, an additional commitment option has strong long-run effects that extend beyond the end of the commitment-contract period. 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.

A growing literature studies the effectiveness of online learning. 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 the 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 from better socioeconomic backgrounds are more likely to succeed in MOOCs. Cacault et al. (2019) show that online live streaming of lectures lowers achievement for low-ability students and increases achievement for high-ability ones.

Regarding learning effectiveness, Banerjee and Duflo (2014) document significant engagement decay in online courses and find evidence for a self-control problem. To combat this issue, Patterson (2018) tests three behaviorally motivated tools and shows that a commitment device can increase effort and performance in an online course, while an alert or a distraction blocking tool fail to motivate online learners. Jaggars and Xu (2016) and Zhang et al. (2017) find that promoting social interaction significantly improves both students' completion rates and their course grades. Kizilcec and Brooks (2017) further survey a broad range of field experiments on MOOCs. Our experiment aims to evaluate the effectiveness of another possible mechanism to encourage student engagement and performance: offering a monetary incentive under different framing.

2. Field setting: XuetangX

XuetangX was launched in China in 2013 as a start-up MOOC platform affiliated with Tsinghua University and the Ministry 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.[4] 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 U.S. universities including MIT, Stanford, and UC Berkeley. Compared with other MOOCs, XuetangX is more public in nature and provides a greater number of free courses and accounts to alleviate education inequality and promote life-long learning in China.[5]

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).[6] We draw on one course from each of these fields, i.e., *Chinese Culture* and *Data Structure*, 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 out 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 a blank 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 and view lecture videos, complete assignments, post a thread in a course forum, or respond to other students' posts. Qiu et al. (2016) describe these activities and summarize 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 viewing lecture videos and submitting assignments decrease significantly over time. Similarly, Feng et al. (2018) find that the likelihood of a user dropping a course is positively correlated

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

[5] 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.

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

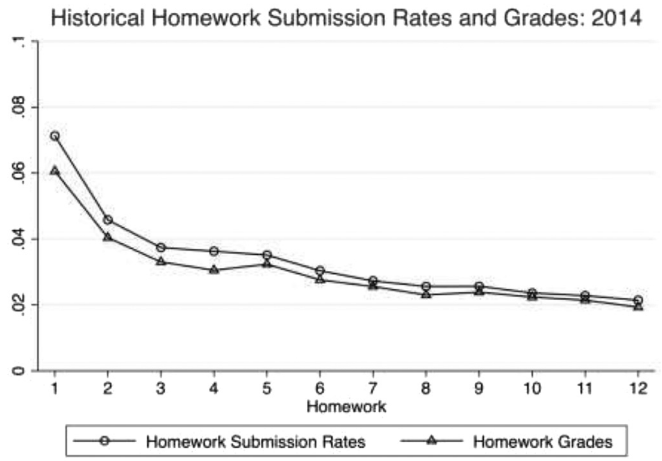


Fig. 1. Historical homework records.

with dropping another course, suggesting an overall engagement decrease rather than effort reallocation among courses. Fig. 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 2 weeks, with a reduction of more than half by midterm. This finding is consistent with engagement patterns found in other MOOC studies (e.g., Banerjee and Duflo 2014).

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 by Qiu et al. (2016), courses in art and literature differ from those in science and engineering in their requirements, difficulty level, and 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. For both courses, users receive a certificate if their final grades exceed 60 out of 100 points.

The first course in our experiment, *Chinese Culture*, is offered by the Department of History at Tsinghua University and lasts from March 2, 2015 to June 22, 2015. The course consists of 16 lectures, with one assignment posted that corresponds to each lecture.[7] Students can complete the assignments anytime 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 assignments around the middle of the semester for two reasons. First, enrollment is not stabilized during the first few weeks and we don't want to risk losing a large number of participants during the experiment; second, the first 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 at 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. Each assignment is due one month after it is posted, and accounts for 5% of a student's final grade (60% in total). In addition, four programming projects account for 40%. Our intervention targets three assignments—assignments 6, 7, and 8; again, we collect data on user activities throughout the whole semester.

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

3.1. Treatments

Each of our treatment groups is given 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 course in the fall 2014 semester—i.e., one semester before our experiment—and find that conditional on submission (nonzero grades), both the mean and median scores are 80% for each course. We therefore consider this to be a feasible target that students can meet with a reasonable amount of effort. 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*.

The participants in our experiment are randomly assigned to 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 that encourages them to complete their assignments. The same email is sent to the treatment groups. For instance, the control group in the *Data Structure* course receives the following message:

> *Data Structure* has been updated to the 6th homework assignment. 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 this 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 offered one of three different levels of payment size: 1 rmb, 10 rmb and 100 rmb. 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. Therefore, one rmb is considered a small amount. We use the one rmb treatment group to test whether a small monetary incentive may crowd out the intrinsic motivation of learning. Moreover, it can help account for other common confounding factors in the literature, such as reference-dependency, saliency, goal-setting, and anchoring.[8] Ten rmb represents a medium amount, which is an acceptable amount as a reward. One hundred rmb is the largest amount and is considered a generous reward. For comparison, student TAs at Tsinghua University are paid 24 rmb per hour.

Regarding framing, findings from prior literature are mixed. Hossain and List (2012) and Hong et al. (2015) find that framing bonus as loss significantly increases factory workers' productivity. Andreoni (1995) finds that positive framing of an incentive significantly increases participants' contributions in public goods. Furthermore, both Fryer et al. (2012) and Levitt et al. (2016a) find that framing incentives in the loss domain is more effective for enhancing students' performance. Finally, Karlan et al. (2016) find no significant effect of incentive framing on individuals' saving behavior, and Chen et al. (2018) similarly find no framing effect on students arrival time for experimental sessions. In our experiment, we 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 assignment that receives a score of at least 80%. Positive-framing participants receive the following message:

> For the next 3 assignments, you will receive an X rmb reward for every assignment grade that is $\geq$80% of the total score.[9]

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

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

To minimize potential collusion between students, we impose a deadline such that to claim the monetary reward, students must submit their assignment within 2 weeks of the assignment's posting date. User activity from the past semester shows that most homework submissions are made within 2 weeks of the assignment posting date.[10] Online Appendix A includes a sample of the experimental emails sent to subjects in the treatment groups.

3.2. Experimental procedure

Table 1A summarizes the experiment timeline and data. We conducted our experiment in the spring of 2015. On April 6, 2015, we sent recruiting emails to enrolled students in the two courses (5714 users in *Chinese Culture* and 9720 users in *Data Structure*). We also posted a recruiting message on the announcement board for each course on behalf of XuetangX. 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 (Chen and Konstan 2015).

[8] We thank an anonymous referee for pointing this out.

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

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

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 1B: Number of Subjects by Treatment and Course

	Chinese Culture	Data Structure	Overall
Control	46	61	107
+ ¥1	48	62	110
+ ¥10	46	63	109
+ ¥100	46	63	109
- ¥1	46	62	108
- ¥10	48	62	110
- ¥100	48	59	107

By April 14, 337 users from *Chinese Culture* and 455 users from *Data Structure* had signed up for our study and completed a survey on their demographic characteristics. Online Appendix C includes pre-experiment survey questions. Participants understood that they had signed up for a study on online learning, but were not told the details or 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. Nine users registered for both courses. We only randomly assign them to treatments for *Data Structure*.

Altogether, our subject group consists of 328 users enrolled in the *Chinese Culture* course and 432 in the *Data Structure* course. Table 2 reports summary statistics for participant demographic characteristics and pre-experiment course performance. The statistics in Table 2 show that our participants are generally young (mean age is 25 years), educated (the majority have a 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 6 weeks of the course (before they sign up for the experiment), we also see that the *Chinese Culture* class in general has a higher participation rate and student performance profile, possibly due to less challenging content and requirements.

The way we recruit participants could introduce selection bias in that the participants are different from the nonparticipants. We formally compare the two groups of users, in particular their gender, age, education background, the number of courses they have taken at the platform, and the number of certificates they have obtained so far.[11] As shown in Appendix Table C, we draw more female users, who are slightly older (although the magnitude of the age difference is negligible) and more educated than users who registered for either of the two courses but did not participate in our study. Interestingly, the participants tend to register fewer courses but manage to gain more certificates than the nonparticipants. As the participants are not perfect representative of the universe of XuetangX users, one should be cautious in extending the treatment effects to a broader population.

For each course, we randomly assign each subject to either the control or one of the six treatment groups (i.e., complete randomization). Table 1B lists the number of subjects by treatment and course. For each of the user demographics and learning experience variables, we conduct two-sided t-tests between our treatment groups and the control group and adjust the p-values by employing the False Discovery Rate of Benjamini and Hochberg (1995) (Appendix Tables B1 and B2).[12] 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 20, right before the 8th (6th) assignment posting for *Chinese Culture (Data Structure)*. As mentioned in Section 3.1, the control group receives an email that encourages them to complete their homework, while 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 2 weeks to complete their assignment 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 2-week period. After each collection, participants are immediately informed how much they have earned from the previous assignment. On average, participants in *Chinese Culture* earn 61.21 rmb from the intervention stage and those in *Data Structure* earn 38.91 rmb.

[11] We could not compare all the variables as in Table 2, as the nonparticipants are not surveyed for their motivation, time commitment, etc.

[12] This method reduces false positives, and is commonly used to address multiple hypothesis testing problem (e.g., Edmonds and Theoharides 2020; Masset et al. 2020).

Table 2
Summary statistics.

Sample	Chinese culture (1)	Data structure (2)	Overall (3)
Male	0.377	0.807	0.623
	(0.485)	(0.395)	(0.485)
Age	27.11	23.72	25.162
	(8.060)	(5.119)	(6.743)
Education			
Below College	0.140	0.0758	0.102
	(0.347)	(0.265)	(0.303)
College	0.648	0.697	0.676
	(0.478)	(0.460)	(0.468)
Master and PhD	0.213	0.227	0.221
	(0.410)	(0.420)	(0.415)
Employment Status			
Student	0.506	0.716	0.626
	(0.501)	(0.451)	(0.484)
Unemployed	0.0528	0.0512	0.0519
	(0.224)	(0.221)	(0.222)
Employed	0.438	0.230	0.319
	(0.497)	(0.421)	(0.466)
Retired	0.00311	0.00233	0.00266
	(0.0557)	(0.0482)	(0.0515)
Subject and MOOC background			
Experience with the subject	1.820	2.248	2.065
	(0.938)	(0.900)	(0.940)
Friends taking the same course	0.0926	0.0626	0.0755
	(0.290)	(0.243)	(0.264)
Time commitment	2.475	2.566	2.527
	(0.564)	(0.541)	(0.553)
Retake this course	0.194	0.367	0.293
	(0.396)	(0.482)	(0.455)
Number of courses taken	1.981	2.248	2.134
	(3.345)	(3.638)	(3.516)
Number of certificates obtained	0.395	0.146	0.253
	(0.828)	(0.523)	(0.682)
Activity before experiment			
Homework score	0.598	0.330	0.445
	(0.431)	(0.436)	(0.453)
Homework submission rate	0.470	0.176	0.302
	(0.412)	(0.286)	(0.374)
Weekly video hours	2.645	1.636	2.069
	(3.390)	(2.810)	(3.110)
Observations	324	431	755

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) pool participants from the two courses and report the overall mean of the variable values.

Our intervention covers a span of three assignments for each course and ends on June 12. After completion of the 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 end on July 1, we send participants a final survey to collect their long-term responses to the experimental manipulation. To encourage participation in the post-experiment surveys, we pay 5 rmb for filling out each survey and also award a 100 rmb prize drawn randomly from the respondents. All payments are transferred to participants through XuetangX, and they were informed of this procedure before participating in the study. Online Appendices C and D contain 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 study long-term effects on the same set of outcome variables after incentives are removed. We exclude one participant from *Data Structure* and four from *Chinese Culture*, as they dropped their respective courses before any monetary incentive occurred.[13] Altogether, we have 324 subjects in *Chinese Culture* and 431 in *Data Structure*.

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

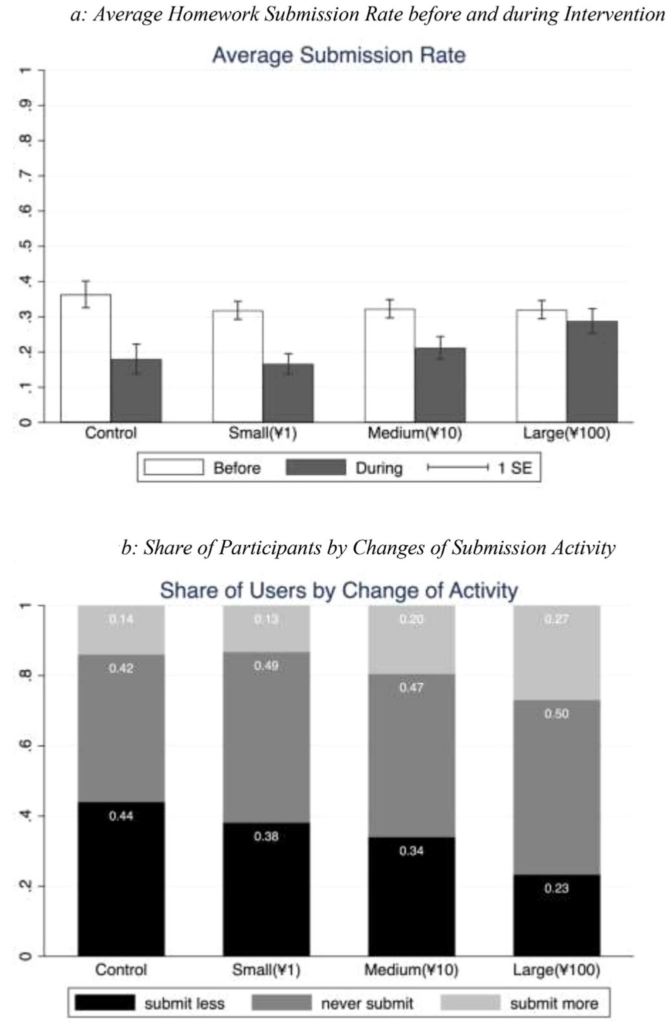


Fig. 2. a: Average homework submission rate before and during Intervention. b: Share of participants by changes of submission activity.

4.1. Treatment effects on grades

Fig. 2 a presents the homework submission rate before and during the intervention. The control group exhibits a significant decay over the course.[14] The average submission rate before the intervention across two courses is 36%; this drops to 18% during the intervention ($p < 0.01$, 2-sided test of proportions). We find a similar decrease for our 1 rmb group (32% to 17%, $p < 0.01$, 2-sided test of proportions) and a smaller drop for 10 rmb group (32% to 21%, $p < 0.01$, 2-sided test of proportions), and that our 100 rmb group exhibits a much smaller and statistically insignificant drop from 32% to 28% ($p = 0.15$, 2-sided test of proportions). Thus, the 100 rmb incentives are effective in maintaining users' motivation to complete homework assignments. There is no evidence that 1 rmb incentive maintains users engagement or even has negative impact on engagement.[15]

We next examine the effect of a monetary incentive at the individual level. Examining within-user changes before and during our intervention, we classify subjects into three types: users who decrease submissions, users who increase submissions, and those who never submits homework. Fig. 2b presents the share of each type of user. We find that, on average 44%

[14] We do not find significant difference between framing the incentive as gain versus loss, and combine the two framing treatments. 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).

[15] When we separately analyze the two courses, we find that in *Chinese Culture*, the decrease in submission rate is insignificant for 10 rmb group (Appendix Figure 1a: 42% to 38%, $p = 0.11$, 2-sided test of proportions). This indicates the difference in engagement across the two courses. *Data Structure* has a much lower baseline submission rate, and a 10 rmb incentive is sufficient to keep users engaged in *Chinese Culture* but not in *Data Structure*. One reason could be the higher cost required to complete assignments in *Data Structure*. 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 and apply a method and algorithm.

Table 3
Homework submission rate during intervention.

	Outcome: Whether homework submitted on time		
	(1)	(2)	(3)
¥1	0.007	0.008	-0.052
	(0.032)	(0.039)	(0.033)
¥10	0.047	0.060*	-0.045
	(0.031)	(0.035)	(0.035)
¥100	0.128***	0.133***	0.013
	(0.032)	(0.035)	(0.036)
¥1 x punish		-0.001	
		(0.035)	
¥100 x punish		-0.027	
		(0.033)	
¥100 x punish		-0.010	
		(0.037)	
¥1 x active			0.122*
			(0.065)
¥10 x active			0.203***
			(0.064)
¥100 x active			0.267***
			(0.066)
User controls	yes	yes	yes
Course FE	yes	yes	yes
Observations	2154	2154	2154
R-squared	0.52	0.52	0.54

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. "Punish" is a dummy variable indicating whether the incentive is introduced as a loss for each assignment that fails to meet the threshold. "Active" is a dummy variable indicating whether a participant's pre-intervention submission rate is above the sample mean. All specifications include course fixed effects and 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.

of users in the control group reduce their submission rate during our intervention period, compared with 38% in the 1 rmb group, 34% in the 10 rmb group, and 23% in the 100 rmb group, with a significant difference for the control and 100 group ($p < 0.01$, 2-sided test of proportions). We further find that 14% of users in the control group increase their submission rate during our intervention, compared with 20% in the 10 rmb group and 27% in the 100 rmb group, with a significant difference for the control and 100 rmb groups ($p < 0.01$, 2-sided test of proportions). Of those who increase their submissions during intervention, only 2% (1%, 5%) of the control (10 rmb, 100 rmb) group had not previously submitted assignments. Overall, our individual-level analysis suggests that the treatment effect (higher submission rate) reflects the maintenance of existing engagement levels rather than any 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:

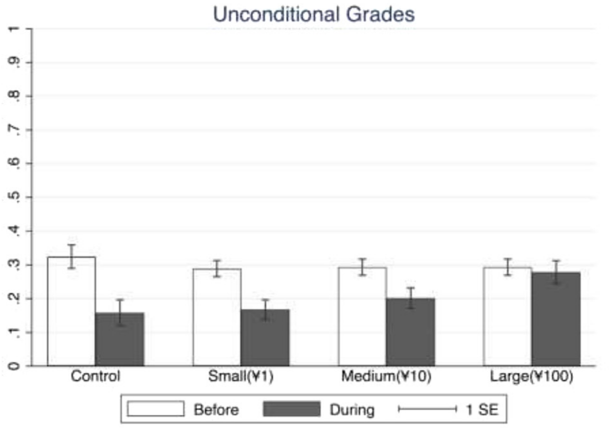
$$Y_{ijt} = \alpha + \sum_j \beta_j \times Treatment_j + X_i + \varepsilon_{ij} \tag{1}$$

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, experience with online learning, and baseline activity before the experiment. 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 Eq. (1) with linear probability models and present the results in Table 3.[16] We pool two courses together in the main analysis and include course fixed effects in all specifications.

Overall, regression results confirm our graphical evidence. First, a 100 rmb incentive significantly increases homework submission. As shown in Column 1, a 100 rmb incentive raises the probability of submission by 12.8 percentage points. A 10 rmb incentive has positive effects but the coefficient is not statistically significant. A small amount of incentive (1 rmb)

[16] Probit or logit models yield similar results. Also, the estimates are not sensitive to the inclusion of week nor homework fixed effects.

a: Unconditional Means of Homework Grades before and during Intervention



b: Homework Grades Conditional on Submission, before and during Intervention

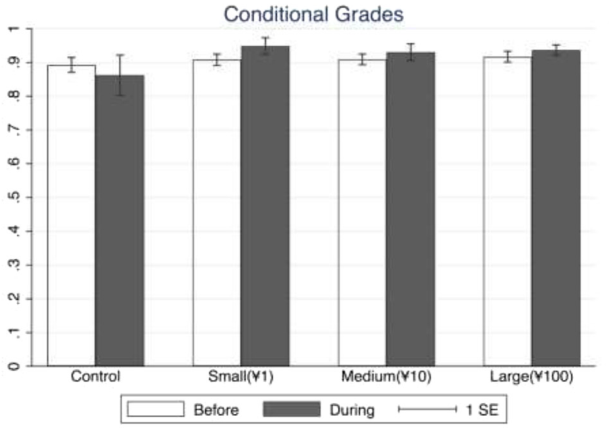


Fig. 3. a: Unconditional means of homework grades before and during intervention. b: Homework grades conditional on submission, before and during intervention.

does not show positive or adverse effects.[17] Second, we include the interaction between incentive amount and frame and present the results in Column 2. 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 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.[18] As shown in Columns 3, the coefficients of the interaction terms 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.[19]

4.2. Treatment effects on grades

In this section, we examine whether a monetary incentive impacts assignment grades during the intervention. Fig. 3a presents the unconditional mean of homework grades and a non-submission is coded as zero. We find results similar to those of our homework assignment submission analysis. For example, the 100 rmb groups exhibit a significant continuance of performance. The 1 rmb group show similar pattern as the control group while the 10 rmb groups perform slightly better than the control.[20]

[17] The effects for the 1 rmb and 10 rmb groups are significantly different from that of the 100 group ($p < .001$).

[18] The pre-intervention activities are also included in the baseline control.

[19] We also repeat the same regression analysis for each course separately (Appendix Table D). The main findings are consistent across courses, except that a 10 rmb incentive has a positive effect in *Chinese Culture* but not *Data Structure*.

[20] Similar to the difference in submission across the two courses, in *Chinese Culture*, both the 10 rmb and 100 rmb groups exhibit a significant continuance of performance while in *Data Structure*, only 100 rmb has a significant impact (Appendix Figure 2).

Table 4
Treatment effects on homework grades.

	Unconditional grade	Grade conditional on submission	Upper bound	Lower bound
	(1)	(2)	(3)	(4)
¥1	0.033	0.101**	0.110***	0.101**
	(0.032)	(0.039)	(0.038)	(0.039)
¥10	0.061*	0.089**	0.114***	0.084**
	(0.031)	(0.039)	(0.037)	(0.039)
¥100	0.139***	0.092**	0.101***	0.084**
	(0.031)	(0.036)	(0.035)	(0.036)
User controls	yes	yes	yes	yes
Course FE	yes	yes	yes	yes
Observations	2085	461	445	432
R-squared	0.51	0.08	0.12	0.08

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 course fixed effects and 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.

Fig. 3 b 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 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 subset of students who are motivated to submit their homework assignments. If so, 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 whether 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 described below.

As a baseline, we fit Eq. (1) to estimate the treatment effects on assignment grades during the intervention period. Table 4 reports OLS estimates for pooling across courses. All specifications include user controls and course fixed effects. 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 an 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 group during our intervention. Column 1 show that a 10 rmb incentive has a marginally significant effect on the unconditional grade, while a 100 rmb incentive significantly improves homework grades relative to the control group. Column 2 shows that all three treatments have positive and significant effects on conditional grades.

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).[21] The estimated upper and lower bounds for all three treatments 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 as 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 of Qiu et al. (2016) to capture their effective viewing time.[22] This measure allows us to measure learning activity and effort, which is difficult to observe or quantify in a traditional offline classroom environment.

The raw data for viewing time have critical flaws in capturing user learning time. For instance, a user sometimes turns on a lecture video but switches to other unrelated tasks without watching the video. Such data may introduce measurement errors and contaminate our estimates. Therefore, we apply the algorithm developed in "Modeling and Predicting Learning

[21] In our context, we consider, for example, the case in which the 12.8% 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 12.8% 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 12.8% grades of the 100 rmb group so that the estimate from the refined and balanced sample represents the lower bound of the true effect.

[22] The raw data for viewing time may fail to capture user effective learning time. For example, a user sometimes turns on a lecture video but switches to other unrelated tasks without watching the video. We therefore apply the algorithm developed in Qiu et al. (2016), which can subtract such "idle time" from effective learning time.

Table 5
Treatment effects on video hours.

	Outcome: ln (weekly video hours)		
	(1)	(2)	(3)
¥1	-0.009	-0.014	-0.047
	(0.029)	(0.034)	(0.036)
¥10	-0.008	-0.015	-0.051
	(0.028)	(0.033)	(0.036)
¥100	0.058**	0.074**	-0.024
	(0.030)	(0.033)	(0.039)
¥1 x punish		0.010	
		(0.030)	
¥100 x punish		0.016	
		(0.027)	
¥100 x punish		-0.033	
		(0.033)	
¥1 x active			0.081
			(0.056)
¥10 x active			0.092*
			(0.056)
¥100 x active			0.193***
			(0.060)
User controls	yes	yes	yes
Course FE	yes	yes	yes
Observations	2154	2154	2154
R-squared	0.31	0.31	0.32

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. "Punish" is a dummy variable indicating whether the incentive is introduced as a loss for each assignment that fails to meet the threshold. "Active" is a dummy variable indicating whether a participant's pre-intervention submission rate is above the sample mean. All specifications include course fixed effects and 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

Behavior in Moocs," in "Proceedings of the Ninth ACM International Conference on Web Search and Data Mininge (Qiu, Tang, Liu, Gong, Zhang, Zhang, Xue, 2016), which can subtract such "idle time" from effective learning time. For example, a "pause" activity triggers students from playing video to "idle" state.

The results in Table 5 show that the 100 rmb treatment increases weekly video time by 5.8% (Column 1). Interestingly, we find 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 nonlinear navigation strategy. As documented by 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 nonlinear navigation implies that better performance does not necessarily come from more hours spent viewing course materials. Results shown in Columns 2 and 3 are consistent with our findings on treatment effects on submission rate, that framing the incentives as rewards or punishments does not have different effects and the treatment effects are largely driven by sustained activity of regular users.

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).[23] 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

[23] In particular, Gneezy and List (2006) shows that the positive reciprocity that arises 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 over to post-experiment periods, generating a negative net effect on the participation rate.

Table 6
Treatment effects after incentives are removed.

	Submission	Unconditional grade	Grade conditional on submission
	(1)	(2)	(3)
¥1	0.034	0.049	0.018
	(0.036)	(0.034)	(0.028)
¥10	0.072*	0.087**	0.051*
	(0.037)	(0.035)	(0.027)
¥100	0.102***	0.106***	0.035
	(0.037)	(0.035)	(0.025)
User controls	yes	yes	yes
Course FE	yes	yes	yes
Observations	3763	3554	806
R-squared	0.45	0.46	0.05

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 course fixed effects and 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.

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. About 2 months remain until the courses end, during which participants in *Chinese Culture* are assigned seven more homework assignments, and those in *Data Structure* four 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. We find that the 100 rmb group continues to increase their homework submission rates and assignment performance. For instance, participants who were offered a 100 rmb incentive are still 10.2 percentage points more likely to submit their homework, and their (unconditional) grades are 0.106 points higher. While the magnitude is lower than during the intervention, we cannot reject the hypothesis that the treatment effects are equal between during- and post-intervention phases—i.e., no significant decay. The 10 rmb group also show some positive effects although the coefficients on submission and conditional grades are marginally significant. 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, suggesting habit formation or certain types of learning about the online course experience.[24] Third, there is no evidence of a decrease in performance, suggesting that offering incentives, either small or large, does not necessarily crowd out students' intrinsic motivation to learn. Also, the fact that the removal of incentives does not reverse student engagement toward the pre-intervention level suggests that the baseline low engagement is suboptimal for the students.

Lastly, we investigate the spillover effect of a monetary incentive to other courses during the same and 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 on 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, Columns 1 to 3). Furthermore, we find that treated participants still outperform the control group in their subsequent semester courses, as measured by obtaining a certificate (Table 7, Column 5).[25]

Our findings on spillover effects lend further support to long-run improvement in student engagement. If offering a 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 participants do not compromise their engagement and performance in other courses, or stop exerting effort once the incentive is removed. Our intervention ap-

[24] Another explanation for the persistent effect is that the value of doing well—e.g., the likelihood of completing the course and receiving a certificate—on later assignments may be higher conditional on doing well on previous assignments. Our analysis of spillover effect does not support this alternative explanation.

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

Table 7
Spillover effects.

	Video hour on other courses during intervention (1)	Video hour on other courses after intervention (2)	Grade of other courses during the same semester (3)	# of enrolled courses in the following semester (4)	Certificate rate in the following semester (5)
¥1	0.045	0.022	-0.005	-0.220	0.035**
	(0.030)	(0.027)	(0.019)	(0.266)	(0.016)
¥10	0.033	0.017	-0.006	-0.408	0.030*
	(0.029)	(0.030)	(0.018)	(0.255)	(0.016)
¥100	0.080**	0.040	0.039*	-0.101	0.039**
	(0.032)	(0.029)	(0.020)	(0.253)	(0.017)
User controls	yes	yes	yes	yes	yes
Course FE	yes	yes	yes	yes	yes
Observations	2154	3763	639	718	507
R-squared	0.11	0.09	0.27	.	0.19

Note. The sample for columns (1) to (3) include participants' video activity and performance in other courses they enrolled in during the same semester. The sample for columns (4) and (5) include participants' enrolment and performance in the semester after our experiment. Columns (4) report Poisson estimates of the treatment effects on the number of courses enrolled in the following semester. Columns (5) report the OLS estimates of the effects on the likelihood of obtaining certificates from enrolled courses. All specifications include course fixed effects and 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.

pears to have helped them learn about the online learning process, or about disciplining themselves—which, in turn, shifts their learning behavior toward a more persistent and sustainable pattern.

5. Discussion

To the best of our knowledge, ours is one of the first 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. 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 finding by 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 using 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 assignments, and therefore the positive treatment effects echo the efficacy of rewarding concrete tasks.

Our results have both practical and academic implications. On a practical level, the platform on which 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 scholarship program to motivate learning.[26]

On a broader level, our findings can also be used by public programs that promote online learning in designing platforms that better utilize the resources invested by teachers, universities, and the public sector in online courses. In designing online public courses, it is useful to understand which 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 8 and 9, respectively).[27] For our geographic data, we use subjects' IP addresses and control for the GDP per capita of the region to ensure that 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.[28] In fact, the *Data Structure* course includes a multistage programming tournament throughout the course, in which winners can access materials (programming projects) that are available exclusively for the Tsinghua computer science department.

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

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

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

Table 8
Heterogeneity by gender.

	Submission (1)	Video Hours (2)	Unconditional grade (3)	Grade conditional on submission (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 9
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.

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 courses for both empirical and logistical reasons. However, it would be interesting to see what effect would occur if the intervention were instead provided at the beginning of each course. With big data on user activity, we might even possibly predict the "hazard rates" for users at any given moment, and design customized instruments, e.g., individualized social information (Shang and Croson, 2009), to keep them engaged and improve learning outcomes. Another potential behavioral mechanism which may promote students' engagement is to create group identity by forming study groups, or prime natural identity (Chen et al., 2014; Mobius et al., 2016) for online learners, and to promote competition between groups (Akerlof and Kranton, 2000; 2005; 2008; 2013). 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.

Acknowledgements

Financial support from Ministry of Education of the People's Republic of China (MOE) Research Center for Online Education, NSFC for Distinguished Young Scholar (61825602) and the National Key Research and Development Program of China (No. 2018YFB1004503) is gratefully acknowledged.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.06.029

References

Acemoglu, D., Laibson, D., List, J.A., 2014. Equalizing superstars: the internet and the democratization of education. Am. Econ. Rev. Pap. Proc. 104.
Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. Q. J. Econ. 115 (3), 715–753.
Akerlof, G.A., Kranton, R.E., 2005. Identity and the economics of organizations. J. Econ. Perspect. 19 (1), 9–32.
Akerlof, G.A., Kranton, R.E., 2008. Identity, supervision, and work groups. Am. Econ. Rev. 98 (2), 212–217.
Akerlof, G.A., Kranton, R.E., 2013. Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being. Princeton University Press.
Andreoni, J., 1995. Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. Q. J. Econ. 110 (1), 1–21.
Angrist, J., Bettinger, E., Bloom, E., King, E., Kremer, M., 2002. Vouchers for private schooling in colombia: evidence from a randomized natural experiment. Am. Econ. Rev. 92 (5), 1535–1558.
Angrist, J., Bettinger, E., Kremer, M., 2006. Long-term educational consequences of secondary school vouchers: evidence from administrative records in colombia. Am. Econ. Rev. 96 (3), 847–862.
Angrist, J., Lang, D., Oreopoulos, P., 2009. Incentives and services for college achievement: evidence from a randomized trial. Am. Econ. J. 1 (1), 136–163.
Angrist, J., Lavy, V., 2009. The effects of high stakes high school achievement awards: evidence from a randomized trial. Am. Econ. Rev. 99 (4), 1384–1414.
Ariely, D., Gneezy, U., Loewenstein, G., Mazar, N., 2009. Large stakes and big mistakes. Rev. Econ. Stud. 76 (2), 451–469.
Banerjee, A.V., Duflo, E., 2014. (Dis) organization and success in an economics MOOC. Am. Econ. Rev. Pap.Proc. 104 (5), 514–518.
Bellés-Obrero, C., 2020. Who is learning? A field experiment comparing three different incentive schemes in the same educational setting. Working Paper.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57 (1), 289–300.
Bettinger, E.P., 2012. Paying to learn: the effect of financial incentives on elementary school test scores. Rev. Econ. Stat. 94 (3), 686–698.
Braun, H., Kirsch, I., Yamamoto, K., 2011. An experimental study of the effects of monetary incentives on performance on the 12th-grade NAEP reading assessment.. Teach. Coll. Rec. 113 (11), 2309–2344.
Cacault, M. P., Hildebrand, C., Laurent-Lucchetti, J., Pellizzari, M., 2019. Distance learning in higher education: evidence from a randomized experiment.
Campos-Mercade, P., Wengström, E., 2020. Threshold incentives and academic performance. Working Paper.
Charness, G., Gneezy, U., 2009. Incentives to exercise. Econometrica 77 (3), 909–931.
Chen, J., Fonseca, M. A., Grimshaw, S. B., 2018. Using norms and monetary incentives to change behavior: a field experiment. Working Paper.
Chen, Y., Konstan, J., 2015. Online field experiments: a selective survey of methods. J. Econ. Sci. Assoc. 1 (1), 29–42.
Chen, Y., Li, S.X., Liu, T.X., Shih, M., 2014. Which hat to wear? Impact of natural identities on coordination and cooperation. Games Econ. Behav. 84, 58–86.
Clark, D., Gill, D., Prowse, V., Rush, M., 2017. Using Goals to Motivate College Students: Theory and Evidence from Field Experiments. Technical Report. National Bureau of Economic Research.
De Paola, M., Scoppa, V., Nisticò, R., 2012. Monetary incentives and student achievement in a depressed labor market: results from a randomized experiment. J. Hum. Cap. 6 (1), 56–85.
Deming, D.J., Goldin, C., Katz, L.F., Yuchtman, N., 2015. Can online learning bend the higher education cost curve? Am. Econ. Rev. Pap.Proc. 105 (5), 496–501.
Edmonds, E., Theoharides, C., 2020. The short term impact of a productive asset transfer in families with child labor: experimental evidence from the philippines. J. Dev. Econ. 146, 102486.
Feng, W., Tang, J., Liu, T. X., 2018. Dropout analysis and prediction for large scale users in MOOCs. Working Paper.
Fryer, R.G., 2011. Financial incentives and student achievement: evidence from randomized trials. Q. J. Econ. 126 (4), 1755–1798.
Fryer, R. G., Levitt, S. D., List, J., Sadoff, S., 2012. Enhancing the efficacy of teacher incentives through loss aversion: a field experiment. Working Paper.
Gneezy, U., List, J.A., 2006. Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments. Econometrica 74 (5), 1365–1384.
Gneezy, U., Meier, S., Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. J. Econ. Perspect. 25 (4), 191–209.
Gneezy, U., Rustichini, A., 2000. Pay enough or don't pay at all. Q. J. Econ. 115 (3), 791–810.
Guo, P.J., Reinecke, K., 2014. Demographic differences in how students navigate through MOOCs. In: Proceedings of the First ACM Conference on Learning @ Scale Conference. ACM, New York, NY, USA, pp. 21–30. doi:10.1145/2556325.2566247.
Hansen, J.D., Reich, J., 2015. Democratizing education? examining access and usage patterns in massive open online courses. Science 350 (6265), 1245–1248.
Hong, F., Hossain, T., List, J.A., 2015. Framing manipulations in contests: a natural field experiment. J. Econ. Behav. Organ. 118, 372–382.
Hossain, T., List, J.A., 2012. The behavioralist visits the factory: increasing productivity using simple framing manipulations. Manage. Sci. 58 (12), 2151–2167.
Hussam, R., Rabbani, A., Reggiani, G., Rigol, N., 2017. Habit formation and rational addiction: a field experiment in handwashing.
Jaggars, S.S., Xu, D., 2016. How do online course design features influence student performance? Comput. Educ. 95, 270–284.
Jalava, N., Joensen, J.S., Pellas, E., 2015. Grades and rank: impacts of non-financial incentives on test performance. J. Econ. Behav. Organ. 115, 161–196.
Karlan, D., McConnell, M., Mullainathan, S., Zinman, J., 2016. Getting to the top of mind: how reminders increase saving. Manage. Sci. 62 (12), 3393–3411.
Kizilcec, R., Piech, C., Schneider, E., 2013. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In: Proceedings of the Third International Conference on Learning Analytics and Knowledge. ACM, pp. 170–179.
Kizilcec, R.F., Brooks, C., 2017. Diverse big data and randomized field experiments in MOOCs. In: Handbook of Learning Analytics, pp. 211–222.
Lee, D.S., 2009. Training, wages, and sample selection: estimating sharp bounds on treatment effects. Rev. Econ. Stud. 76 (3), 1071–1102.
Leuven, E., Oosterbeek, H., Van der Klaauw, B., 2010. The effect of financial rewards on students' achievement: evidence from a randomized experiment. J. Eur. Econ. Assoc. 8 (6), 1243–1265.
Levitt, S.D., List, J.A., Neckermann, S., Sadoff, S., 2016. The behavioralist goes to school: leveraging behavioral economics to improve educational performance. Am. Econ. J 8 (4), 183–219.

Levitt, S.D., List, J.A., Sadoff, S., 2016. The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment. Technical Report. National Bureau of Economic Research.
Masset, E., García-Hombrados, J., Acharya, A., 2020. Aiming high and falling low: the sada-northern ghana millennium village project. J. Dev. Econ. 143, 102427.
Meier, S., 2007. Do subsidies increase charitable giving in the long run? Matching donations in a field experiment. J. Eur. Econ. Assoc. 5 (6), 1203–1222.
Mobius, M., Rosenblat, T., Wang, Q., 2016. Ethnic discrimination: evidence from china. Eur. Econ. Rev. 90, 165–177.
Patterson, R.W., 2018. Can behavioral tools improve online student outcomes? Experimental evidence from a massive open online course. J. Econ. Behav. Organ. 153, 293–321.
Qiu, J., Tang, J., Liu, T.X., Gong, J., Zhang, C., Zhang, Q., Xue, Y., 2016. Modeling and predicting learning behavior in moocs. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mininge.
Royer, H., Stehr, M., Sydnor, J., 2015. Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company. Am. Econ. J 7 (3), 51–84.
Seaton, D., Chuang, I., Mitros, P., Pritchard, D., et al., 2014. Who does what in a massive open online course? Commun. ACM 57 (4), 58–65.
Shang, J., Croson, R., 2009. A field experiment in charitable contribution: the impact of social information on the voluntary provision of public goods. Econ. J. 119 (540), 1422–1439.
Zhang, D.J., Allon, G., Van Mieghem, J.A., 2017. Does social interaction improve learning outcomes? Evidence from field experiments on massive open online courses. Manuf. Serv. Oper. Manage. 19 (3), 347–367.