**Submitted to**
**manuscript**

# Does Social Interaction Improve Service Quality? Field Evidence from Massive Open Online Education

Dennis J. Zhang, Gad Allon, Jan A. Van Mieghem
Kellogg School of Management, Northwestern University, Evanston, IL 60208

September 21, 2015

This paper studies how service providers can construct social interaction among participants and quantify the causal impact of that interaction on service quality. We focus on education and analyze whether encouraging social interaction among students improves learning outcomes in Massive Open Online Courses (MOOCs), which are a new service delivery channel with universal access at reduced, if not zero, cost.

We analyze two randomized experiments in a MOOC with 24,225 students from 183 countries. The first experiment studies large-group interaction by encouraging a random subset of students to visit the course discussion board. The majority of students treated by this experiment had higher social engagement, higher quiz completion rates, and higher course grades. Using this treatment as an instrumental variable, we estimate that one additional board visit causally increases the probability that a student finishes the quiz in the subsequent week by up to 3.5%. The second experiment studies small-group interaction by encouraging a random subset of students to conduct one-on-one discussions. Students who followed through and actually conducted pairwise discussions did increase their quiz completion rates and quiz scores by 10% in the subsequent week. Combining results from these two experiments, we provide recommendations for designing social interaction mechanisms to improve service quality.

*Key words*: Service Operations, Massive Open Online Courses (MOOCs), Social Interaction.

## 1. Introduction

Education is an important service sector of the economy that many readers of this journal participate in yet has received scant research attention in our field. The educational service since the early days of The School of Athens was built on the idea of a service provider (i.e., the instructor) delivering the content and orchestrating a discussion among participants. The service quality, measured by learning outcomes, depends critically on the quality of discussion and the instructor's ability to engage participants in the learning process (i.e., the Socratic method). The main issue with this method is that it is too expensive: college student loan debt in the U.S., reaching \$1.2 trillion dollars, is the second largest debt in the U.S. right after the mortgage debt.[1]

[1] http://www.newyorkfed.org/newsevents/news/research/2015/rp150217.html

Massive Open Online Courses (MOOCs) were proposed as a new delivery channel that makes education universally accessible at reduced, if not zero, cost. The MOOC channel allows instructors to build virtual classrooms for hundreds of thousands of students with almost zero distribution cost. A typical MOOC involves a series of video lectures, weekly assignments, and a discussion board moderated by instructors. In November 2012, the New York Times proclaimed 2012 to be the year of the MOOC.[2] The excitement was based on the promise that the world now had access to a wide variety of courses from the best universities. And indeed, as a result of efforts conducted by academia and industry, many MOOC platform providers (such as Coursera, Udacity and edX) have collaborated with top academic institutions in the world to deliver free online courses to learners anywhere on the planet.

However, the hype over MOOCs disappeared quickly. On December 12, 2013, a Washington Post article asked: "Are MOOCs Already Over?"[3] The main problem is low completion rates: a widely cited study by Christensen et al. (2013) from the Graduate School of Education at University of Pennsylvania finds that, on average, only 5 percent of students who registered for a MOOC offered by University of Pennsylvania actually completed it. Opinions diverge on whether low completion rates imply a failure of the MOOC learning model or whether they merely indicate that the community of MOOC learners is diverse and not every participant intends to complete a course. Regardless, researchers and MOOC providers are certainly interested in methods for improving learning outcomes, measured by quiz and course grades and completion rates.

In this paper, we take an operational perspective and ask a fundamental question: does large-group social engagement in discussion, a key component of the Socratic method, causally improve MOOCs' service quality? If so, do benefits accrue equally to all students or disproportionally to some groups? Moreover, can small-group social engagement (not supported in existing MOOC platforms) benefit students and improve service quality? Last, how do these findings suggest better ways to design social interactions in the MOOC channel?

A key, and noble, aspiration of service operations researchers is to improve service quality without increasing costs. Recent service operations literature acknowledges the role that social interaction among the participants of a service plays in determining its quality. Thus, while this paper focuses on education, a specific and important service sector, the question is much broader: how can service providers construct social interactions among their participants and how can one validate, using data and randomized experimentation, that social interaction is indeed the cause of the improvement in service quality?

[2] http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

[3] http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/12/12/are-moocs-already-over/

It is important to remember that simple observational studies fall short in answering our questions due to omitted variable biases: merely comparing the learning outcomes of students who visit discussion boards with those who do not can show that learning and visits are correlated, but this cannot provide convincing evidence of a causal effect because the decision to visit the discussion board can be endogenous. For instance, students who visit the discussion board more often may be more motivated and in turn more likely to finish the course. Observing then that students who are more socially engaged have better learning outcomes may reflect that higher motivation (but not social engagement per se) leads to better completion rates or grades. Establishing that social engagement has a *causal* effect (beyond correlation) on learning outcomes is the main hurdle in empirical research and requires careful statistical techniques to account for omitted variable biases, often overlooked by past MOOC research (Lamb et al. 2015 and Anderson et al. 2014). To do so, we designed and conducted two randomized experiments during our own MOOC, which is a five-week-long introductory Operations Strategy course that was first offered in April 2015 on the Coursera platform. That first offering had $24,225$ registered students of which $13,726$ visited actual course materials at least once. The first experiment focuses on large group interaction, while the second one targets small groups.

Each student received an email that previewed materials in Week 2 and asked for feedback on materials delivered in Week 1 through a survey. In the large-group interaction experiment, we encouraged a random subset of students who have filled out that survey to visit the discussion board more often through additional encouragement text and survey questions, shown in Appendix A. The experiment results provide strong evidence that social engagement in online discussion boards benefits the majority of students in MOOCs. Compared to the control group, students treated by this experiment increased their visits to, and their number of posts on, the discussion board by 26.5% and 96.8%, respectively. Treated students also had 10% higher student quiz completion rates and achieved an average total course grade of 45 points versus 39 points by the control group. Using this random encouragement assignment as an instrumental variable, we estimate that one additional board visit increases the probability that a student finishes the quiz in the subsequent week by up to 3.5%. We also find that this effect decreases over time—3.5%, 1.4%, 0.9%, and 0.8% for Week 2, 3, 4, and 5. Last, we show that social engagement in the discussion board has no effect on student quiz scores.

Coursera allows students to register for a course for free or for a fee. The latter is called the signature track and those students receive a certificate after finishing the course with a satisfactory grade. Our empirical evidence in previous experiment also suggests that, in our course, signature-track students *do not* benefit from social engagement in the discussion board. We provide two empirically justified interpretations of this no-effect result: (1) social engagement in the discussion board can only improve learning outcomes up to a certain level. Therefore, only students with low social engagement levels originally, which are more likely to be non-signature-track students, benefit from increasing social engagement in the discussion board; (2) the fee is already a powerful commitment device for signature-track students since they have to pass a certain grade hurdle to receive the certificates. Therefore any other form of motivation, such as social engagement, has less influence on their learning outcomes. This no-effect result implies that for-profit MOOC platforms may not benefit much from better designed discussion boards since their paying customers, who are motivated by the fee and usually already have high social engagement levels, are not more likely to finish the course if they are more socially engaged in the online discussion.

Moreover, to test the effect of other forms of social engagement on learning outcomes, we conducted the small-group interaction experiment. In particular, at the beginning of this course, we asked students whether they were interested in being paired up with another student for a one-on-one discussion outside the MOOC platform. During the small-group interaction experiment in Week 3, we randomly selected a subset of students who were willing to conduct this pairwise discussion and sent them email invitations to pair them up.

The results of this experiment provide empirical evidence that students who engage in one-on-one discussion improve their quiz completion rates by 10% to 7% and quiz scores by 10% to 2% in subsequent weeks. We caution, however, that the fraction of students who actually held pairwise discussions after receiving the invitation was very small (i.e., about 7%). Therefore, there seems to be no effect of receiving the invitation of this pairwise discussion on learning outcomes: the cost of one-on-one social engagement to students in MOOCs is so high that the majority of those who were invited did not follow up with their assigned partners. This indicates that, when designing mechanisms to create or improve small group social engagement in MOOCs, educators and researchers should focus on reducing the transaction cost of those mechanisms to students.

The remainder of this paper is organized as follows: Section 2 reviews the literature on MOOCs from various fields, such as computer science, education, and economics; In Section 3, we explain our course and experiment. Sections 4 and 5 provide our analysis of the large-group and small-group interaction experiments, respectively. We discuss implications of our analysis for better designing the MOOC channel in Section 6 and conclude the paper in Section 7.

## 2. Literature Review

Massive Open Online Courses have drawn considerable attention from various literature, including Computer Science (Anderson et al. 2014), Education (Breslow et al. 2013), Economics (Banerjee and Duflo 2014), and Operations Management (Terwiesch and Ulrich 2014). Many papers point out that the main problem associated MOOCs is their low completion rates (Kizilcec et al. 2013, Breslow et al. 2013, and Anderson et al. 2014). Breslow et al. 2013 report that, among 155,000 students who enrolled in a course from Harvard on edX, only 7,100 of them finished it (4.5%), while Anderson et al. (2014) show that, on average, only 2% to 10% of students finished the courses among all students who have signed up in three computer science courses from Stanford on Coursera. Christensen et al. (2013) demonstrate that among 14 MOOCs held by University of Pennsylvania, only 2% to 14% of students who at least interacted with the course once "survived" till the last week. While *New York Times* proclaimed 2012 to be the year of MOOC,[4] it later "demythified" MOOCs and credited their failures to low completion rates.[5]

Because of the importance of completion rates in MOOCs, there is a considerable amount of literature that tries to understand the drivers of drop-outs and how to increase completion rates. In particular, many papers believe that the key to improve completion rates is social engagement among students in the discussion board (Kizilcec et al. 2013, Anderson et al. 2014, and Lamb et al. 2015). For instance, Anderson et al. (2014) design a new badge system in discussion boards to reward students' participation by assigning them different badges, and show that this badging system increases students' social engagement levels. Lamb et al. (2015) study several reminding methods and their impacts on discussion board participation rates of students. Kizilcec et al. (2013) carefully analyzes various types of students on MOOCs and conclude from observational data that "participation on the board creates a positive feedback loop for some learners, as they are provided with social and informational inputs that help them stay on their trajectory towards completion."

In this paper, instead of studying how to design a more engaging discussion board, we take a step back and ask the following questions: does social engagement in the discussion board *causally* increase learning outcomes in MOOCs? If so, which groups of students are affected the most? Do other forms of social engagement outside the MOOC platform benefit students? In order to precisely answer theses causality questions, we design and implement two field experiments to conduct our study. Therefore, this is one of the few papers that conducts large-scale field experiments in operations settings (Buell and Norton 2011).

Since our paper focuses on students' drop-outs in a service system, we are connected to a large body of literature in Operations Management studying customers' abandonment and attrition behaviors in service systems (Naor 1969, Allon et al. 2011, and Veeraraghavan and Debo 2011). The main difference between our study and past literature is that, while past papers often assume that customers individually make rational attrition and abandonment decisions (Allon et al. 2011, Aksin et al. 2013 and Yu et al. 2013), we assume that students are constantly influenced by each

[4] http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

[5] http://www.nytimes.com/2014/11/02/education/edlife/demystifying-the-mooc.html

other through social interactions. We thus focus on how students' learning outcomes are affected by their social engagement.

Finally, our paper is related to the emerging literature in operations management which studies various operations problems in the presence of social interactions among agents. Candogan et al. (2012) study a dynamic pricing problem when customers' consumption depends on their friends' consumptions in a social network. Jing (2011) and Papanastasiou and Savva (2014) provide insights into dynamic pricing strategies for new products when customers may strategically delay their purchasing decisions due to social learning effects. While all papers mentioned above focus on theoretical properties of operating systems in the presence of social interactions, we use field experiments to offer empirical evidence of social interaction effects in service systems.

## 3. Background and Experiment Setup

### 3.1. Coursera Platform and Our Course

Coursera is a for-profit education technology company that partners with universities to offer MOOCs. Courses on Coursera cover a wide variety of subjects, such as physics, engineering, humanities, medicine, biology, social sciences, and business. As of now, Coursera has partnered with more than 120 top institutions around the world and offered 571 distinct courses. These courses in total had 22.2 million enrollments from students in 190 countries, who, in aggregation, spent 343 million minutes of learning on Coursera's platform (Coursera 2015). This makes Coursera one of the largest MOOC providers, along with edX and Udacity.

The current business model of Coursera is to offer paid verified certificates to students: when taking a course, students can choose to take the course for free or pay a fee and enroll in the *signature-track*. Students who are enrolled in the signature-track and successfully pass the completion requirements will get verified certificates signed by both Coursera and the institution that designs the course. Students who pass the completion requirement but do not pay will often receive statements of accomplishment which are not signed nor verified by either Coursera or partner institutions. Most courses on Coursera cost between \$39 to \$129 to enroll in the signature-track. On top of the signature-track, Coursera also offers *specializations*, which is a set of coursers and a final project in one area. Specializations normally include 4 to 9 courses and cost between \$500 to \$1000. According to a post by an instructor,[6] the Data Science Specialization from John Hopkin's University, one of the most popular specializations on Coursera, had 71,589 signature-track enrollments in 2014 with \$3.5 million revenue.

Our course is a five-week-long business course inspired by our five-week-long EMBA elective course. Each week, there are four video lectures and a quiz. Video lectures are about 5 to 10 minutes,

[6] http://simplystatistics.org/2015/02/05/johns-hopkins-data-science-specialization-top-performers/

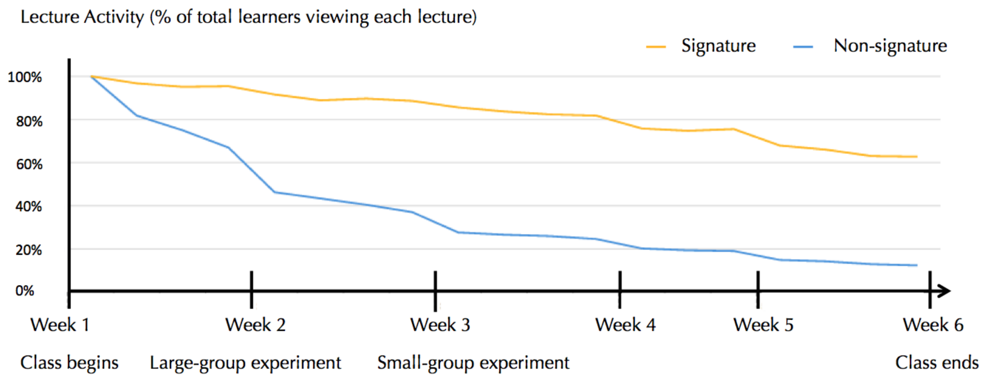



Figure 1: Fraction of students that viewed each video over time. (100% = 9,812 students.)

making the lecture time in each week between 20 to 40 minutes. Students also are required to do a case study or a project in each of the first four weeks. Students can choose between the case study, where they are asked to analyze the case provided by us, or a project, where we provide guiding questions for students to analyze their own companies/organizations. The course has a total score of 100; the minimal passing grade is 70 and students pass with distinction if they achieve 90 or above. During its first offering in April 2015, the course had in $24,225$ registered students from 183 different countries among which 480 enrolled in the signatures track; $13,726$ registered students actually visited the actual course website at least once; 9,812 watched at least one lecture; 4,221 submitted at least one exercise; and 3,598 browsed the discussion boards at least once. Figure 1 shows the fraction of the 9,812 learners who viewed each video over time. Out of the 433 students who completed the course, 233 were signature-track students.

### 3.2. Large-group Interaction Experiment

In order to identify the causal effect of social engagement in the discussion board on learning outcomes, we used an "encouragement design," which was widely used in Economics literature (Duflo and Saez 2003). Our large-group interaction experiment encourages a random subset of students to visit the discussion board through randomly assigned survey questions and notifications. The treatment in this experiment is whether a student who completed the survey received this encouragement.

At the beginning of Week 2, each student received an email that previewed materials in Week 2 and asked for feedback on materials delivered in Week 1 through a survey. Each email includes a unique embedded link to the survey website that contains the student user id on Coursera. This URL embedding technology allows us to record students' user ids along with their survey responses. Therefore, we can match students' survey responses on our survey website with their social engagement levels and learning outcomes on Coursera through their user ids. Moreover, when mentioning the survey in the email, we use embedded HTTP link instead of directly displaying the URL, which hides the URL from students and in turn make it less likely for them to feel that they are in an experiment (even though they explicitly consented to participate in an experiment when signing up for the course).

When students arrived at the survey website, we randomly assigned them to two versions of surveys: treatment version and control version. The control version of the survey is one-page-long with four questions asking students' feedback in the first week. The treatment survey is two-pages-long: the first page is the same as that in the control survey, while the second page encourages students to go to the discussion board. The encouragement consists of two parts: First, there is bold text at the top of the page that explicitly asks students to visit the discussion board to get most out of our course. Second, there are two questions that ask students how many posts and visits they had in the first week. These questions were designed to "shame" students if their social engagement levels in the first week did not meet the described standard in the question. The encouragement text and survey questions served as a remainder for students and set norms for desired levels of social engagement in the discussion board. The second page of the treatment version is included in Appendix A.

There are about $3,500$ students who have at least interacted with the course once after the first week. Among those students, 335 of them, about 10% of the total population, completed the survey. 175 students received the treatment version, while 160 students were directed to the control version. If focusing on signature-track or non-signature-track students separately, we do not observe any distributional differences in observed characteristics between the total population and our experiment participants. In the rest of paper, we will refer to those 335 students as experiment participants.

Last, let us discuss alternative encouragement design that one could consider. In particular, one may think that we should use direct randomized treatments, such as sending out two versions of emails, so that our sample can be the total population (i.e., $3,500$ students) instead of a sub-sample of it. There are three problems associated with direct treatment encouragement design. First, in March 2015, the Coursera platform did not allow instructors to send emails to a random subset of students. In fact, there is no infrastructure in Coursera to conduct any form of A/B testing. Therefore, the only randomization that we could do was to randomize students at the landing page of the survey based their user ids embedded in URLs. Second and more importantly, past research in Education has shown that asking questions through surveys is much more effective in incentivizing students to participate in the discussion board in MOOCs than sending emails (Lamb et al. 2015). Third, as we will discuss in our next experiment, directly assigning treatments to a large population on MOOC will result in severe compliance problems: majority of students who receive the email will simply ignore it, which decreases the power of the experiment significantly.

### 3.3. Small-group Interaction Experiment

At the beginning of our course, we conduct a demographic survey to ask students whether they would like to participate in a one-on-one discussion with their fellow "classmates." Out of 2,436 students who attempted the demographics survey, 1,477 students were willing to participate. Therefore, we refer to those 1,477 students as our participants in the small-group interaction experiment. The survey also asked those student for their preferred dates and communication channels to later match them with similar preferences.

We matched students into pairs (to conduct the pairwise discussion) based on three criteria: (1) they are within 3 timezones of each other; (2) they have at least one common preferred communication channel; and (3) they have at least one common preferred day in a week. Using various randomized algorithms (i.e., genetic search and hill climbing), we were able to match 600 groups out of those 1,477 students. We randomly selected 473 pairs that we invited for this one-on-one discussion.

At the middle of the second week, we sent out an email to those randomly selected 473 pairs of students (the detailed email is included in Appendix A) to inform them of (1) the contact email of their partner, (2) their preferred communication channel, (3) their preferred day in a week, and (4) *they needed to schedule and conduct the one-on-one discussion during the third week*. This email also included a link for students to report whether they had successfully scheduled the discussion and, if not, whether they would like to be re-matched.

Among the selected 473 pairs, 27 pairs (i.e., 54 students) reported to have successfully conducted the one-on-one discussion. Among all other groups, 75 students reported back that their partners did not respond. We did not receive any responses from the remaining 817 students that received the invitation email. In Section 5, we discuss the detailed classification of students based on their responses.

## 4. Large-group Interaction Experiment: Analysis and Findings

We first discuss our measures of students' social engagement levels and learning outcomes. We then provide empirical evidence that the encouragement treatment in our large-group interaction experiment has significant effects on students' social engagement levels as well as their learning outcomes. Last, using our treatment as an instrumental variable, we estimate the effects of social engagement levels of students on their learning outcomes.

### 4.1. Measurements and Summary Statistics

In this section, we discuss the direct effects of our encouragement treatment on students' social engagement levels in the discussion board and their learning outcomes. The discussion board on Coursera normally consists of a sequence of threads: each thread starts with an initial post from either students or instructors, which can then be followed by a sequence of posts. Each student can take four actions in the discussion board: (1) read, (2) post, (3) comment, and (4) vote. First, a student can *read* each thread. Second, a student can *post* a new post in a thread as shown in Zone 1 of Figure 2. Moreover, a student can *comment* on an existing post, as indicated in Zone 2 of Figure 2. The difference between comments and posts is subtle: posts are supposed to be a reply or a contribution to the whole thread, while comments represent replies to a specific existing post in a thread. However, we find that students often use them interchangeably. Last, a student can up-vote or down-vote other student's posts by clicking the button as shown in Zone 3 of Figure 2.

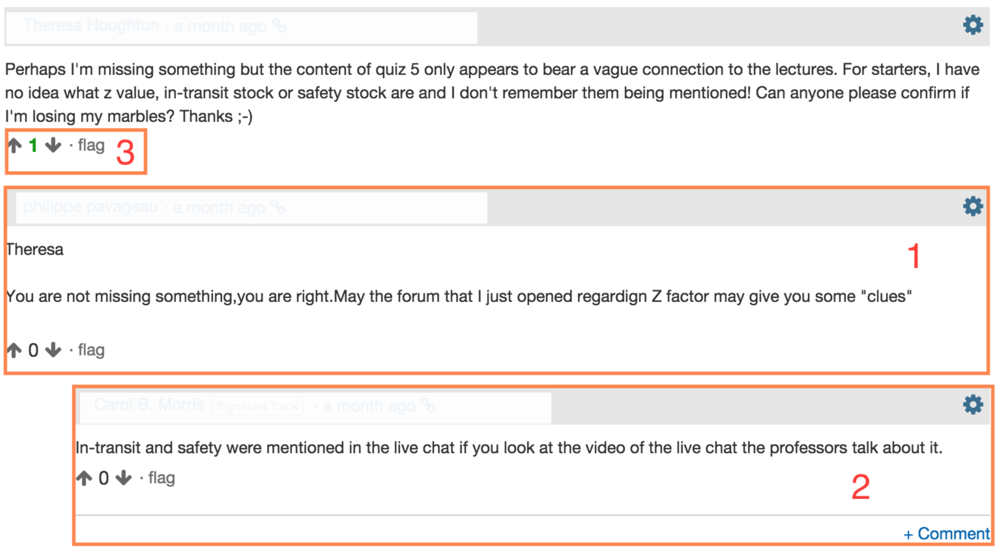



Figure 2: Actions that student can take in the discussion board

**Social Engagement:** Given the structure of the board and the actions of students, we define social engagement levels of a student in each time period $j$ with two measures: First, we use $v_{ij}$ to denote the total number of visits to the board that student $i$ has during the time interval $j$. In other words, $v_{ij}$ does not only measure the frequency that a student visits the board but also how actively the student is in each visit (i.e., how many threads she has read and posted). Second, we define $p_{ij}$ as the number of posts and comments that student $i$ started during time interval $j$. As aforementioned, we use the sum of numbers of posts and comments since students often use posts and comments interchangeably. These two definitions are widely used in past literature studying social engagement in the discussion board of MOOCs (Anderson et al. 2014).

**Learning Outcomes:** Similar to many other Coursera courses, our Coursera course consists of 4 video lectures and one quiz each week. Students can attempt the quiz in a week at most three times, and their maximum scores among all attempts for one quiz are used as their final scores for that quiz. We divide students' learning outcomes in this course into two parts: quiz completion rates and quiz scores in a week. We define $q_{ij} = 1$ if the student has attempted the quiz at least once in Week $j$. Moreover, we use $g_{ij}$ to represent student $i$'s quiz score at Week $j$.

Table 1 displays the summary statistics associated with the encouragement design, broken into four columns. Column (1) has the statistics for students who have received the treatment, while Column (2) provides the statistics for students who are in the control group. Column (3) shows the statistics for the entire student population who has participated in the experiment.

Panel A of Table 1 represents the demographics characteristics of each group. Each student's demographics consists of 3 variables based on data from Coursera's internal system: (1) signature track status, (2) English locale, and (3) email announcement status. The signature-track status is 1 if the student pays $69 to enroll in the signature-track. A student's English locale status is 1 if her operating system has English as its default language. Last, the email announcement status equals to 1 if the student agrees to receive email announcements from Coursera courses. Because the treatments are randomly assigned, the means of observed characteristics such as signature track status, system locale, and email announcement status, are not statistically different among the three groups.

Panel B of Table 1 shows that our encouragement strategy had a dramatic effect on the social engagement levels of students: the treated students on average visited the board 17.91 times and posted 1.06 times in Week 2 (i.e., the week immediately after the treatment). In contrast, students in the control group on average only visited the board 13.58 times and had 0.66 posts during that week. Moreover, we can see, that the difference between the social engagement levels of students in treatment and control groups do not disappear until Week 5. It really surprised us that such a simple nudge, as in our encouragement design, had a profound effect on students' social engagement levels in the discussion board in the long run.

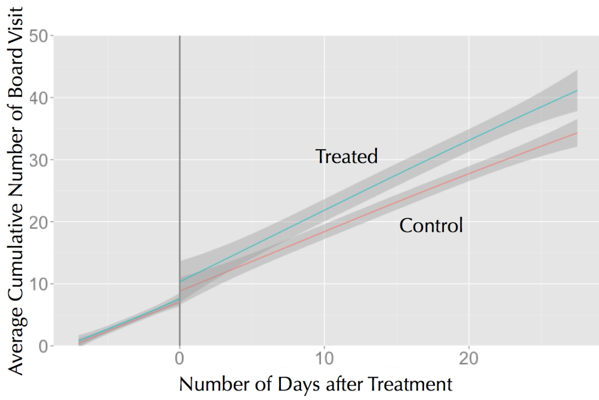
Panel C presents the effect of our encouragement treatment on learning outcomes. It shows that students in the treatment group are more likely to complete the quiz after the treatment: in Week 2 (i.e., right after the treatment), 70% of students in the treatment group completed the quiz versus only 56% of students in the control group. Moreover, the effect of our treatment on students' completion rates also persists until Week 5. However, students in the treatment group do not necessarily have higher quiz scores than their counterparts in the control groups after the treatment.

Last, Figure 3 visualizes the effect of our encouragement treatment on students' social engagement levels. Panel (a) of the figure displays average cumulative numbers of visits for students in the treatment and control groups over time. (0 on the horizontal axis represents the time when the treatment was assigned.) It is evident that, before the assignment, these two groups have similar cumulative number of visits, while, after the treatment, treated students have significantly higher

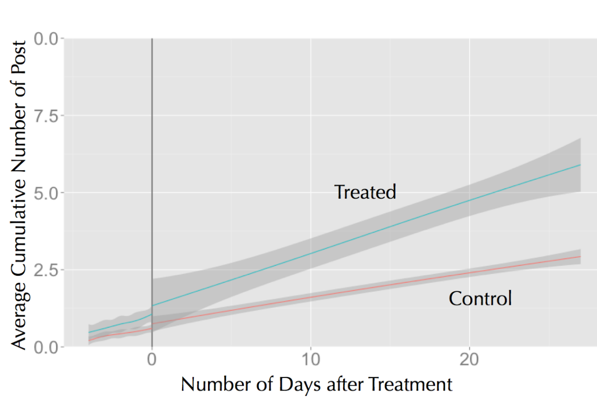
| | Experiment Participants | | |
|---|---|---|---|
| | **Treated Group** (1) | **Untreated Group** (2) | **All Participants** (3) |
| | **Panel A: Demographics Characteristics** | | |
| Signature Track | 0.24 | 0.23 | 0.23 |
| | (0.03) | (0.03) | (0.02) |
| English Locale | 0.53 | 0.53 | 0.53 |
| | (0.04) | (0.04) | (0.03) |
| Email Announcement | 0.99 | 0.99 | 0.99 |
| | (0.01) | 0.01 | (0.00) |
| | **Panel B: Social Engagement (Week 2 - 5)** | | |
| Number of Visits in Week 2 | 17.91 | 13.58 | 15.84 |
| | (2.89) | (1.82) | (1.78) |
| Number of Visits in Week 3 | 17.41 | 13.47 | 15.53 |
| | (3.25) | (2.30) | (2.02) |
| Number of Visits in Week 4 | 16.85 | 12.66 | 14.84 |
| | (5.44) | (2.40) | (3.06) |
| Number of Visits in Week 5 | 16.94 | 14.19 | 15.63 |
| | (4.20) | (2.91) | (2.60) |
| Number of Posts in Week 2 | 1.06 | 0.66 | 0.87 |
| | (0.21) | (0.10) | (0.12) |
| Number of Posts in Week 3 | 1.41 | 0.55 | 1.00 |
| | (0.47) | (0.11) | (0.25) |
| Number of Posts in Week 4 | 1.37 | 0.62 | 1.01 |
| | (0.69) | (0.14) | (0.37) |
| Number of Posts in Week 5 | 1.10 | 0.68 | 0.90 |
| | (0.27) | (0.13) | (2.16) |
| | **Learning Outcomes (Week 2 - 5)** | | |
| Quiz Completion in Week 2 | 0.70 | 0.56 | 0.63 |
| | (0.03) | (0.04) | (0.03) |
| Quiz Score in Week 2 | 3.63 | 3.74 | 3.68 |
| | (0.14) | (0.15) | (0.10) |
| Quiz Completion in Week 3 | 0.53 | 0.44 | 0.49 |
| | (0.04) | (0.04) | (0.03) |
| Quiz Score in Week 3 | 2.6 | 2.68 | 2.63 |
| | (0.11) | (0.11) | (0.08) |
| Quiz Completion in Week 4 | 0.40 | 0.36 | 0.38 |
| | (0.04) | (0.04) | (0.03) |
| Quiz Score in Week 4 | 3.96 | 3.83 | 3.90 |
| | (0.16) | (0.16) | (0.11) |
| Quiz Completion in Week 5 | 0.35 | 0.30 | 0.34 |
| | (0.04) | (0.04) | (0.03) |
| Quiz Score in Week 5 | 3.16 | 3.15 | 3.16 |
| | (0.12) | (0.13) | (0.09) |
| Total Score | 44.46 | 39.27 | 41.98 |
| | (2.59) | (2.90) | (1.94) |
| Course Completion Rate | 0.29 | 0.27 | 0.28 |
| | (0.04) | (0.03) | (0.02) |
| Observations | 175 | 160 | 335 |

1. Standard errors are in parentheses.
2. Signature tract is 1 if the student enrolled in the Signature Track. English locale is 1 if the student's operating system is in English. Email announcement is 1 if the student agrees to receive email announcements from the course.
3. The quiz score in each week is the average of maximum quiz score across all students who attempted the quiz.

Table 1: Descriptive statistics of the large-group interaction experiment, by groups.




(a) Cumulative number of visits




(b) Cumulative number of posts

Figure 3: Cumulative numbers of visits and posts before and after the treatment

Panel (a) shows the average cumulative number of visits for students in treatment (green line) and control (red line) groups before and after the treatment. The raw data is the time of each visit for all students. Since the visits can happen any time, the lines are drawn by fitting a generalized additive models (GAM) of average cumulative visits over time. The shades represent the standard errors of the average cumulative visits at any time. Panel (b) represents the average cumulative number of posts for students in treatment and control groups. Similarly, the line is the extrapolated average cumulative number of posts using GAM model while the shaded zone represents standard errors.

cumulative number of visits. Moreover, it is easy to see that the differences between average cumulative numbers of visits of students in those two groups are increasing over time. This shows that the encouragement treatment has a long-lasting effect on students' social engagement levels. Similarly, Panel (b) of Figure 3 plots average cumulative number of posts for students in the treatment and control groups before and after the treatment. It also demonstrates that the encouragement has a significant and persistent effect on students' cumulative number of posts.

### 4.2. Reduced-form Differences

We first analyze the effects of our encouragement treatment on students' social engagement levels and learning outcomes. We consider a set of simple reduced-form regression specifications. Let $D_i$ denote the treatment status of student $i$ with $D_i = 1$ representing that student $i$ has received the encouragement treatment. $S_i$ is the dummy for enrolling in the signature track. The average effects of our treatment on the number of visits can be captured by the following specifications:

$$v_{ij} = \text{Poisson}(\alpha_1 + \beta_1 D_i + D_i * S_i + S_i + W_j + \epsilon_{ij}), \tag{1}$$

where $v_{ij}$ is the number of visits for student $i$ at Week $j$, and $W_j \in \{2, 3, 4, 5\}$ represents the week-level control. Notice that we have included the interaction between our treatment effect ($D_i$) and the signature status ($S_i$) since we assume that non-signature track and signature-track students may experience different treatment effects.[7]

Similarly, since $p_{ij}$ represents the number of posts that student $i$ has in Week $j$, we model the treatment effects on students' posting behaviors using the following specification:

$$p_{ij} = \text{Poisson}(\alpha_2 + \beta_2 D_i + D_i * S_i + S_i + W_j + \epsilon_{ij}). \tag{2}$$

The estimates of interest are $\beta_1$ and $\beta_2$. $\beta_1 > 0$ shows that the treatment has a positive effect on students' probabilities to visit the board, while $\beta_2 > 0$ demonstrates that the treatment encourages students to post more in the discussion board. The regression also controls for other observed characteristics (i.e., English locale and email announcement).[8] All standard errors in the analysis are clustered at the week-level to account for weekly fluctuations, which can be caused by different difficulty levels across time.

The estimates for $\beta_1$ and $\beta_2$ are reported in Column (1) and (2) in Panel A of Table 2. It shows that receiving the encouragement treatment significantly increases students' numbers of visits and posts in the discussion board after the treatment. Specifically, receiving the treatment increases a students' weekly number of visits to the discussion board by 2.95 and weekly number of posts in the discussion board by 0.55. Notice that the impact of our treatment is large in relative terms: the average weekly number of visits and posts for all students are 15.46 and 0.94, while the marginal effect of the encouragement treatment on the weekly number of visits and posts are 2.95 and 0.55. This is also confirmed in Figure 3 where the cumulative number of posts at the end of the course for students in the treatment group is around 5, while that for students in control group is only 2.5. Moreover, these results set the foundation for us to use our randomly assigned encouragement treatment as an instrumental variable to estimate the average effect of students' social engagement on their learning outcomes since these results show that our randomly assigned treatment is correlated with students' social engagement levels.

Last but not least, we also note that the number of visits and posts for students may be over-dispersed (i.e., the standard deviations are larger than the mean) and have many zeros (i.e., on average more than 30% of students have zero visits and posts during a week). In order to account for potential biases introduced by those distributional characteristics into our Poisson counting models, we also use other specifications. In particular, we do not find any statistical different results when using zero-inflated Poisson models (models to account for zero-inflation) or Negative Binomial models (models to account for over dispersion) to estimate Equation (1) and (2).

[7] Our results do not qualitatively change if we drop the interaction term.

[8] Since our treatment is randomly assigned, it is unlikely that we have omitted variable biases. In this case, adding controls is merely a method to increase the efficiency of our estimators.

| | Dependent variable | | | |
|---|---|---|---|---|
| | Social Engagement | | Learning Outcomes | |
| | **Number of Visits** (1) Poisson | **Number of Posts** (2) Poisson | **Quiz Completion Rate** (3) Probit | **Quiz Score** (4) OLS |
| **Panel A: Average Effect of Social Engagement and Learning Outcome (Week 2 - 5)** | | | | |
| Treatment | $0.243^{***}$ | $0.405^{***}$ | $0.237^{***}$ | 0.092 |
| | (0.068) | (0.103) | (0.065) | (0.133) |
| Signature Track | $1.660^{***}$ | $1.660^{***}$ | $1.949^{***}$ | $0.143^{*}$ |
| | (0.251) | (0.299) | (0.073) | (0.040) |
| Observations | 1340 | 1340 | 1340 | 617 |
| **Panel B: Average Effect of Social Engagement and Learning Outcome (Week 1)** | | | | |
| Treatment | 0.003 | 0.474 | 0.272 | 0.64 |
| | (0.262) | (0.382) | (0.183) | (0.050) |
| Signature Track | -0.038 | 0.057 | (omitted) | -0.017 |
| | (0.418) | (0.433) | (omitted) | (0.092) |
| Observations | 335 | 335 | 335 | 287 |

(1) Standard errors are in parentheses.
(2) . $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
(3) All regressions control for the interaction effect of the treatment and signature track status, email announcement status, and language.
(4) Standard errors are corrected for clustering at the weekly level.

Table 2: Reduced-form estimates of our treatment effects on social engagement levels and learning outcomes in the large-group interaction experiment.

We then move to the effect of our encouragement treatment on learning outcomes. Let $q_{ij} = 1$ represent that student $i$ has finished the quiz in Week $j$. And let $s_{ij}$ denote student $i$'s quiz score in Week $j$ if she has completed the quiz. We model students' quiz completion rates and scores as follows:

$$q_{ij} = \text{Probit}(\alpha_3 + \beta_3 D_i + D_i * S_i + S_i + W_j + \epsilon_{ij}), \tag{3}$$

and

$$s_{ij} = \alpha_4 + \beta_4 D_i + D_i * S_i + S_i + W_j + \epsilon_{ij}. \tag{4}$$

We are interested in $\beta_3$ and $\beta_4$. $\beta_3 > 0$ means that the encouragement treatment has a positive effect on students' quiz completion rates, while $\beta_4 > 0$ represents that the treatment has a positive effect on students' quiz scores. The results are reported in Column (3) and (4) of Panel (A) in Table 2. Consistent with the summary statistics in Table 1, the encouragement treatment has a significantly positive effect on students' quiz completion rates: students who received the treatment were 5.54% more likely to complete the weekly quiz in subsequent weeks. Again, this impact is substantial compared to 46.1%, the average quiz completion rates of all students in the sample from Week 2 to 5. Moreover, we do not observe a significant effect of the treatment on the scores that students achieve in their quizzes. This is also consistent with the results from Table 1, where in some weeks (i.e., Week 2 and Week 3), students in the treatment group even had lower average scores than students in the control group.

Table 1 showed no distributional differences between observed characteristics of students in treatment and control groups. One may think that there may be unobserved characteristics that are correlated with both our random assignments and learning outcomes, which in turn could bias our estimators. We provide a simple Placebo test to address this issue. In particular, we estimate the treatment effects on students in Week 1, before the treatment is even assigned, with Equations (1), (2), (3), and (4). If our results are robust, we should expect to find that there is no effects of our treatment in Week 1. Panel B of Table 2 demonstrates the results. As expected, whether students would be selected into the treatment group *in the future* did not have any effects on students' social engagement levels or learning outcomes in the first week of the course.

### 4.3. Estimating the Causal Effects of Social Engagement in Discussion Boards

In this section, we focus on the effect of social engagement in discussion boards on learning outcomes. Notice that $v_{ij}$ measures student $i$'s social engagement levels in week $j$. We cannot directly estimate the effect of social engagement levels (i.e., $v_{ij}$) on learning outcomes (i.e., $q_{ij}$ and $s_{ij}$) since students can take the quiz any time in a week and the social engagement measures may include students' activities in the discussion board after they complete the quiz. For example, if student 1 finished the quiz in Week 2 on Wednesday, while she visited the discussion board on Tuesday and Friday. In this case, $v_{12} = 2$, while only the visit on Tuesday within $v_{12}$ should causally affect the student's completion rate.

Therefore, to account for this issue, we re-define the measure of social engagement (i.e., weekly number of visits) in two ways: (1) $\bar{v}_{ij}$: student $i$'s visits to the discussion board before she finished the quiz in Week $j$. For students who have not finished quiz in Week $j$ at all, we impute their times of finishing the quiz in Week $j$ as the average time of all students who have finished the quiz; (2) $\tilde{v}_{ij}$: student $i$'s visits to the discussion board before the beginning of Week $j$.

The first measure is more efficient since it includes both social engagement levels prior to Week $j$ and in Week $j$ prior to the time student took the quiz, However the imputation may introduce biases (White and Carlin 2010). While the second measure is less efficient since it only includes visits to the board prior to the week of the quiz, it does not rely on imputation. Therefore, we present estimated effects based on both measures.

We posit the following set of specifications to explain our instrumental variable (IV) strategy of estimating the effects of social engagement on learning outcomes:

$$q_{ij} = \text{Probit}(\alpha + \gamma_j^1 V_{ij} + X_i + u_{ij}), \tag{5}$$

and

$$s_{ij} = \alpha + \gamma_j^2 V_{ij} + X_i + u_{ij}, \tag{6}$$

where $V_{ij} \in \{\bar{v}_{ij}, \tilde{v}_{ij}\}$ represent either of the two measures of social engagement, and $X_i$ is demographics control for student $i$ (i.e., signature track status, English locale, and email announcement statue). These equations state that an individual's quiz completion rate and quiz score in Week $j$ are potentially influenced by her social engagement levels prior to the quiz. More importantly, we assume that this influence may be heterogeneous across time. In other words, the average effect in Week $j$ may be different from that in Week $j+1$ (i.e., $\gamma_j \neq \gamma_{j+1}$). We have assumed, at least for now, that the average treatment effect is homogeneous across individuals (i.e., $\gamma_{ij} = \gamma_{kj}\ \forall i, j, k$).

Evidently, directly estimating the effect of social engagement levels on learning outcomes through Equation (5) may introduce omitted variable biases. The omitted variable biases are caused by non-controlled variables (i.e., omitted) that are correlated with students' social engagement levels as well as learning outcomes, such as students' motivation and ability. To address this causal identification issue, we use aforementioned randomized encouragement treatment (i.e., $D_i$) as an instrumental variable (IV). This is a typical IV setup with homogeneous treatment effects. Therefore, we need to check both inclusion and exclusion assumptions of our instruments:

Assumption 1. *Inclusion restriction assumption: $D_i$ is correlated with $V_{ij}$ (i.e., $\bar{v}_{ij}$ and $\tilde{v}_{ij}$).*

This assumption states that our encouragement treatment indeed incentivized students to visit the board more often. Both the summary statistics in Table 1 and the estimated results in Table 2 indicate that this assumption is satisfied. The second assumption that we need to satisfy is exclusion restriction:

Assumption 2. *Exclusion restriction assumption: $u_{ij}$ is independent of $D_i$.*

This assumption means that the encourage treatment only affects learning outcomes through students' social engagement levels. This assumption can be divided into two parts: First, $D_i$ is not correlated with any other variables (not in $X_i$) that affect learning outcomes directly, such as students' motivation and ability. This part is satisfied directly from the construction of $D_i$ since we randomly assigned treatments to students. Second, the treatment does not directly affect learning outcomes for treated students conditional on their social engagement levels (i.e., $cov(q_{ij}, D_i|V_{ij}) = 0$ and $cov(s_{ij}, D_i|V_{ij}) = 0$). Since we did not explicitly mention any other motivational sentences except telling students to attend the discussion board more often, it seems reasonable to assume the second part. Moreover, we did not give any additional information about the course materials in the treatment page.

With these two assumptions, Equations (5) and (6) can be estimated with standard IV setup, and the average treatment effects of students' visit to the discussion board on their quiz completion rates and scores can be identified. The estimates of interest are $\gamma_j^i$ where $i \in \{1,2\}$ and $j \in \{1,2,3,4\}$.

| | **Dependent variable** | | | |
|---|---|---|---|---|
| | **IV specification** | | **Reduced-form specification** | |
| | **Quiz completion rate** (1) Probit | **Quiz score** (2) Linear | **Quiz completion rate** (3) Probit | **Quiz score** (4) Linear |
| **Panel A: Average Effect of Social Engagement on Learning Outcomes (Week 2)** | | | | |
| Number of Visits | $0.0899^{***}$ | $-0.0726$ | $0.0287^{***}$ | 0.0023 |
| | (0.0131) | (0.113) | (0.0110) | (0.0037) |
| Observations | 335 | 212 | 335 | 212 |
| **Panel B: Average Effect of Social Engagement on Learning Outcomes (Week 3)** | | | | |
| Number of Visits | $0.0343^{***}$ | 0.0146 | $0.0166^{***}$ | 0.0014 |
| | (0.0055) | (0.0246) | (0.0047) | (0.0013) |
| Observations | 335 | 163 | 335 | 163 |
| **Panel C: Average Effect of Social Engagement on Learning Outcomes (Week 4)** | | | | |
| Number of Visits | $0.0215^{***}$ | 0.0312 | $0.0157^{***}$ | $0.0018^{*}$ |
| | (0.0048) | (0.141) | (0.0034) | (0.0009) |
| Observations | 335 | 128 | 335 | 128 |
| **Panel D: Average Effect of Social Engagement on Learning Outcomes (Week 5)** | | | | |
| Number of Visits | $0.0203^{***}$ | 0.0005 | $0.0179^{***}$ | $0.0016^{*}$ |
| | (0.0043) | (0.0163) | (0.0033) | (0.0006) |
| Observations | 335 | 114 | 335 | 114 |

(1) Standard errors are in parentheses.
(2) . $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
(3) All regressions control for signature track status, email announcement status, and language.

Table 3: IV-estimated effect of social engagement in the discussion board ($\bar{v}_{ij}$) on learning outcomes

$\gamma_j^1$ is the average treatment effect of board visits on students' quiz completion rates in Week $j$, while $\gamma_j^2$ represents the average treatment effect of board visits on students' quiz scores in Week $j$.

Table 3 and 4 show the estimated results related to both measures of social engagements. In particular, Column (1) of Table 3 and 4 show that, under both measures, the effects of board visits (i.e., social engagement in the discussion board) on students' quiz completion rates are significantly positive. In particular, in our five-week-long course, one more cumulative visit to the discussion board increases students' completion rates of quizzes by 3.5%, 1.4%, 0.9%, and 0.8% in Week 2, 3, 4, and 5 respectively. Interesting, the effects of social engagement levels on learning outcomes are decreasing over time.

It is also useful to compare the estimated effects through IV approach with the estimates from the naive reduced-form specifications in Column (3) of Table 3 and 4. The reduced-form estimators reported in Column (3) are consistently smaller than the IV estimators in Column (1). This shows that a naive reduced-form estimator of the effects of social engagement levels on quiz completion rates is likely to be biased downwards, which suggests that there exist confounding factors that are positively correlated with quiz completion rates and negatively correlated with social engagement levels.

Moreover, it is evident from Column (2) of Table 3 and 4 that more board visits have no effects on students' quiz scores throughout the course. This is not surprising for two reasons: (1) students are

| | Dependent variable | | | |
|---|---|---|---|---|
| | IV specification | | Reduced-form specification | |
| | Quiz completion rate (1) Probit | Quiz score (2) Linear | Quiz completion rate (3) Probit | Quiz score (4) Linear |
| **Panel A: Average Effect of Social Engagement on Learning Outcomes (Week 3)** | | | | |
| Number of Visits | $0.0254^{***}$ | $-1.281$ | $0.0172^{***}$ | $0.0012^{*}$ |
| | (0.0042) | (0.2002) | (0.0036) | (0.0006) |
| Observations | 335 | 163 | 335 | 163 |
| **Panel B: Average Effect of Social Engagement on Learning Outcomes (Week 4)** | | | | |
| Number of Visits | $0.0157^{***}$ | 0.0618 | $0.0096^{***}$ | $0.0016^{*}$ |
| | (0.0059) | (0.568) | (0.0025) | (0.0007) |
| Observations | 335 | 128 | 335 | 128 |
| **Panel C: Average Effect of Social Engagement on Learning Outcomes (Week 5)** | | | | |
| Number of Visits | $0.0132^{***}$ | 0.0004 | $0.0102^{***}$ | $0.0013^{***}$ |
| | (0.0024) | (0.0114) | (0.0023) | (0.0005) |
| Observations | 335 | 114 | 335 | 114 |

(1) Standard errors are in parentheses.
(2) . $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
(3) All regressions control for signature track status, email announcement status, and language.

Table 4: IV-estimated effect of social engagement in the discussion board ($\tilde{v}_{ij}$) on learning outcomes

not allowed to discuss quiz related problems in the discussion board; (2) our board lacks sufficient tagging and recommendation functionalities; therefore, it is difficult for students to extract needed information from it. Column (4) of Table 3 and 4 provides the estimates of social engagement levels on quiz scores through simple OLS estimators. The OLS coefficients are significantly positive, which may lead researchers to an incorrect conclusion that social engagement in the discussion board also improves students' quiz scores. Hence, our results demonstrate that it is important to use experiment-based methods to extract reliable causal inference in MOOC design and analysis.

### 4.4. Differential Treatment Effects and Subgroup Analysis

So far, we have assumed that the average effect of our encouragement treatment on students' social engagement levels and learning outcomes is homogeneous across individuals. In practice, different students may respond differently to our treatments, and the effects of social engagement levels on learning outcomes could vary among students. To investigate these issues, we use the subgroup analysis, which is widely used in clinical trial literature (Rothwell 2005). We then discuss the implication of differential treatment effects on our IV estimators.

**Reduced-form Treatment Effect**: Subgroup analysis helps us to understand important heterogeneity in treatment effects between different subsets of the population. In particular, we can divide the population based on students' signature-track status, English locale status, and email announcement status. We only provide subgroup analysis on the first measure, i.e., the signature-

| | Dependent variable | | | |
|---|---|---|---|---|
| | Social Engagement | | Learning Outcomes | |
| | **Number of Visits** (1) Poisson | **Number of Posts** (2) Poisson | **Quiz Completion Rate** (3) Probit | **Quiz Score** (4) OLS |
| **Panel A: Non-signature-track students (Week 2- 5)** | | | | |
| Treatment | $0.243^{***}$ | $0.405^{***}$ | $0.236^{***}$ | 0.010 |
| | (0.069) | (0.103) | (0.065) | (0.134) |
| Observations | 1028 | 1028 | 1028 | 341 |
| **Panel B: Signature-track students (Week 2 - 5)** | | | | |
| Treatment | $0.159^{***}$ | $0.728^{***}$ | $-0.143$ | $-0.179$ |
| | (0.059) | (0.117) | (0.088) | (0.038) |
| Observations | 312 | 312 | 312 | 276 |

(1) Standard errors are in parentheses.
(2) . $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
(3) All regressions control for signature track status, email announcement status, and language.
(4) Standard errors are corrected for clustering at the weekly level.

Table 5: Sub-group analysis of reduced-form estimates

track status, since we do not find any statistically significant heterogeneities in treatment effects between students with different characteristics in other two controls.[9]

Table 5 provides the effects of our encouragement treatment on students' social engagement levels and learning outcomes for signature-track and non-signature-track students separately. Panel (a) shows the treatment effect on social engagement levels in the discussion board as well as learning outcomes for *non-signature-track students*. It is evident that the encouragement has significantly positive effects on non-signature-track students' social engagement levels and their quiz completion rates. This result is consistent with what we get from the general population analysis in Table 2.

Panel (b) of Table 5 provides the corresponding treatment effects of the encouragement for signature-track students. Column (1) and (2) show that the encouragement also has a positive and significant effect on signature-track students' social engagement levels. In other words, both non-signature-track and signature-track students' likelihood of visiting the board and posting on the board increases after receiving treatment. Interesting, focusing only on Column (3) and (4) in Panel (b), we find that, despite that signature-track students who received treatment had higher social engagement levels, they were not more likely to finish the quiz nor scored higher in the quiz.

In summary, the subgroup analysis suggests that our encouragement treatment has significant effects of encouraging both signature-track and non-signature-track students to visit the discussion board and post more on it. This is intuitive since, as illustrated in Section A, our formulation of the encouragement does not depend on any information related to signature-track status. Interestingly, only non-signature-track students have better learning outcomes (i.e., higher quiz completion

[9] We report this null result since past literature has emphasized on the danger related to p-value fishing with sub-group analysis (Assmann et al. 2000).

| | Quiz completion rate | |
|---|---|---|
| | **Non-signature-track Students** | **Signature-track students** |
| **Panel A: Average Effect of Social Engagement on Learning Outcomes (Week 3)** | | |
| Number of Visits | $0.0341^{***}$ | 0.0188 |
| | (0.0057) | (0.0117) |
| Observations | 257 | 78 |
| **Panel B: Average Effect of Social Engagement on Learning Outcomes (Week 4)** | | |
| Number of Visits | $0.0231^{***}$ | $-0.0046$ |
| | (0.0048) | (0.0114) |
| Observations | 257 | 78 |
| **Panel C: Average Effect of Social Engagement on Learning Outcomes (Week 5)** | | |
| Number of Visits | $0.0195^{***}$ | $-0.0013$ |
| | (0.0024) | (0.0088) |
| Observations | 257 | 78 |

(1) Standard errors are in parentheses.
(2) . $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
(3) All regressions control for signature track status, email announcement status, and language.

Table 6: Sub-group analysis of the social engagement effects on learning outcomes over signature-track status.

rate) after receiving the treatment. This suggests that non-signature-track students benefit from increasing social engagement levels while signature-track students may not.

**IV Estimator:** The previous heterogeneity in encouragement effects suggests that treatment effects of social engagement levels on learning outcomes may be different between signature-track and non-signature-track students. Therefore, we re-estimate our IV estimators in Equation (5) and (6) on signature-track and non-signature-track students separately based on the second measure of social engagement levels, i.e., $\tilde{v}$.

Table 6 shows our estimated effects of social engagement levels on learning outcomes for non-signature-track and signature-track students respectively. Comparing Column (1) and (3) in Table 6, we can see that social engagement in the discussion board increases students' completion rates only for non-signature-track students. This is consistent with our previous reduced-form estimators, which indicate that the encouragement has positive and significant effects on learning outcomes only for non-signature-track students. This subgroup analysis reveals one important feature of social engagement on discussion board: signature-track students, in our context, do not benefit from higher social engagement levels.

We provide two interpretations of this no-effect result: (1) social engagement in the discussion board can only improve learning outcomes up to a certain level, and therefore only students who originally have low social engagement levels (i.e., non-signature-track students) benefit from increasing social engagement in the discussion board; (2) the fee paid by signature-track students is already a powerful commitment device, and therefore any other form of motivation, such as social engagement, has less influence on them.

| | Non-signature-track Students | Signature-track students |
|---|---|---|
| | **Average Effect of Social Engagement on Quiz Completion Rate (Week 3)** | |
| **Low Social Engagement** | $0.694^{***}$ | 0.048 |
| | (0.021) | (0.044) |
| **High Social Engagement** | -0.019 | 0.016 |
| | (0.081) | (0.007) |

(1) Standard errors are in parentheses.
(2) . $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Sub-group analysis of the social engagement effects over signature-track status and original social engagement levels

To empirically justify these two interpretations, we divide the students into four groups based on (1) their signature-track status and (2) their initial social engagement level as measured by whether their number of visits to the discussion board is in the bottom 50% across all students during the first week. We re-estimate the effects of social engagement on learning outcomes for students in each group separately. Table 7 demonstrate the results. We observe that signature-track students, regardless of their social engagement levels in the first week, do not benefit from social engagement in the discussion board. Moreover, we also observe that only non-signature-track students who have low initial social engagement levels benefit from the increasing social engagement levels after the treatment. These results empirically justify both of our interpretations.

Last, we may further assume that the treatment effect is heterogeneous even within signature-track students and non-signature-track students. In other words, $\gamma_{ij}^1 \neq \gamma_{kj}^1$, where $\gamma_{ij}^1$ represents the effect of student $i$'s board visits on her quiz completion rate in Week $j$. In this case, our original estimators $\gamma_j^1$ and $\gamma_j^2$ no longer estimate the average treatment effect. Let $v_{ij}(0)$ denote the number of visits that student $i$ has before quiz in Week $j$ if she did not receive the treatment. Similarly, $v_{ij}(1)$ represents the number of visits that student $j$ has before quiz in Week $j$ if she received the treatment. As the literature on differential treatment effects has documented (Angrist and Imbens 1995), with simple monotonicity assumption, the estimated effects, i.e., $\gamma_j^1$ and $\gamma_j^2$, represent *the local average treatment effects of treatment compliers*, the effect of social engagement on learning outcomes for compilers, students with $v_{ij}(1) > v_{ij}(0)\ \forall j$. The monotonicity assumption is as follows:

Assumption 3. *Monotonicity assumption: For each individual $i$, $v_{ij}(1) \geq v_{ij}(0)\ \forall j$.*

This assumption states that receiving the encouragement treatment cannot decrease students' probabilities to visit the discussion board. While, for each student $i$ and each week $j$, we cannot observe both $v_{ij}(0)$ and $v_{ij}(1)$ simultaneously, this assumption sounds very plausible in our situation. In particular, our treatment page, as shown in Appendix A, does not contain any content that discourages students from visiting the discussion board.

| | **Receiving Invitations** | | | | |
|---|---|---|---|---|---|
| | **Group 1** (1) | **Group 2** (2) | **Group 3** (3) | **Treated Group** (4) | **Control Group** (5) |
| | | | **Panel A: Demographics Characteristics** | | |
| Signature Track | 0.24 | 0.21 | 0.09 | 0.11 | 0.11 |
| | (0.06) | (0.05) | (0.01) | (0.02) | |
| English Locale | 0.41 | 0.51 | 0.44 | 0.47 | 0.45 |
| | (0.07) | (0.06) | (0.02) | (0.02) | (0.02) |
| Email Announcement | 0.98 | 1,00 | 0.99 | 0.99 | 0.99 |
| | (0.02) | (0.00) | (0.02) | (0.00) | (0.00) |
| | | | **Panel B: Learning Outcomes (Week 2 - 5)** | | |
| Quiz Completion in Week 3 | 0.54 | 0.47 | 0.17 | 0.21 | 0.21 |
| | (0.07) | (0.06) | (0.01) | (0.01) | (0.02) |
| Quiz Score in Week 3 | 2.66 | 2.63 | 2.60 | 2.62 | 2.67 |
| | (0.11) | (0.12) | (0.06) | (0.05) | (0.06) |
| Quiz Completion in Week 4 | 0.46 | 0.36 | 0.13 | 0.16 | 0.17 |
| | (0.07) | (0.06) | (0.01) | (0.01) | (0.02) |
| Quiz Score in Week 4 | 4.16 | 3.73 | 3.65 | 3.74 | 3.93 |
| | (0.23) | (0.17) | (0.10) | (0.09) | (0.10) |
| Quiz Completion in Week 5 | 0.37 | 0.32 | 0.10 | 0.13 | 0.14 |
| | (0.07) | (0.05) | (0.01) | (0.01) | (0.02) |
| Quiz Score in Week 5 | 3.15 | 3.08 | 3.15 | 3.02 | 3.26 |
| | (0.23) | (0.18) | (0.12) | (0.09) | (0.11) |
| Observations | 54 | 75 | 817 | 946 | 531 |

1. Standard errors are in parentheses.
2. The quiz score in each week is the average of maximum quiz score across all students who attempted the quiz.

Table 8: Descriptive statistics of the small-group interaction experiment, by groups.

## 5. Small-group Interaction Experiment: Analysis and Findings

In this section, we analyze the results from our small-group interaction experiment: giving access to one-on-one discussion technology to a random subset of students. For each student $i$, we denote $A_i = 1$, if student $i$ receives the invitation email, and $G_i = 1$, if student $i$ received the invitation and successfully conducted the pairwise discussion. Moreover, we use $W_i = 1$ to indicate that student $i$ wants to conduct the one-ob-one discussion after they have been assigned to a partner. Notice that a student completes the feedback form if either she successfully conducted the discussion or she wanted to be re-matched to another partner. Hence, we assume that all students who have completed the feedback form wanted to conduct the one-on-one discussion. We deliberately did not re-match these students so that we can not only estimate the effect of receiving the invitation email (i.e., the intention-to-treat effect) but also examine the effect of actually conducting the discussion (i.e., the treatment effect).

We divide the students into four groups based on their characteristics, i.e., $(A_i, G_i, W_i)$:

(1) Treated Group 1: students who were assigned to a partner and conducted the one-on-one discussion ($A_i = 1$, $G_i = 1$ and $W_i = 1$).

(2) Treated Group 2: students who were assigned to a partner, wanted to conduct the discussion but did not due to issues of their partners ($A_i = 1$, $G_i = 0$ and $W_i = 1$).

(3) Treated Group 3: students who were assigned to a partner but did not want to conduct the discussion ($A_i = 1$, $G_i = 0$ and $W_i = 0$).

(4) Control: students who were not invited ($A_i = 0$, $G_i = 0$ and $W_i = 0$).

Table 8 displays the summary statistics associated with these groups of students. Column (1), (2), (3), and (4) correspond to treated group 1, treated group 2, treated group 3, and group of all treated students (i.e., students who received the invitation) respectively. Column (5) provides summary statistics of students in the control group.

Panel A of Table 8 represents the demographics characteristics of each group. It is evident that there are no differences between characteristics in Column (4) and (5). This is expected since we randomly select the subset of students to send the invitation email. There are considerable differences between the signature-track ratios in Column (1), (2) and (3). This implies that signature-track students are more likely to conduct the one-on-one discussion after receiving the invitation. Moreover, among 946 students who were invited, only 54 of them conducted the discussion, and 129 of them wanted to conduct the discussion. In other words, the effort cost of social engagement was so high that the majority of students (i.e., more than 85%) who received the invitation did not want to conduct the pairwise discussion.

In Panel B, we present learning outcomes of students in different groups after Week 3 since students were instructed to have the discussion during Week 3. Comparing Column (4) and (5), we can see that students who received the invitation did not have higher quiz completion rates nor higher quiz scores after Week 3. In other words, there is no effect of receiving the invitation on learning outcomes. In contrast, comparing Column (1) and (2), it is evident that, among students who wanted to conduct the one-on-one discussion, those who actually did it had higher quiz completion rates and quiz scores after Week 3.

To formally analyze the effect of receiving the invitation and conducting the pairwise discussion on learning outcomes, we consider a set of reduced-form specifications:

$$q_{ij} = \text{Probit}(\alpha + \theta^1 A_i + X_i + W_j + u_{ij}) \ \forall\ i, \tag{7}$$

$$s_{ij} = \alpha + \theta^2 A_i + X_i + W_j + u_{ij} \ \forall\ i, \tag{8}$$

$$q_{ij} = \text{Probit}(\alpha + \theta^3 G_i + X_i + W_j + u_{ij}) \ \forall\ i \text{ s.t. } W_i = 1, \tag{9}$$

$$s_{ij} = \alpha + \theta^4 G_i + X_i + W_j + u_{ij} \ \forall\ i \text{ s.t. } W_i = 1, \tag{10}$$

where $q_{ij}$ and $s_{ij}$ represent student $i$'s quiz completion rates and scores in Week $j$, $X_i$ is the vector of student-specific controls including her signature track status, email announcement status, and language locale, and $W_j$ is the week level control.

The estimators of interest are $\theta^1$, $\theta^2$, $\theta^3$, and $\theta^4$. $\theta^1$ and $\theta^2$ represent the average treatment effects of **receiving the invitation email** on students' quiz completion rates and quiz scores, while $\theta^3$ and $\theta^4$ are the average effect of **conducting the one-on-one discussion** on students' quiz completion rates and quiz scores. In other words, $\theta^1$ and $\theta^2$ are intention-to-treat effects of the one-on-one discussion treatment, while $\theta^3$ and $\theta^4$ are the actual average treatment effects.

| | **Quiz Completion Rate** | **Quiz Score** |
|---|---|---|
| | (1) Probit | (2) OLS |
| **Panel A: Average effect of receiving the invitation on learning outcomes** | | |
| Treatment | $-0.0001$ | $-0.031$ |
| | (0.018) | (0.019) |
| Signature Track | $1.949^{***}$ | $2.044^{**}$ |
| | (0.026) | (0.211) |
| Observations | 4431 | 4431 |
| **Panel B: Average effect of conducting the discussion on learning outcomes** | | |
| Treatment | $0.248^{**}$ | $0.304^{*}$ |
| | (0.076) | (0.152) |
| Signature Track | $2.202^{***}$ | $2.072^{*}$ |
| | (0.194) | (0.216) |
| Observations | 384 | 384 |

(1) Standard errors are in parentheses.
(2) . $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
(3) All regressions also control for the interaction effect of the treatment and signature track status, email announcement status, and language.

Table 9: Reduced-form estimates for small-group interaction experiment

Table 9 demonstrates the estimation results. Panel A of Table 9 shows that receiving the invitation does not increase students' quiz completion rates nor their quiz scores. This no-effect result is surprising since the one-on-one discussion should provide more collaborative learning than merely visiting the discussion board and we find there is a significantly positive effect of encouraging students to visit the discussion board on learning outcomes. Our interpretation of this "contradiction" between Experiment 1 and 2's results is that the effort of benefiting from an intervention in MOOCs is important: while the cost for students to visit the board after receiving the encouragement is a few clicks, it requires much more efforts for students who received the invitation to actually coordinate and conduct the one-on-one discussion with another student. Therefore, the benefit of receiving the invitation is tiny since most students were not able to conduct the discussion. This interpretation is empirically justified by our data: while more than 80% of students who received the encourage in Experiment 1 have visited the board at least once after the encouragement, only 7% of students have conducted the pairwise discussion after receiving the invitation.

From Panel B of Table 9, it is evident that there is a large effect of actually conducting the one-on-one discussion on students' completion rates, and more importantly, student quiz scores. This shows that other forms of social interactions among students, such as one-on-one discussion, could also improve learning outcomes. It is interesting that conducting the discussion has a significant positive effect on students' quiz scores while higher level of social engagement in the discussion board does not. One possible explanation is that students can effectively exchange information and collaboratively study quiz questions during the one-on-one discussion while they could not achieve this through the discussion board.

## 6. Discussion

There are several important insights that we derive from our experiments and analysis:

**Encouragement Design:** Experiment 1 demonstrates that a simple encouragement treatment increased treated students' visits to, and their number of posts on, the discussion board by 26.5% and 96.8%, in subsequent weeks respectively. This indicates that even simple encouragements at the beginning of a course can have a profound effect on how students interact with the course. Moreover, our simple encouragement also increased treated students' quiz completion rates by 10% and their total grades by 6 points (compared to an average of 39 for the control group). Therefore, it is important for MOOC researchers and practitioners to study and adopt similar encouragement mechanisms at the beginning of their courses to improve learning outcomes.

**The discussion board and quiz completion rates**: Our large-group interaction experiment provides strong evidence that social engagement in online discussion boards benefits the majority of students in MOOCs. In particular, one additional visit to the discussion board (i.e., visits to read or to post) causally increases a student's quiz completion probability by 3.5% in the subsequent week. This result validates the importance of better designed discussion boards that improve students' learning outcomes, e.g., by providing badging systems (Anderson et al. 2014) or implementing a post recommendation system (Breslow et al. 2013).

Moreover, our results also demonstrate that the effects of social engagement levels on learning outcomes is steadily decreasing over time. In other words, one visit to the discussion board at the beginning of a course is worth much more than one at the end. One possible explanation is the self-selection mechanism: students who still remain in the course are more motivated on average to complete the quiz. Therefore, the marginal effect of one more visit on those students' completion rates are smaller than that on all students. This implies that MOOC researchers and practitioners should focus primarily on designs that improve the discussion board participation rates in the early stage of a course.

**Social engagement and signature track:** The sub-group analysis of our large-group interaction experiment suggests that, in our course, signature-track students *do not* benefit from social engagement in the discussion board. We provide two empirically justified interpretations of this no-effect result: (1) social engagement in the discussion board can only improve learning outcomes up to a certain level, and therefore only students who originally have low social engagement levels (i.e., non-signature-track students) benefit from increasing social engagement in the discussion board; (2) the service fee is already a powerful commitment device for signature-track students since they have to pass a certain grade hurdle to receive the certificates. Therefore any other form of motivation, such as social engagement, has less influence on learning outcomes. This no-effect result implies that for-profit MOOC platforms may not benefit much from better designed discussion boards since their paying customers, who are already motivated by the fee and usually already have high social engagement levels, are not more likely to finish the course if they are more socially engaged in the online discussion.

**The discussion board and quiz scores**: Moreover, our large-group interaction experiment's results also show that students' social engagement levels have no effects on their quiz scores. This is not surprising in our course since students, restricted by the honor code, are not allowed to discuss the quiz before it is closed. Moreover, we show that the effect of social engagement levels on quiz scores estimated by simple OLS specifications is likely to be positive and significant. This may lead researchers to incorrectly conclude that social engagement in the discussion board increases not only students' probabilities to complete quizzes but also their quiz scores. This result has two implications for MOOC designers: (1) making decisions solely based on the correlation between social engagement levels and learning outcomes may be misleading; (2) it is important to re-design the discussion board so that students can actually learn more form the board and achieve higher scores in quizzes.

**One-on-one discussion and learning outcomes:** The results of our small-group interaction experiment provide evidence that students who engage in one-on-one discussion improve their quiz completion rates and quiz scores. We caution, however, that the fraction of students who actually held pairwise discussions after receiving the invitation was very small (i.e., about 7%). Therefore, there seems to be no effect of receiving the invitation of this pairwise discussion on learning outcomes: the cost of one-on-one social engagement to students in MOOCs is so high that the majority of them who were invited did not follow up with their assigned partners. This indicates that, when designing mechanisms to create or improve small group social engagement in MOOCs, educators and researchers should focus primarily on reducing the transaction cost of those mechanisms to students.

Moreover, our results suggest that conducting one-on-one discussion not only improves students' quiz completion rates but also increases their final quiz scores. This is in stark contrast with Experiment 1 where we found that social engagement in the discussion board only increases students' completion rates. One interpretation is that one-on-one discussion is a more direct communication mechanism that is more conducive to collaborative learning. This result encourages MOOC practitioners and researchers to think "out of the box" and look for other direct communication channels for students to socially engage with each other to improve learning outcomes.

## 7. Conclusion

This paper analyzes two randomized experiments to study how social interaction among MOOC students impacts learning outcomes. We highlight the two main results of our analysis. First, we show that large-group social engagement in MOOC discussion boards improves the service quality for the majority of students (i.e., non-signature-track students). In contrast, we also show that in our course signature-track students do *not* benefit from social engagement in the discussion board, and provide two empirically justified explanations. Second, we demonstrate that small-group social engagement in one-on-one discussions also improves service quality by increasing quiz completion rates as well as quiz grades. However, we caution that the cost incurred by students of conducting such small-group social engagement may be substantial given that only a small subset (i.e., 7% in our case) of the invited students actually participated.

Given the demonstrated improvements of service quality from social interactions, one natural next step is to study how our insights can be used to design better operating systems for MOOCs. For example, what is the optimal timing and frequency of encouraging students to socially interact with each other? Given that social engagement in our discussion board does not improve learning outcomes of signature-track students, it is desirable to design other forms of social interactions that benefit all students. Last, it is interesting to extend our findings to other service sectors and study the effect of participants' social interactions on the service quality in other settings.

## References

Zeynep Aksin, Baris Ata, Seyed Morteza Emadi, and Che-Lin Su. Structural estimation of callers' delay sensitivity in call centers. *Management Science*, 59(12):2727–2746, 2013.

Gad Allon, Achal Bassamboo, and Itai Gurvich. We will be right with you: Managing customer expectations with vague promises and cheap talk. *Operations research*, 59(6):1382–1394, 2011.

Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web*, pages 687–698. International World Wide Web Conferences Steering Committee, 2014.

Joshua Angrist and Guido Imbens. Identification and estimation of local average treatment effects, 1995.

Susan F Assmann, Stuart J Pocock, Laura E Enos, and Linda E Kasten. Subgroup analysis and other (mis) uses of baseline data in clinical trials. *The Lancet*, 355(9209):1064–1069, 2000.

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

Lori Breslow, David E Pritchard, Jennifer DeBoer, Glenda S Stump, Andrew D Ho, and Daniel T Seaton. Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, 8(1):13–25, 2013.

Ryan W Buell and Michael I Norton. The labor illusion: How operational transparency increases perceived value. *Management Science*, 57(9):1564–1579, 2011.

Ozan Candogan, Kostas Bimpikis, and Asuman Ozdaglar. Optimal pricing in networks with externalities. *Operations Research*, 60(4):883–905, 2012.

Gayle Christensen, Andrew Steinmetz, Brandon Alcorn, Amy Bennett, Deirdre Woods, and Ezekiel J Emanuel. The MOOC phenomenon: who takes massive open online courses and why? *Available at SSRN 2350964*, 2013.

Coursera. Community — coursera. `https://www.coursera.org/about/community`, 2015. (Visited on 06/08/2015).

Esther Duflo and Emmanuel Saez. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics*, 118(3):815–842, 2003.

Bing Jing. Social learning and dynamic pricing of durable goods. *Marketing Science*, 30(5):851–865, 2011.

René F Kizilcec, 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*, pages 170–179. ACM, 2013.

Anne Lamb, Jascha Smilack, Andrew Ho, and Justin Reich. Addressing common analytic challenges to randomized experiments in moocs: Attrition and zero-inflation. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pages 21–30. ACM, 2015.

Pinhas Naor. The regulation of queue size by levying tolls. *Econometrica*, pages 15–24, 1969.

Yiangos Papanastasiou and Nicos Savva. Dynamic pricing in the presence of social learning and strategic consumers. *History*, 2014.

Peter M Rothwell. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *The Lancet*, 365(9454):176–186, 2005.

Christian Terwiesch and Karl T Ulrich. Will video kill the classroom star? The threat and opportunity of massively open on-line courses for full-time MBA programs. *The Threat and Opportunity of Massively Open On-Line Courses for Full-Time MBA Programs (July 16, 2014)*, 2014.

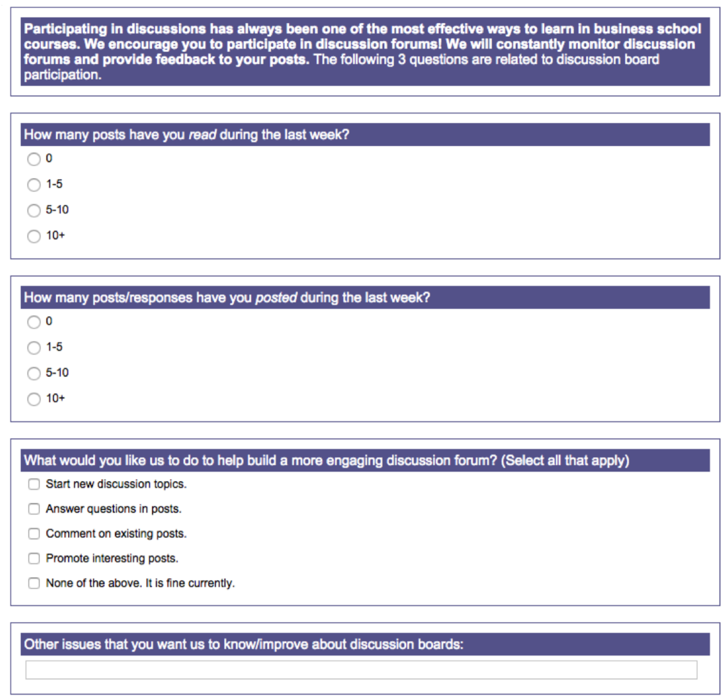
Senthil K Veeraraghavan and Laurens G Debo. Herding in queues with waiting costs: Rationality and regret. *Manufacturing & Service Operations Management*, 13(3):329–346, 2011.

Ian R White and John B Carlin. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in medicine*, 29(28):2920–2931, 2010.

Qiuping Yu, Gad Allon, and Achal Bassamboo. How do delay announcements shape customer behavior? An empirical study. *Management Science, under review*, 2013.

# Appendix

## A. Experiments: Text and Questions


**Participating in discussions has always been one of the most effective ways to learn in business school courses. We encourage you to participate in discussion forums! We will constantly monitor discussion forums and provide feedback to your posts.** The following 3 questions are related to discussion board participation.

How many posts have you *read* during the last week?

- 0
- 1-5
- 5-10
- 10+

How many posts/responses have you *posted* during the last week?

- 0
- 1-5
- 5-10
- 10+

What would you like us to do to help build a more engaging discussion forum? (Select all that apply)

- Start new discussion topics.
- Answer questions in posts.
- Comment on existing posts.
- Promote interesting posts.
- None of the above. It is fine currently.

Other issues that you want us to know/improve about discussion boards:

Figure 4: Second page of the survey: encouragement text and questions

Figure 4 shows the second of page of the survey, which is our treatment in Experiment 1 that consists of a short paragraph with encouragement text and four questions related to the discussion board in Week 1.

Experiment 2 consisted of an email, shown in Figure 5, that we sent to invite participating students to one-on-one group discussion is as follows. We have masked the course name as “MOOC”, guiding questions as “guiding question”, the survey url as “URL” and our names as “the instructor” for reviewing purposes.

*Dear MOOC Students,*

*Thank you both for participating in group discussions. We have matched you with another student (cc-ed) based on your time and communication preferences. Both of you have indicated that you would like to use **Facebook Messenger** to discuss with each other. Please conduct the discussion by the end of third week (April 18th). Feel free to discuss any topic related to the course, but here are some guiding questions:*

1. *Guiding Question 1.*
2. *Guiding Question 2.*
3. *Guiding Question 3.*

*This is our first time asking students to conduct their own group discussions. We want this experience to be as useful to you as possible, so please fill out the following survey after deciding the time and communication method for your discussion. We will use this information to re-match unsuccessful matches and improve future matches: URL*

*Thank you, and good luck with your discussion. The instructor.*

Figure 5: One-on-one discussion invitation email