

Does Social Interaction Improve Learning Outcomes? Field Evidence from Massive Open Online Education

Dennis J. Zhang

Olin Business School, Washington University in St. Louis, denniszhang@wustl.edu

Gad Allon

Wharton School, University of Pennsylvania, gadallon@wharton.upenn.edu

Jan A. Van Mieghem

Kellogg School of Management, Northwestern University, vanmieghem@kellogg.northwestern.edu

May 17, 2016

This paper studies how service providers can design 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 three randomized experiments in a MOOC with more than 30,317 students from 183 countries. Two experiments study large-group interaction by encouraging a random subset of students to visit the course discussion board. The majority of students treated in these experiments had higher social engagement, higher quiz completion rates, and higher course grades. Using these treatments as instrumental variables, we estimate that one additional board visit causally increases the probability that a student finishes the quiz in the subsequent week by up to 4.3%. The third experiment studies small-group interaction by encouraging a random subset of students to conduct one-on-one synchronous discussions. Students who followed through and actually conducted pairwise discussions increased their quiz completion rates and quiz scores by 10% in the subsequent week. Combining results from these three experiments, we provide recommendations for designing social interaction mechanisms to improve service quality.

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

History: Submitted on September 21, 2015; Revised on May 17, 2016; Accepted on July 22, 2016.

1. Introduction

Education is an important service sector of the economy that many readers of this journal participate in yet it 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 expensive: college student loan debt in the U.S., reaching \$1.2 trillion, is the second largest debt after mortgage debt.¹

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 marginal distribution cost. A typical MOOC involves a series of video lectures, weekly assignments, and a discussion board moderated by instructors. 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, EdX and Udacity) have collaborated with top academic institutions around the world to deliver free online courses to learners anywhere on the planet.

The hype over MOOCs, however, 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) found 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 grades and completion rates.

In this paper, we take an operational perspective and ask a fundamental question: does large-group social interaction in the discussion board, a key component of the Socratic method, causally improve learning outcomes in MOOCs?⁴ If so, do benefits accrue equally to all students or disproportionately to some groups? Moreover, can synchronous social interaction (not supported in existing MOOC platforms) benefit students and improve service quality? Lastly, how do these findings suggest better ways to design social interaction in MOOCs as well as in general services?

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 participants plays in determining service quality. Thus, while this paper focuses on specific

¹ <http://www.newyorkfed.org/newsevents/news/research/2015/rp150217.html>

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/>

⁴ Our main question and the corresponding results apply to MOOC platforms where students communicate asynchronously on the discussion board and are evaluated based on quiz/assignment grades. All three major MOOC platforms, Coursera, EdX and Udacity, satisfy these two conditions.

and important service sector, education, the question is much broader: how can service providers construct social interaction among their participants and how can one validate, using data and randomized experimentation, that social interaction indeed causes service quality to improve?

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 interaction has a *causal* effect (beyond correlation) on learning outcomes as well as on service quality 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 three randomized experiments during two offerings of our own MOOC, which is a five-week-long introductory Operations Strategy course that was offered in April and September 2015 on the Coursera platform. The first offering had 24,005 registered students of which 13,726 visited actual course materials at least once. This offering contains two experiments: *The first experiment focuses on large-group interaction, while the second one targets small-group interaction.* We conducted another large-group experiment in the second offering of our class to test the robustness of our results.

In the first offering, 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 survey respondents 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 encouraged in 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 (IV), we estimate that *one additional board visit causally increases the probability that a student finishes the quiz in the subsequent week by 0.5% to 1.0%.*

While this large-group interaction experiment focuses on MOOCs, we believe that the main insight – service providers can construct social interaction to improve service quality – can be

applied to other service settings. In order to test the robustness of this main result and generalize it to a broader population, we conducted another large-group interaction experiment with a different encouragement in our second offering of the MOOC. In this new experiment, we estimate that, for an average student, one more board visit increases the quiz completion rate in the subsequent week by 1.2% to 4.3%. In other words, we reproduced our main result with a different experimental set-up and a completely different sample population, which demonstrates the robustness of our main results.

We further analyze the large-group interaction experiment in the first offering and provide several insights specific to the MOOC industry. First, we find that the impact of social interaction on quiz completion rates decreases over time—1.0%, 0.6%, 0.5% for Week 3, 4, and 5. Second, using the same IV, we show that social interaction in the discussion board has no effect on student quiz scores. This result suggests that social facilitation (Zajonc et al. 1965 and Falk and Ichino 2006), a well documented mechanism for students impacting each other in the physical classroom, could be the main mechanism in MOOCs. Third, major MOOC platforms (i.e., Coursera, EdX and Udacity) allow students to register for a course for free or at a cost. On Coursera, the latter is called signature-track and those students receive a certificate after finishing the course with a satisfactory grade. Our experiment suggests that, in our course, signature-track students *do not* benefit from social interaction in the discussion board.

Moreover, we conducted a *small-group interaction experiment* in the first MOOC offering to test the effect of other forms of social engagement on learning outcomes. In particular, at the beginning of the 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 email invitations to pair them up.

The results of this experiment provide empirical evidence that students who engaged in one-on-one discussion improved their quiz completion rates by 7% to 10% as well as quiz scores by 2% to 10% 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 education service and MOOCs from various fields and develops our hypotheses; 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 in the first offering of the class, respectively; we present the analysis of our large-group experiment in the second offering of the class as a robustness test in Section 6; we discuss implications of our analysis for better designing the MOOC channel and conclude the paper in Section 7.

2. Literature Review and Hypothesis Development

Social interaction in educational service has recently received a great deal of attention in the education and economics literature (Epple and Romano 1998, Sacerdote 2001, Zimmerman 2003, Calvó-Armengol et al. 2009 and Lavy and Schlosser 2011). The fundamental question in this literature is how social interaction among students in the physical classroom can impact their test scores and social behaviors. All past papers focus on the physical classroom in elementary (Hoxby 2000), secondary (Burke and Sass 2013) and post-secondary levels (Sacerdote 2001).

While this is a rich literature, the direction and the size of such effect is still not clear: some researchers show that social interaction among peers can positively impact their test scores (Sacerdote 2001), while others find that there is no impact (Lyle 2007), or the impact is negative (Imberman et al. 2012). The main reason for these seemingly contradictory results is that social engagement can impact students' behaviors through many mechanisms and different settings may have different dominant mechanisms. For instance, one well-documented mechanism is *social learning*: students can learn directly from their high-ability peers so that the impact of social interaction on learning outcomes is positive (Sacerdote 2001). Another important mechanism for positive impact is *social facilitation*: students are more likely to exert effort and participate in tests if they observe other students doing the same task (Falk and Ichino 2006). In contrast, *invidious comparison* is a mechanism that provides a negative effect of social interaction on learning outcomes: under this mechanism, students are harmed by the presence of better students in the same classroom (Hoxby and Weingarth 2005).

In this paper, instead of studying the educational service in the physical classroom, we focus on Massive Open Online Courses, which is a virtual educational service. MOOCs have drawn considerable attention from various literature, including Computer Science (Anderson et al. 2014), Education (Breslow et al. 2013), and Operations Management (Terwiesch and Ulrich 2014). Since it is unclear which mechanism is dominant in the MOOC service, the fundamental question that we would like to answer first is: does social interaction in the discussion board *causally* improve learning outcomes in MOOCs? In particular, we hypothesize that:

HYPOTHESIS 1. *Social interaction among students on the discussion board increases their test participation rates and scores.*

The literature also demonstrates considerable heterogeneity in the effects of social interaction on learning outcomes (Imberman et al. 2012 and Fruehwirth 2014). Imberman et al. (2012) use the Hurricane Katrina as a natural experiment to show that high-achieving students benefit from interacting with other high-achieving students while they are hurt by interacting with low-achieving students. Fruehwirth (2014) finds that the magnitude of the effects experienced by student i differ both by student i 's achievement level and peers' achievement levels. We expect similar heterogeneity may exist in the MOOC service. To better design social interaction in MOOCs, we need to understand which group of students benefit the most from such interaction. In particular, since the computer science literature demonstrates that signature-track students have much higher achievement levels than non-signature-track students on Coursera (Anderson et al. 2014), we hypothesize that the estimated effects of social interaction on learning outcomes may be heterogeneous across signature-track and non-signature-track students:

HYPOTHESIS 2. The effect of social interaction on the discussion board over learning outcomes is heterogeneous across signature-track and non-signature-track students.

One fundamental difference between social interaction in MOOCs and physical classrooms is the synchronization of communication. While face-to-face discussion is normally synchronous, communication on the discussion board is often asynchronous. In order to understand whether synchronization plays a vital role in designing social interaction, we construct and test the third hypothesis:

HYPOTHESIS 3. The effect of face-to-face synchronous social interaction on learning outcomes is different from the effect of asynchronous social interaction on the discussion board.

Since our paper focuses on preventing students' drop-outs by constructing social interaction in a service system, we are also 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 the 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 other through social interaction. We thus focus on how learning outcomes are affected by their social interaction.

Finally, our paper is related to the emerging literature in operations management which studies various operations problems in the presence of social interaction among agents. Candogan et al. (2012) study a revenue management problem when customer 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 interaction, 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 receive verified certificates signed by both Coursera and the institution that designs the course. Signature-track students typically are more committed to finishing the course and they tend to have higher ability and achievement levels (Koller et al. 2013). Students who pass the completion requirement but do not pay often receive statements of accomplishment that are not signed nor verified by either Coursera or partner institutions. Most courses on Coursera charge 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 courses and a final project in one area. Specializations normally include 4 to 9 courses and cost between \$500 to \$1000. According to one instructor,⁵ 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 on operations strategy. Each week, there are four video lectures and a quiz. Video lectures are about 5 to 10 minutes, making the lecture time in each week between 20 to 40 minutes. 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 24,005 registered students from 183 different countries among which 480 enrolled in the signature-track; 13,726 registered students actually visited the actual course website at least once; 9,575 watched at least one lecture; 4,221 submitted at least one exercise; and 3,598 browsed the discussion board at least once. Figure 1 shows the fraction of the 9,575 learners who viewed each video over time. Out of the 433 students who completed the course, 233 were signature-track students.

⁵ <http://simplystatistics.org/2015/02/05/johns-hopkins-data-science-specialization-top-performers/>

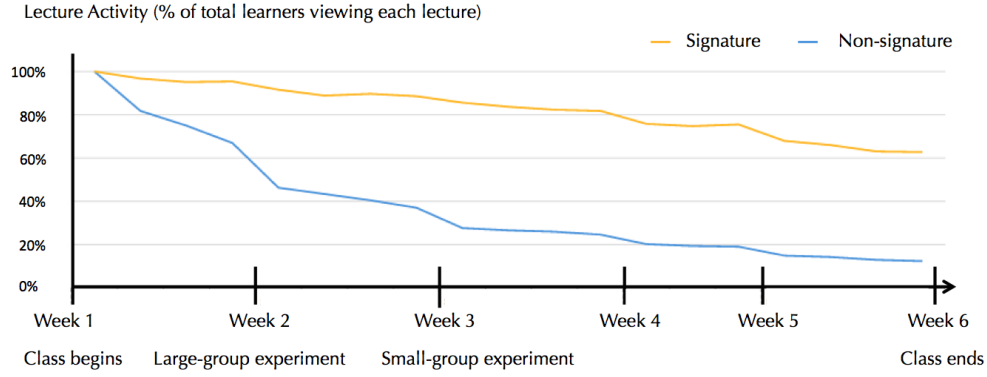


Figure 1: Fraction of students that viewed each video over time. (100% = 9,575 students.)

3.2. Large-group Interaction Experiment

In order to identify the causal effect of social interaction in the discussion board on learning outcomes, we used an “encouragement design” that is widely used in the Economics literature (Duflo et al. 2007). Our large-group interaction experiment encourages a random subset of students to visit the discussion board through randomly assigned survey questions and notifications. The encouragement 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’s 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 an embedded HTTP link instead of directly displaying the URL. This 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: encouragement 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 the most out of the course. Second, there are two questions that ask students how many posts and visits they made 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 at the top. The encouragement text and survey questions served as a reminder for students and set norms for desired levels of social engagement in the discussion board. The second page of the encouragement version is included in Appendix A.

There were about 5,118 students who interacted with the course at least once after the first week. Among those students, 335 of them, about 7% of the 5,118, completed the survey. 175 students received the treatment version, while 160 students were directed to the control version. In the rest of paper, we will refer to those 335 students as participants in the large-group interaction experiment.

Last but not least, let us briefly discuss the analysis of this experiment and some alternative designs. Hypothesis 1 focuses on the **average treatment effect (ATE)** of social interaction on learning outcomes. However, the direct comparison of students receiving the encouragement with those in the control group only provides us **the intent-to-treat (ITT) effect**: the effect of receiving the encouragement on learning outcomes. This is because we face a two-sided non-compliance problem: students who received the encouragement are not guaranteed to go to the discussion board while students who are in the control group may voluntarily visit the board. As shown in Section 4.3, we use the random assignment as an instrument to estimate the ATE.

One may think that we should design a randomized encouragement that guarantees near-perfect compliance so that we do not need the IV strategy to estimate the ATE. One possible design is to close the discussion board for students in the control group. Despite the technical difficulty of implementing these interventions, we think that such “forced design” may introduce additional biases to the analysis. For instance, a student in the control group may infer that the class has low quality because of the lack of discussion boards and in turn is less likely to exert effort in the class. Therefore, to avoid introducing additional biases, we adopt this encouragement design and use the IV analysis to correct for partial compliance (Duflo et al. 2007).

3.3. Small-group Interaction Experiment

At the beginning of the first offering, we conducted a demographic survey and asked 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 students 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 time zones of each other; (2) they have at least one common preferred

communication channel; and (3) they have at least one common preferred day of the 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 to this one-on-one discussion.

In 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) *that 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 analyze this experiment and discuss the detailed classification of responses.

4. Large-group Interaction Experiment: Analysis and Findings

We first discuss our measures of social engagement levels and learning outcomes. We then estimate the intent-to-treat effect of our encouragement design. In particular, we provide empirical evidence that the encouragement in our large-group interaction experiment has significant effects on social engagement levels as well as learning outcomes. Using the encouragement as an IV, we test our Hypothesis 1 and provide our main result: we show that social interaction on the discussion board has positive effects on learning outcomes. Lastly, we extend our empirical model to include heterogeneous treatment effects and test our Hypothesis 2. We demonstrate that only non-signature-track students are affected by social interaction.

4.1. Measurements and Summary Statistics

In this section, we discuss the direct effects of our encouragement on social engagement levels and learning outcomes. The Coursera discussion board 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 user can take four actions on the discussion board: (1) read, (2) post, (3) comment, and (4) vote. First, a user can *read* each thread. Second, a user can *post* a new post in a thread as shown in Zone 1 of Figure 2. Moreover, a user 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

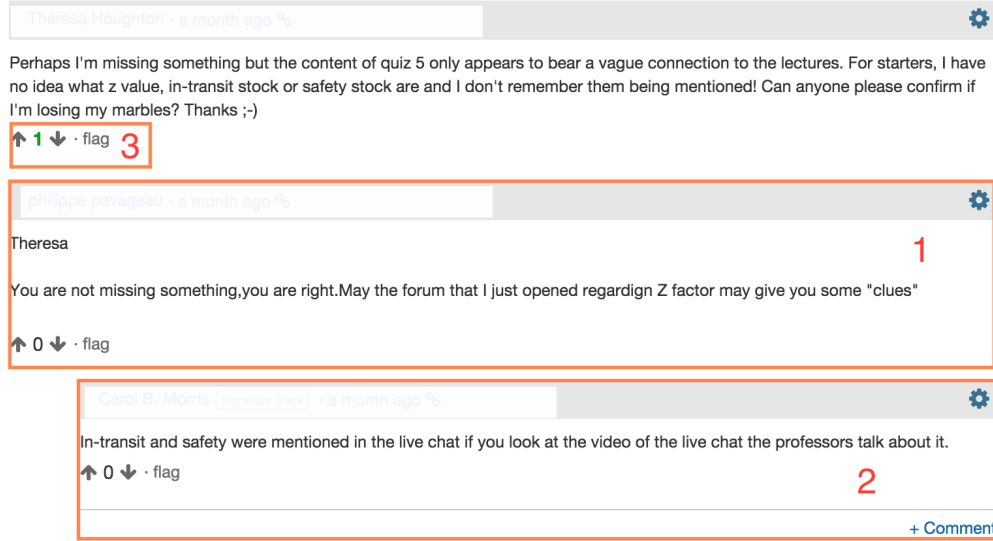


Figure 2: Actions that students can take in the discussion board

a specific existing post in a thread. However, we find that students often use them interchangeably. Lastly, a user can up-vote or down-vote other users' posts by clicking buttons as shown in Zone 3.

Social Interaction: Given the structure of the board and the actions of students, we define social interaction of a student in each time period j with two measures: First, we use v_{ij} to denote the total number of pages on the discussion board visited by student i during the time interval j . In other words, v_{ij} not only measures the frequency that a student visits the board but also how active the student is in each visit (i.e., how many threads and posts 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 number of posts and comments since students often use posts and comments interchangeably. These two definitions are widely used in past literature studying social interaction 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 up to three times in a week, and their maximum scores among all attempts for one quiz are used as their final scores for that quiz. We divide 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 final quiz score at Week j .⁶

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 in the control group. Column (3) shows the statistics

⁶ Our results do not qualitatively change if we use the average quiz scores (across all trials) or the number of trials to measure learning outcomes.

	Experiment Participants			
	Treated Group	Untreated Group	All Participants	Students in Week 2
	(1)	(2)	(3)	(4)
Panel A: Demographics Characteristics				
Signature Track	0.24 (0.03)	0.23 (0.03)	0.23 (0.02)	0.07 (0.00)
English Locale	0.53 (0.04)	0.53 (0.04)	0.53 (0.03)	0.59 (0.01)
Email Announcement	0.99 (0.01)	0.99 0.01	0.99 (0.00)	0.99 (0.00)
Panel B: Social Engagement (Week 2 - 5)				
Number of Visits in Week 2	17.91 (2.89)	13.58 (1.82)	15.84 (1.78)	4.57 (0.21)
Number of Visits in Week 3	17.41 (3.25)	13.47 (2.30)	15.53 (2.02)	4.33 (0.24)
Number of Visits in Week 4	16.85 (5.44)	12.66 (2.40)	14.84 (3.06)	3.97 (0.31)
Number of Visits in Week 5	16.94 (4.20)	14.19 (2.91)	15.63 (2.60)	4.39 (0.3)
Number of Posts in Week 2	1.06 (0.21)	0.66 (0.10)	0.87 (0.12)	0.25 (0.02)
Number of Posts in Week 3	1.41 (0.47)	0.55 (0.11)	1.00 (0.25)	0.25 (0.02)
Number of Posts in Week 4	1.37 (0.69)	0.62 (0.14)	1.01 (0.37)	0.23 (0.03)
Number of Posts in Week 5	1.10 (0.27)	0.68 (0.13)	0.90 (2.16)	0.30 (0.02)
Panel C: Learning Outcomes (Week 2 - 5)				
Quiz Completion in Week 2	0.70 (0.03)	0.56 (0.04)	0.63 (0.03)	0.35 (0.01)
Quiz Score in Week 2	3.63 (0.14)	3.74 (0.15)	3.68 (0.10)	3.63 (0.02)
Quiz Completion in Week 3	0.53 (0.04)	0.44 (0.04)	0.49 (0.03)	0.22 (0.01)
Quiz Score in Week 3	2.6 (0.11)	2.68 (0.11)	2.63 (0.08)	2.61 (0.02)
Quiz Completion in Week 4	0.40 (0.04)	0.36 (0.04)	0.38 (0.03)	0.16 (0.01)
Quiz Score in Week 4	3.96 (0.16)	3.83 (0.16)	3.90 (0.11)	3.79 (0.04)
Quiz Completion in Week 5	0.35 (0.04)	0.30 (0.04)	0.34 (0.03)	0.13 (0.00)
Quiz Score in Week 5	3.16 (0.12)	3.15 (0.13)	3.16 (0.09)	3.11 (0.04)
Total Score	44.46 (2.59)	39.27 (2.90)	41.98 (1.94)	18.87 (0.38)
Course Completion Rate	0.29 (0.04)	0.27 (0.03)	0.28 (0.02)	0.10 (0.00)
Observations	175	160	335	5118

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.

for the entire student population who has participated in the experiment, while Column (4) shows the statistics for all 5,118 students who were still active at the beginning of Week 2, when the survey was sent out.

Panel A of Table 1 represents the demographic characteristics of each group. Student demographics consist 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. Lastly, the email announcement status equals 1 if the student agrees to receive email announcements from Coursera courses. Because the encouragement 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 Column (1), (2) and (3). However, we do observe that students who participated in the experiment (i.e., Column (3)) are slightly different from other students in the class (i.e., Column (4)). In particular, there are 23% signature-track students among all participants while there are only 7% signature-track students among the entire student population. This may limit our results in this section to those students who want to provide feedback. In Section 6, we analyze another large-group experiment in the second offering of our class that does not rely on students’ willingness to give feedback to demonstrate that our results can be generalized to a broader population.

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. Furthermore, the difference between the social engagement levels of students in treatment and control groups does not disappear until Week 5. It really surprised us that the simple nudge in our encouragement design had such profound effect on social engagement levels on the discussion board in the long run.

Panel C presents the effect of our encouragement 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 completion rates 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.

Finally, Figure 3 visualizes the effect of our encouragement on social engagement levels. Panel (a) 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

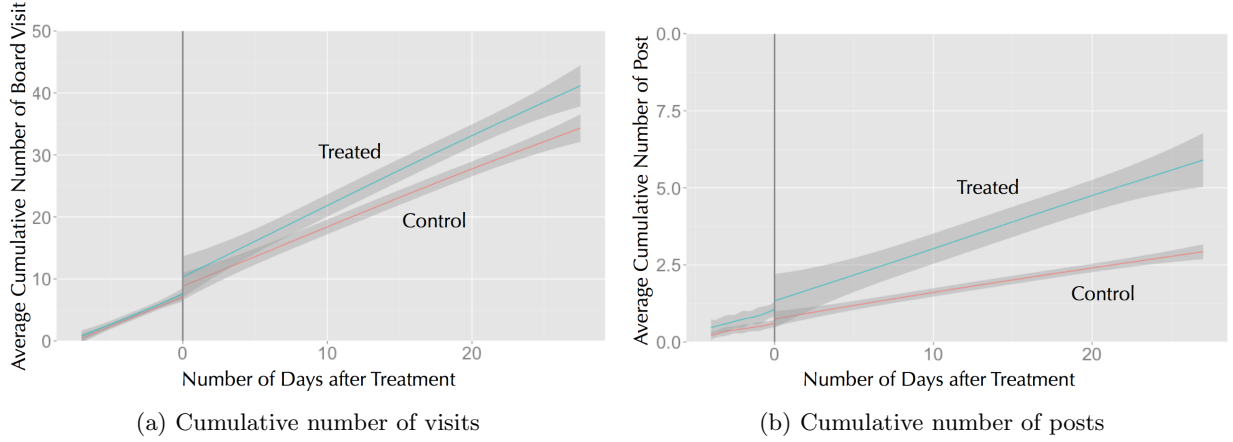


Figure 3: Cumulative numbers of visits and posts before and after the treatment

Panel (a) shows the average cumulative number of visits for students in treatment (green line) and control (red line) groups before and after the encouragement. 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 shaded areas represent the standard errors 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.

evident that these two groups have similar cumulative number of visits before the assignment while encouraged students have significantly higher cumulative number of visits after the encouragement. Moreover, the difference between average cumulative numbers of visits of students in those two groups increases over time. This shows that the encouragement has a long-lasting effect on social engagement levels. Similarly, Panel (b) 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. Intent-to-treat (ITT) Effect

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

$$v_{ij} = \text{Poisson}(\alpha_1 + \beta_1 D_i + W_j + X_i + \epsilon_{ij}), \quad (1)$$

where v_{ij} is the number of visits for student i at Week j , $W_j \in \{2, 3, 4, 5\}$ represents the week-level control, and X_i is a demographics control student i (i.e., signature-track status (S_i), English locale and email announcement status).⁷

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

$$p_{ij} = \text{Poisson}(\alpha_2 + \beta_2 D_i + W_j + X_i + \epsilon_{ij}). \quad (2)$$

The estimates of interest are β_1 and β_2 : $\beta_1 > 0$ shows that the treatment has a positive effect on student probability to visit the board, while $\beta_2 > 0$ demonstrates that the encouragement treatment invites students to post more in the discussion board. The regression also controls for other observed characteristics (i.e., English locale and email announcement).⁸ Since this analysis is based on a panel of students over 4 weeks, the standard errors of the same student could be correlated. We address this issues with several more complex specifications in Appendix B.

The estimates for β_1 and β_2 are reported in Column (1) and (2) in Panel A of Table 2. It shows that receiving the encouragement treatment significantly increases numbers of visits and posts in the discussion board after the encouragement. Specifically, receiving the encouragement increases a student's 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 encouragement 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 the control group is only 2.5. These results show that our randomly assigned encouragement is correlated with social engagement levels. Hence, these results set the foundation to use our randomly assigned encouragement treatment as an IV to estimate the average treatment effect of social interaction on learning outcomes.

Column (3) of Table 2 shows that the encouragement increases a student's number of posts even controlling for her number of visits to the discussion boards. This demonstrates that there are two channels that our encouragement may improve students' activities in the discussion boards: (1) our encouragement can increase a student's number of visits to the board and in turn her number of posts; (2) our encouragement can increase a student's probability to post if she is already on the discussion board. The marginal effect of our encouragement estimated in Column (2) of Table 2 is

⁷ Our results do not qualitatively change if we include the interaction term between D_i and S_i in the model.

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	Number of Posts		Quiz Completion Rate	Quiz Score
	(1) Poisson	(2) Poisson	(3) Poisson	(4) Probit	(5) OLS
Panel A: Average Effect of Social Engagement and Learning Outcome (Week 2 - 5)					
Encouragement Treatment	0.192*** (0.0141)	0.613*** (0.0604)	0.298*** (0.0639)	0.178* (0.0765)	−0.0270 (0.0703)
Signature Track	1.611*** (0.0143)	1.874*** (0.0599)	1.342*** (0.0652)	1.740*** (0.107)	−0.0077 (0.0709)
Marginal Effect	2.944***	0.550***	0.271***	0.0537*	−0.0270
Observations	1340	1340	1340	1340	617
Panel B: Average Effect of Social Engagement and Learning Outcome (Week 1)					
Encouragement Treatment	−0.0133 (0.206)	0.321 (0.202)	0.257 (0.187)	0.275 (0.183)	0.059 (0.0417)
Signature Track	0.985*** (0.209)	0.908*** (0.304)	0.394 (0.203)	(omitted) (omitted)	0.0158 (0.0461)
Observations	335	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 signature-track status, email announcement status, and English locale.

Table 2: Intent-to-treat effects of our encouragement on social engagement levels and learning outcomes in the large-group interaction experiment.

almost twice of that in Column (3), 0.550 vs. 0.271. This demonstrates that the impact from our encouragement on board activities through these two channels is almost identical.

Last but not least, the number of visits and posts may be over-dispersed (i.e., the standard deviations are larger than the mean) and has 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 statistically different results when using zero-inflated Poisson models (i.e., models to account for zero-inflation) or negative binomial models (i.e., models to account for over dispersion) to estimate Equation (1) and (2). Furthermore, student performance may be correlated over time, which implies that the standard errors of each student's observations may be correlated over weeks. In Appendix B, we present several robustness tests to show that our results do not qualitatively change even if we allow standard errors of each student's observations to be correlated over time.

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 quiz completion rates and grades as follows:

$$q_{ij} = \text{Probit}(\alpha_3 + \beta_3 D_i + W_j + X_i + \epsilon_{ij}), \quad (3)$$

and

$$s_{ij} = \alpha_4 + \beta_4 D_i + W_j + X_i + \epsilon_{ij}. \quad (4)$$

We are interested in β_3 and β_4 : $\beta_3 > 0$ means that the encouragement treatment has a positive effect on quiz completion rates; $\beta_4 > 0$ means that the encouragement treatment increases quiz grades. The results are reported in Column (4) and (5) of Panel (A) in Table 2. Consistent with the summary statistics in Table 1, the encouragement treatment has a significantly positive effect on quiz completion rates: students who received the encouragement were 5.38% 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 participants from Week 2 to 5. Moreover, we do not observe a significant effect of the encouragement on the scores that students achieved 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 shows no distributional differences between the observed characteristics of students in treatment and control groups. 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 are 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 social engagement levels nor learning outcomes in the first week of the course.⁹

4.3. Average Treatment Effect of Social Engagement on Learning Outcomes

So far, we show that our encouragement treatment has a positive effect on both social interaction and learning outcomes. We then move to our main hypothesis: *can social interaction among students improve their learning outcomes?* In particular, we want to estimate the impact of student i 's number of visits to the discussion boards before Week j , denoted as $\tilde{v}_{ij}$, on her quiz completion rate in Week j , q_{ij} .¹⁰ Notice that we cannot obtain this effect by comparing the treatment and the control group due to the two-sided non-compliance problem mentioned in Section 3.2.

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

$$q_{ij} = \text{Probit}(\alpha + \gamma_j^1 \tilde{v}_{ij} + X_i + u_{ij}), \quad (5)$$

and

$$s_{ij} = \alpha + \gamma_j^2 \tilde{v}_{ij} + X_i + u_{ij}, \quad (6)$$

⁹ Another way to address these concerns is to employ a difference-in-difference (DID) model and treat week 1 and week 2-5 as the pre-period and post-period. Our results do not change qualitatively under such DID approach.

¹⁰ Our results do not change qualitatively if we measure social interaction by the number of visits in the previous week

where $\tilde{v}_{ij}$ represents the number of visits to the discussion board by student i before Week j , and X_i is a demographics control for student i (i.e., signature-track status, English locale, and email announcement statue). Equations 5 and 6 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., $\gamma_j \neq \gamma_{j+1}$).

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 social engagement levels as well as learning outcomes, such as student motivation and ability. To address this causal identification issue, we use aforementioned randomized encouragement treatment (i.e., D_i) as an 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 $\tilde{v}_{ij}$.*

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 is exclusion restriction:

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

This assumption means that the encouragement treatment only affects learning outcomes through 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 student 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 and observed controls (i.e., $cov(q_{ij}, D_i | \tilde{v}_{ij}, X_i) = 0$ and $cov(s_{ij}, D_i | \tilde{v}_{ij}, X_i) = 0$). 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 encouragement page.

With these two assumptions, Equations (5) and (6) can be estimated with the standard IV setup, and the average treatment effects of student visits to the discussion board on their quiz completion rates and scores can be identified (Angrist and Pischke 2008).¹¹ The estimates of interest are γ_j^i where $i \in \{1, 2\}$ and $j \in \{1, 2, 3, 4\}$. γ_j^1 is the average treatment effect of board visits on quiz

¹¹ We use conditional maximum-likelihood estimator to estimate the Probit model with IVs.

	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	0.0253*** (0.0042)	−0.268 (8.407)	0.0171*** (0.0036)	0.0012* (0.0006)
Marginal Effect	0.0101***	−0.268	0.00670***	0.0012*
Observations	335	163	335	163
Panel B: Average Effect of Social Engagement on Learning Outcomes (Week 4)				
Number of Visits	0.0158*** (0.0055)	0.0605 (0.542)	0.0097*** (0.0025)	0.0016* (0.0007)
Marginal Effect	0.00612***	0.0618	0.0037***	0.0016*
Observations	335	128	335	128
Panel C: Average Effect of Social Engagement on Learning Outcomes (Week 5)				
Number of Visits	0.0132*** (0.0023)	0.0008 (0.0117)	0.0103*** (0.0023)	0.0012* (0.0005)
Marginal Effect	0.00503***	0.0004	0.00385***	0.0012*
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 English locale.

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

completion rates in Week j , while γ_j^2 represents the average treatment effect of board visits on quiz scores in Week j . Notice that we have assumed that the treatment effect is homogeneous across individuals. In Section 4.4, we show that we actually estimate the local average treatment effect (LATE) if the treatment effect is heterogeneous. In Section 6, we then demonstrate how we generalize the LATE estimator by conducting another experiment with a different encouragement.

Table 3 shows the estimated results related to both measures of social engagement. In particular, Column (1) of Table 3 shows that, under both measures, the average treatment effects of board visits on quiz completion rates are significantly positive. In particular, in our five-week-long course, one more cumulative visit to the discussion board increases quiz completion rates by 1.0%, 0.6%, and 0.5% in Week 3, 4, and 5, respectively. Therefore, we find empirical evidence that is consistent with our Hypothesis 1. In other words, the positive effect of social interaction overweighs its negative effect, which suggests that the underlying dominant mechanism may be social learning, social facilitation or a combination of both.

Moreover, it is evident from Column (2) of Table 3 that more board visits have no effects on quiz scores throughout the course. This may be due to two reasons: (1) students are not allowed to discuss quiz related problems on the discussion board; (2) our board lacks sufficient tagging and recommendation functionalities; therefore, it is difficult for students to extract needed information. This suggests that, after visiting the discussion board, students are more likely to exert effort (i.e., taking the test) while they do not necessarily learn more knowledge (i.e., earning a higher test

score). This provides empirical evidence that social facilitation, instead of social learning, may be the dominant mechanism through which students impact each other on the discussion board.

Finally, it is also useful to compare the estimated effects through the IV approach with the estimates from the naive reduced-form specifications in Column (3) and (4) of Table 3. 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. This suggests that there exist confounding factors that are positively correlated with quiz completion rates and negatively correlated with social engagement levels. Furthermore, comparing Column (4) to Column (1) shows that simple OLS regression may lead researchers to an incorrect conclusion that social interaction in the discussion board also improves quiz scores. Hence, it is important to use experiment-based methods to extract reliable causal inference in MOOC design and analysis.

4.4. Differential Treatment Effects

So far, we have assumed that the average treatment effect of social interaction on learning outcomes is homogeneous across individuals. In practice, students may benefit differently from social interaction. As suggested in Hypothesis 2, signature-track students with high achievements may benefit less from social interaction. Moreover, heterogeneous treatment effects may indicate that our IV analysis only estimates the effect for the compliance population, students who are more likely to go the discussion board when receiving the encouragement. In this section, we first provide the framework of heterogeneous treatment effect and discuss its implication on our estimators. We then conduct a widely-used subgroup analysis (Rothwell 2005) to test our Hypothesis 3.

Differential Treatment Effects and Local Average Treatment Effect: Let $v_{ij}(1)$ denote the expected number of board visits from student i in week j when receiving the treatment, $v_{ij}(1) = E[v_{ij}|D_i = 1]$. Similarly, $v_{ij}(0) = E[v_{ij}|D_i = 0]$. Based on student responses to the encouragement treatment, they can be classified into three groups:

(1) **Never-takers:** students who never interact socially on the discussion board, regardless of whether they received the encouragement. In other words, student i belongs to this group if $v_{ij}(1) = v_{ij}(0) = 0 \forall j > 1$;

(2) **Always-takers:** students who always interact on the discussion board, regardless of whether they received the encouragement. Student i belongs to this group if $v_{ij}(1) = v_{ij}(0) > 0 \forall j > 1$;

(3) **Compilers:** students go to the discussion board more often if they received the encouragement. Student i is a complier if $v_{ij}(1) > v_{ij}(0) \forall j > 1$.

As the literature on differential treatment effects has documented (Angrist and Imbens 1995), with one additional monotonicity assumption, our previous estimated treatment effects, γ_j^1 and γ_j^2 ,

represent the *local average treatment effects* of social interaction on learning outcomes for those *compliers*. 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 a student's probability to visit the discussion board. While we cannot observe both $v_{ij}(0)$ and $v_{ij}(1)$ simultaneously for each student i and each week j , 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. In Section 6, we analyze another large-group experiment with a different IV to generalize our results beyond this compliance population.

Subgroup analysis: Subgroup analysis helps us understand important heterogeneity in treatment effects between different subsets of the population. As suggested in Hypothesis 3, we are interested in how social interaction may impact signature-track and non-signature-track students differently.¹²

In order to conduct our IV analysis and estimate the treatment effects, we must demonstrate that Assumption 1 is satisfied for both signature-track and non-signature-track students: our encouragement treatment has positive effects on both signature-track and non-signature-track students' social engagement levels. Panel (a) of Table 4 provides the answer: it shows that our encouragement treatment has significant effects on encouraging both signature-track and non-signature-track students to visit the discussion board and post more often. This is intuitive since, as illustrated in Appendix A, our formulation of the encouragement does not depend on any information related to signature-track status.

After verifying Assumption 1 for both subgroups, we can then re-estimate our IV estimators in Equation (5) and (6) on signature-track and non-signature-track students separately. Panels B to D of Table 4 shows our estimated effects of social engagement levels on learning outcomes for non-signature-track and signature-track students, respectively. Comparing Column (1) and (2) in Panels B to D of Table 4, we can see that social interaction in the discussion board increases completion rates only for non-signature-track students. This subgroup analysis reveals one important feature of social engagement on the discussion board: signature-track students, in our context, do not benefit from higher social engagement levels. This result is consistent with the past literature (Imberman et al. 2012): when mixing students with high and low achievements, students with high achievements may not benefit from social interaction.

¹² We also compute subgroup analysis based on other demographic information (i.e., English locale status and email announcement status). We do not find any statistically significant heterogeneities across these characteristics. We report this null result since the past literature has emphasized on the danger related to p-value fishing with sub-group analysis (Assmann et al. 2000).

	Dependent variable			
	Signature-track Students		Non-signature-track Students	
	Number of Visits	Number of Posts	Number of Visits	Number of Posts
	(1) Poisson	(2) Poisson	(3) Poisson	(4) Poisson
Panel A: Average Intent-to-treat Effect (Week 2- 5)				
Treatment	0.249*** (0.0223)	0.413*** (0.103)	0.162*** (0.100)	0.735*** (0.117)
Marginal effects	1.968***	0.166***	6.426***	1.845***
Observations	1028	1028	312	312
Quiz Completion Rate				
	(1) IV Probit		(2) IV Probit	
Panel B: Average Effect of Social Engagement on Learning Outcomes (Week 3)				
Number of Visits	0.0341*** (0.0056)		0.0190 (0.0118)	
Marginal Effects	0.00135***		0.00492	
Panel C: Average Effect of Social Engagement on Learning Outcomes (Week 4)				
Number of Visits	0.0231*** (0.0049)		−0.0032 (0.0171)	
Marginal Effects	0.00817***		−0.0012	
Panel D: Average Effect of Social Engagement on Learning Outcomes (Week 5)				
Number of Visits	0.0195*** (0.0024)		−0.0004 (0.0112)	
Marginal Effects	0.00686***		−0.0002	
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 English locale.

Table 4: Sub-group analysis

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 received the invitation email, and $G_i = 1$, if student i received the invitation and successfully conducted the pairwise discussion. Moreover, we use $M_i = 1$ to indicate that student i was motivated to conduct the one-to-one discussion after they had been assigned to a partner. Notice that a student completed 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 completed the feedback form were motivated 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, M_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 $M_i = 1$).

Receiving Invitations					
	Group 1	Group 2	Group 3	Treated Group	Control Group
	(1)	(2)	(3)	(4)	(5)
Panel A: Demographics Characteristics					
Signature Track	0.24 (0.06)	0.21 (0.05)	0.09 (0.01)	0.11 (0.02)	0.11 (0.02)
English Locale	0.41 (0.07)	0.51 (0.06)	0.44 (0.02)	0.47 (0.02)	0.45 (0.02)
Email Announcement	0.98 (0.02)	1.00 (0.00)	0.99 (0.02)	0.99 (0.00)	0.99 (0.00)
Panel B: Learning Outcomes (Week 2 - 5)					
Quiz Completion in Week 3	0.54 (0.07)	0.47 (0.06)	0.17 (0.01)	0.21 (0.01)	0.21 (0.02)
Quiz Score in Week 3	2.66 (0.11)	2.63 (0.12)	2.60 (0.06)	2.62 (0.05)	2.67 (0.06)
Quiz Completion in Week 4	0.46 (0.07)	0.36 (0.06)	0.13 (0.01)	0.16 (0.01)	0.17 (0.02)
Quiz Score in Week 4	4.16 (0.23)	3.73 (0.17)	3.65 (0.10)	3.74 (0.09)	3.93 (0.10)
Quiz Completion in Week 5	0.37 (0.07)	0.32 (0.05)	0.10 (0.01)	0.13 (0.01)	0.14 (0.02)
Quiz Score in Week 5	3.15 (0.23)	3.08 (0.18)	3.15 (0.12)	3.02 (0.09)	3.26 (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 5: Descriptive statistics of the small-group interaction experiment, by groups.

(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 $M_i = 1$).

(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 $M_i = 0$).

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

Table 5 displays the summary statistics associated with these groups of students. Columns (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 5 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 selected the subset of students to send the invitation email to. There are also no statistically significant differences between Column (1) and (2). This means that all students who wanted to conduct the discussion (i.e., $\{i|M_i = 1\}$) are homogeneous, regardless whether they have done so (i.e., $G_i = 1$ or $G_i = 0$). This suggests that students in Group 1 and 2 are identical except students in Group 1 are lucky to have partners that also want to conduct discussion. This is formally summarized in the following random partner-no-show assumption:

ASSUMPTION 4. *Random-no-show assumption:* $cov(X_{ij}, G_i | M_i = 1) = 0 \quad \forall X_{ij}$ such that $cov(X_{ij}, q_{ij} | M_i = 1) > 0$ or $cov(X_{ij}, s_{ij} | M_i = 1) > 0$.

There are considerable differences between the signature-track ratios of students who wanted to conduct the discussion (i.e., Columns (1) and (2)) and students who did not (i.e., Column (4)). 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 conducted the discussion, and 129 wanted to conduct the discussion. In other words, the effort cost of social interaction 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 Columns (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 did conduct the discussion 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:

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

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

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

$$s_{ij} = \alpha + \theta^4 G_i + W_j + X_i + u_{ij} \quad \forall i \text{ s.t. } \mathbf{M}_i = \mathbf{1}, \quad (10)$$

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.

Since the email invitation is randomly assigned, $cov(q_{ij}, A_i) = 0$ and $cov(s_{ij}, A_i) = 0$. This allows us to interpret the estimators θ^1 and θ^2 as causal effects. Notice that the sample in Equation (7) and (8) includes all students in the experiment (i.e., students in Group (3) and (4)), while the sample in Equation (9) and (10) only includes students who want to conduct the discussion (i.e., students in Group (1) and (2)). This is because Assumption 4 states that $cov(q_{ij}, G_i | M_i = 1) = 0$ and $cov(s_{ij}, G_i | M_i = 1) = 0$, and by focusing on all students with $M_i = 1$, we can interpret θ^3 and θ^4 as causal effects.¹³

¹³ This approach is the same as using M_i as an instrumental variable to estimate the effect of G_i on q_{ij} and s_{ij} . This is because the instrument M_i has perfect compliance in Group (1) and (2): students whose partners did not show up did not conduct the discussion while students whose partners showed up conducted the discussion.

	Quiz Completion Rate (1) Probit	Quiz Score (2) OLS
Panel A: Average effect of receiving the invitation on learning outcomes (Week 3 - 5)		
Treatment	-0.0001 (0.018)	-0.108 (0.0676)
Signature Track	1.952*** (0.026)	0.0886 (0.0656)
Observations	4431	755
Panel B: Average effect of conducting the discussion on learning outcomes (Week 3 - 5)		
Treatment	0.217** (0.069)	1.324*** (0.206)
Signature Track	2.265*** (0.218)	0.549* (0.224)
Marginal Effects	0.0593***	1.324***
Observations	387	159

(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 signature-track status, email announcement status, and English locale.

Table 6: Reduced-form estimates for small-group interaction experiment

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

Table 6 reports the estimation results: Panel A shows that receiving the invitation does not increase quiz completion rates nor 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” 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 effort 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 encouragement in the large-group interaction experiment have visited the board at least once after the encouragement, only 7% of students did conduct the pairwise discussion after receiving the invitation.

From Panel B of Table 6, it is evident that there is a large effect of actually conducting the one-on-one discussion on quiz completion rates, and more importantly, quiz scores.¹⁴ This shows that

¹⁴ These estimated treatments are only on students who want to conduct the one-to-one discussion. Those students represent roughly 61% of all students who fill out the demographics survey.

other forms of social interaction 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 quiz scores while higher level of social interaction in the discussion board does not. This empirical evidence supports our Hypothesis 3: synchronous interaction indeed benefits students differently than asynchronous interaction. One possible mechanism of this difference is that students can effectively exchange information and socially learn from each other during one-on-one discussions while they could not achieve this through the asynchronous communication on the discussion board.

Lastly, the effect on quiz scores could be due to cheating: students could discuss and finish the quiz together during the meeting. In order to test this mechanism, we compute the difference between quiz completion times of two students in each pair and denote this as inter-group quiz completion difference. Student pairs in Group 1 who successfully conduct the discussion have inter-group quiz completion difference ranging from 6.68 hours to 146.22 hours with an average difference at 104.88 hours, while student pairs in Group 2 who did not conduct the discussion have inter-group quiz completion difference from 4.08 hours to 200.15 hours with an average difference at 76.3 hours. Since the pairs of students who conducted the discussion do not have lower inter-group quiz completion difference, this empirical evidence suggests that students did not cheat and finish quizzes together in the pairwise discussion.

6. Generalizability

The major concern associated with our main result is generalizability. In particular, there are three sample selection processes that may limit us from generalizing our results to the entire student population. First, the survey request in our large-group interaction experiment was sent out at the beginning of the second week. It means that our sample does not include students who dropped out in the first week and therefore our result cannot apply to those students. Second, our large-group interaction experiment only includes students who were willing to give us feedback and in turn fill out our survey. These students may be more motivated than an average student, and our result on these students may not be representative. Third, our main result relies on an IV estimation strategy. There is a price to pay for the IV analysis: our estimated treatment effect of social interaction on learning outcomes is only on the compliance population. In other words, our results cannot be generalized to students who are *not* more likely to visit the discussion board if they receive the nudge.

In order to overcome these issues, we offered our class again in September, 2015 and re-conducted our large-group interaction experiment in this new offering. In this new offering, we had 6,312 registered students among which 149 enrolled in the signature-track; 3,365 visited the course at least once; 2,389 watched at least once lecture; and 798 browsed the discussion boards at least once. We made three key changes to address these three issues in the new offering:

	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 1 - 5)				
Encouragement Treatment	0.390*** (0.0277)	0.473* (0.117)	0.244*** (0.0557)	0.0426 (0.0549)
Signature Track	2,041*** (0.0301)	1.947*** (0.122)	1.624*** (0.0999)	0.107 (0.0710)
Marginal Effect	0.677***	0.0466***	0.0577***	0.0426
Observations	3130	3130	3130	675

(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 English locale.

Table 7: Reduced-form estimates of our treatment effects on social engagement levels and learning outcomes in the large-group interaction experiment (Second offering).

(1) The nudge is offered in our demographics survey instead of our Week 1 feedback survey. In this way, our sample also includes students who might drop out during Week 1.

(2) We told students that they would have a chance to win a \$25 Visa gift card if they fill out the demographics survey. Therefore, we attract not only students who want to give us feedback but also students who are motivated by this monetary reward.

These two changes have increased our sample size dramatically. In the first offering of the class, we have 335 participants in the large-group experiment. In this new experiment, we have 626 participants out of 3,365 learners who visited the course at least. This gives us a much higher 19% participation rate. The third modification is:

(3) Instead of using the shaming-expectation nudge, in this new experiment, we use a monetary reward: the treatment group receives a page telling them that the *top students* in the discussion board will receive a Visa gift card.

We use this new encouragement in the hope that the new compliance population could be different than the original one. In particular, we hypothesize that the old shaming-expectation nudge would work better on students who originally had low social engagement levels since these students experienced both the shaming and the expectation effects. Therefore, in this new nudge, we emphasize that only top participants in the discussion board can receive the gift cards. In this way, we hypothesize that the nudge will be more effective on students who originally have high social engagement levels since only those students think that they may be a top participant and win the gift card.

In order to conduct the IV analysis with this new encouragement, we first focus on the ITT effect of the nudge on social interaction and learning outcomes. Table 7 reports the result: it shows that

	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 B: Average Effect of Social Engagement on Learning Outcomes (Week 2)				
Number of Visits	0.126*** (0.0293)	0.0212 (0.0299)	0.0475*** (0.0107)	0.0027 (0.0046)
Marginal Effect	0.0434***	0.0212	0.0142***	0.0027
Observations	626	147	626	147
Panel A: Average Effect of Social Engagement on Learning Outcomes (Week 3)				
Number of Visits	0.0933*** (0.0129)	-0.0154 (0.0269)	0.0466*** (0.0085)	-0.0006 (0.0033)
Marginal Effect	0.0277***	-0.0154	0.0101***	-0.0006
Observations	626	101	626	101
Panel C: Average Effect of Social Engagement on Learning Outcomes (Week 4)				
Number of Visits	0.0716*** (0.0104)	0.0050 (0.265)	0.0367*** (0.0084)	0.0056. (0.0032)
Marginal Effect	0.0201***	0.0050	0.0059***	0.0056.
Observations	626	77	626	77
Panel D: Average Effect of Social Engagement on Learning Outcomes (Week 5)				
Number of Visits	0.0605*** (0.0064)	-0.0064 (0.0418)	0.0338*** (0.0059)	0.0012 (0.0026)
Marginal Effect	0.0197***	-0.0064	0.0040***	0.0012
Observations	626	62	626	62

(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 English locale.

Table 8: IV-estimated effect of social interaction in the discussion board ($\bar{v}_{ij}$) on learning outcomes (Second offering).

the monetary nudge also has a significant effect on both social interaction and learning outcomes, which means that we satisfy the inclusion restriction and can conduct our IV analysis properly.

We then focus on the average treatment effect of social interaction on learning outcomes in Table 8. It shows that social interaction increases learning outcomes measured by quiz completion rates. Moreover, the social interaction induced by the encouragement has no effect on quiz scores. These two results are consistent with previous results in Table 3. Hence, our main result is robust under a completely different student population with a completely different sample selection process and a completely different encouragement. Even though this cannot prove that our results indeed can be applied to the entire population, it demonstrates that our main result is not limited by our selection processes and IV strategy: it can be applied to a broader population.¹⁵

¹⁵ One possible concern related to this new nudge is that students may interpret the nudge as “top participants in the class” instead of “top participants on the discussion board” will receive gift cards. In this case, the nudge may directly affect the outcome variable, which invalidates the analysis. However, this concern is unlikely to happen due to two reasons. First, we emphasized in the statement that the performance is measured on discussion board. In the entire paragraph containing this statement, we only discussed the discussion board and avoided mentioning any other parts of the course. Second, if the students indeed interpret the statement as “top participants in the class (not only on the discussion board) will receive gift cards”, the nudge should have an effect on the quiz scores. However, Table 8 shows that this is not the case.

7. Discussion and Conclusion

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

Encouragement Design: The large-group interaction experiment in the first MOOC offering 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 experiments also provide 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 up to 4.3% in the subsequent week. This result validates the importance of better designed discussion boards that improve 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 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 is smaller than that on other 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 interaction in the discussion board. This is consistent with the past literature that shows that high achieving students do not benefit from interacting with low achieving students (Imberman et al. 2012). This no-effect result implies that, even though social interaction on the discussion boards may affect the non-signature-track and convince them to pay for the class later on, such interaction does not affect the "current" customers of the course, the signature-track students. This suggests that for-profit MOOC platforms should think outside of the box to experiment and design other kinds of social interaction that could affect both signature-track and non-signature-track students.

The discussion board and quiz scores: Our large-group interaction experiment results also show that social engagement levels have no effects on 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, as discussed, this result suggests that social facilitation may be the dominant underlying mechanism: students who go to the discussion board observe others' efforts and therefore are more likely to exert efforts. This has two implications: (1) Since MOOC students are more likely to benefit from the social facilitation mechanism, the MOOC platform should provide other channels for students observing each other's efforts. For example, MOOC platforms may consider providing billboards documenting the average progress of students. (2) It is important to re-design the discussion board so that students can learn from each other and achieve higher test scores.

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 discussions 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 quiz completion rates but also increases final quiz scores. This is in stark contrast with the large-group interaction experiments where we found that social interaction in the discussion board only increases completion rates. One interpretation is that one-on-one discussion is synchronous and more conducive to collaborative learning. This result encourages MOOC practitioners and researchers to look for other direct communication channels for students to socially engage with each other and improve learning outcomes.

Given the demonstrated improvements of service quality from social interaction, 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 interaction that benefit all students. Lastly, it is interesting to extend our findings to other service sectors and study the effect of participants' social interaction on service quality in other settings. We suggest that future researchers start with online services where randomized encouragement treatment can be designed, tracked and analyzed with relatively low costs.

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.
- Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.
- 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.
- 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.
- Mary A Burke and Tim R Sass. Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1):51–82, 2013.
- Antoni Calvó-Armengol, Eleonora Patacchini, and Yves Zenou. Peer effects and social networks in education. *The Review of Economic Studies*, 76(4):1239–1267, 2009.
- 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*, 2013.
- Coursera. Community — coursera. <https://www.coursera.org/about/community>, 2015. (Visited on 06/08/2015).
- Esther Duflo, Rachel Glennerster, and Michael Kremer. Using randomization in development economics research: A toolkit. *Handbook of development economics*, 4:3895–3962, 2007.
- Dennis Epple and Richard E Romano. Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, pages 33–62, 1998.
- Armin Falk and Andrea Ichino. Clean evidence on peer effects. *Journal of Labor Economics*, 24(1):39–57, 2006.
- Jane Cooley Fruehwirth. Can achievement peer effect estimates inform policy? a view from inside the black box. *Review of Economics and Statistics*, 96(3):514–523, 2014.

- Caroline Hoxby. Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research, 2000.
- Caroline M Hoxby and Gretchen Weingarth. Taking race out of the equation: School reassignment and the structure of peer effects. Technical report, Working paper, 2005.
- Scott A Imberman, Adriana D Kugler, and Bruce I Sacerdote. Katrina’s children: Evidence on the structure of peer effects from hurricane evacuees. *The American Economic Review*, pages 2048–2082, 2012.
- Bing Jing. Social learning and dynamic pricing of durable goods. *Marketing Science*, 30(5):851–865, 2011.
- Daphne Koller, Andrew Ng, Chuong Do, and Zhenghao Chen. Retention and intention in massive open online courses: In depth. *Educause Review*, 48(3):62–63, 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.
- Victor Lavy and Analia Schlosser. Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, pages 1–33, 2011.
- David S Lyle. Estimating and interpreting peer and role model effects from randomly assigned social groups at west point. *The Review of Economics and Statistics*, 89(2):289–299, 2007.
- 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.
- Bruce Sacerdote. Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, 116(2):681–704, 2001.
- 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.
- Qiuping Yu, Gad Allon, and Achal Bassamboo. How do delay announcements shape customer behavior? An empirical study. *Management Science*, under review, 2013.
- Robert Boleslaw Zajonc et al. *Social facilitation*. Research Center for Group Dynamics, Institute for Social Research, University of Michigan, 1965.
- David J Zimmerman. Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics*, 85(1):9–23, 2003.

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

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

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.

	Models			
	(1) Poisson	(2) Random-Effect Poisson	(3) AR(1) Poisson	(4) Week 2
Panel A: ITT Effect of Encouragement Treatment on Number of Visits (Week 2 - 5)				
Encouragement Treatment	0.192*** (0.0141)	0.340. (0.195)	0.181*** (0.0222)	0.243*** (0.0279)
Signature Track	1.611*** (0.0143)	1.639*** (0.245)	1.544*** (0.0223)	1.053*** (0.0277)
Marginal Effects	2.944***	5.322.	2.780***	3.799***
Panel B: ITT Effect of Encouragement Treatment on Number of Posts (Week 2 - 5)				
Encouragement Treatment	0.613*** (0.0604)	0.458* (0.215)	0.557* (0.0967)	0.435*** (0.122)
Signature Track	1.874*** (0.0599)	1.872*** (0.246)	1.785*** (0.0956)	1.169*** (0.118)
Marginal Effects	0.550***	0.474***	0.503***	0.369***
Observations	1340	1340	1340	335

(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.

Table 9: Intent-to-treat effect of our encouragement treatment on social interaction with different robust specifications.

B. Robustness

In this Appendix, we demonstrate that our intent-to-treat result in our large-group experiment – the encouragement treatment has a positive effect on social interaction – is robust to different specifications of standard error correlations.

In Section 4.2, Equation 1 estimates the intent-to-treat effects of encouragement treatment on social interaction, which verifies Assumption 1 for our main IV analysis. The major concern associated with these specifications is that student performance may be correlated over weeks. In other words, since each student has 4 observations over 4 weeks in these specifications, the standard errors of 4 observations for each student may be correlated over time. This correlated standard error structure may bias our estimators. In order to address this problem, we employ three more complex specifications to demonstrate the robustness of our results:

(1) We use a random-effect model to directly model each student’s performance and alleviate the concerns related to student-level correlations. The new specification for Equation 1 is $v_{ij} = \text{Poisson}(\alpha_1^1 + \beta_1^1 D_i + S_i + W_j + R_i + \epsilon_{ij})$, where R_i is the student-specific random effect.

(2) We directly model standard errors over time for each student as an autoregressive (AR) process. In particular, we assume that the standard errors of our pooled OLS regression in Equation 1 has an AR(1) structure: $\epsilon_{i,j+1} = \kappa_1 + \kappa_2 \epsilon_{ij} + \zeta_{ij}$. This specification allows us to correct our estimators against autocorrelations of one student’s observations over time.

(3) We focus only on Week 2 observations and test whether the encouragement treatment has an effect on social interaction in Week 2. In this way, we eliminate the correlation problem since each student only has one observation. However, our estimators become less efficient since we do not use all the data.

Table 9 presents the results of these three specifications along with the original specification in Equation 1. As we can see, the encouragement treatment has a positive intent-to-treat effect on social interaction in all

specifications. The magnitude of the new estimators in Columns (2), (3), (4) are within 72% to 162% of the original OLS specification in Column (1). This demonstrates that the original result – our encouragement indeed increases social interaction – is robust and we can use it as an assumption for our main IV analysis.¹⁶

¹⁶ Notice that we do not need to replicate the same analysis for the IV analysis since the IV analysis is week-by-week and in turn does not have correlated standard error issues.