I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15964

Does the Closeness of Peers Matter? An Investigation Using Online Training Platform Data and Survey Data

Xin Gu
Haizheng Li

FEBRUARY 2023

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15964

Does the Closeness of Peers Matter? An Investigation Using Online Training Platform Data and Survey Data

Xin Gu
Georgia Institute of Technology

Haizheng Li
Georgia Institute of Technology and IZA

FEBRUARY 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Does the Closeness of Peers Matter? An Investigation Using Online Training Platform Data and Survey Data*

We study peer effects in online training participation using unique data from a large-scale online teacher training program. The platform data allow us to observe the accurate duration of attendance for every individual-lecture pair. We classify peer groups as close peers, local peers, and global peers based on their relationships. By controlling for unobserved heterogeneity, we find positive effects of close and local peer appearance on trainees' joining a lecture and on their length of stay in the lecture. However, global peers generate a negative but economically insignificant impact. Peer effects differ by group and increase with the relationship closeness. Using the survey data, we investigate the mechanisms of peer influences and find that social interactions facilitate online peer effects. Peer pressure and reputation concerns also help explain our findings. Our results shed new light on how peer effects can be utilized to improve the effectiveness of online learning.

JEL Classification: I21, J24, M53

Keywords: peer effects, online training

Corresponding author:
Xin Gu
School of Economics
Georgia Institute of Technology
221 Bobby Dodd Way
Atlanta, GA 30332
USA
E-mail: xin.gu@gatech.edu

* We thank YouChange China Social Entrepreneur Foundation for the data. We are grateful to Dylan Brewer, Patrick McCarthy for helpful comments. All errors are our own.

1 Introduction

Most studies about peer effects in education are related to the in-person environment. However, very few studies investigate peer effects in the online setting. Since the outbreak of Covid-19, it has become common to host classes, training programs, and meetings online, triggering the explosive growth of online education. For example, during the peak of the Covid-19 pandemic, most schools were forced to move instructions online. As a result, the effectiveness of online education has increasingly drawn attention to scholars, governments, and the public at large (Brown and Liedholm, 2002, Figlio et al., 2013, Alpert et al., 2016).

In this paper, we study peer effects in online training program participation. Participating in an online program as planned is the key first step to improve the quality of online learning (Hrastinski, 2009). We use both the survey data and the data recorded by the digital instruction platform of a large-scale online teacher training program. The unique data allow us to overcome the challenges that are common in identifying peer effects. Based on their relationship, we categorize peers into *global peers*, *local peers*, and *close peers* to estimate their influences. This structure of peer groupings helps us understand the mechanisms of peer effects.

To our best knowledge, our paper is among the first studies exploring high-frequency data in studying peer effects in online education. This study contributes to the literature first by controlling for unobserved individual and lecture heterogeneity in online participation when identifying peer effects. The inclusion of individual and lecture fixed effects helps control for unobserved confounders that are common to peer groups. Second, our measures of participation include not only the attendance of a lecture but also the length of stay in a lecture, thus allowing us to study peer effects at different levels of participation. Third, we can distinguish peer effects across a spectrum of the strength of the social ties amount peer groups. Finally, with the program survey data, we can further investigate the working channels of peers in influencing each other. Our results shed new light on how peer effects influence the learning effectiveness of online education programs.

Our results show that the appearance of local and close peers generates positive influences on attendance of a lecture and on the time of stay in a lecture. In contrast, global peers have a negative but economically insignificant impact on participation . The results imply that local and close peers are complements in participation decisions, while global peers are substitutes. We discover that peer effects increase with the closeness of peer groups, which can be explained by peer pressure, reputation concerns, and social interactions.

The rest of the paper is organized as follows. Section 2 introduces a theoretical framework. Section 3 describes the data. Section 4 presents and discusses the results. Section 5 investigates the potential working channels. Section 6 concludes.

2 A Theoretical Framework

We extend Blume et al. (2015) to model online peer effects in linear social interactions. Assume a trainee i takes an action ω_{il} in the lth lecture of a course to maximize the utility:

$$\max_{\omega_{il}} \left[px_i + \gamma x_l + \delta f\left(x_{-i}\right) + z_{il}\right]\omega_{il} - \frac{1}{2}\omega_{il}^2 - \sum_{g=k,r,s}\left[\frac{\phi^g}{2}\left(\omega_{il} - \sum_{j\in\Omega_{-i}^g}\alpha_{ig}\omega_{jl}\right)^2\right]. \quad (1)$$

The utility function is additively separable. The first term denotes the utility gain from action ω_{il}. The utility gain depends on x_i that includes observed personal traits of i, x_l that represents lecture and instructor heterogeneity, x_{-i} that contains peer characteristics, and z_{il} that captures the random shock. $f\left(x_{-i}\right)$ is a function of i's peer characteristics, reflecting the contextual effects defined by Manski (1993). It is generally difficult to observe peer traits in an online setting, for example, when trainees do not turn on video or put on their actual picture. Thus, for simplicity, we assume no contextual effects on individual action ω_{il}, i.e., $\delta = 0$ in the model. The term $\frac{1}{2}\omega_{il}^2$ denotes convex costs of taking the action ω_{il}. The last term is the sum of the squared distance between i's action ω_{il} and the weighted averages of i's peer actions $\sum_{j\in\Omega_{-i}^g}\alpha_{ig}\omega_{jl}$ in group $\Omega_{-i}^g, g = k, r, s$, to describe the prosocial behaviors that generate peer effects. α_{ig} is the weight of actions between i and his or her peers in group Ω_{-i}^g, assumed to be identical for all $j \in \Omega_{-i}^g$. We define group Ω_{-i}^g's collective action as $\omega_{jl}^g \equiv \sum_{j\in\Omega_{-i}^g}\omega_{jl}$. Furthermore, we classify i's peers $-i$ into three groups Ω_{-i}^k, Ω_{-i}^r, and Ω_{-i}^s. Without loss of generality, we denote Ω_{-i}^k as *global peers*, Ω_{-i}^r as *local peers*, and Ω_{-i}^s as *close peers*.

We follow the social network literature and measure peer effects at multiple levels with different strengths of social links (Calvó-Armengol et al., 2009, Goldsmith-Pinkham and Imbens, 2013, Bramoullé et al., 2009, 2014). There exists other literature looking into peer effects from more than one group at the same time, for example, from one's study group and roommates (Jain and Kapoor, 2015) or from short-term peers and long-term peers (Patacchini et al., 2017). Our group divisions reflect the scope, the size, and the intimacy of peers, and are common in reality. For example, global peers could be i's unknown colleagues with a large group size. Local peers could be some known colleagues with whom i occasionally collaborates. They usually come from other business units with a middle group size. Close peers could be those with whom i stays closely. They work in the same unit with the smallest group size. We let ϕ^g measure the strength of i's social ties with group Ω^g_{-i}. Intuitively, unknown colleagues have too few chances to meet each other, and thus have the weakest connections with i. Collaborators sometimes work on a joint project, forming moderate linkages with i. Team members interact most regularly, developing the strongest relationships with i. Therefore, we assume $|\phi^k| < |\phi^r| < |\phi^s|$ to reflect the relationships between group Ω^g_{-i} and i. As stronger relationships may lead to larger peer effects, the means that can strengthen peer relationships, like social interactions, may drive our results in the later section.

We use the absolute terms of ϕ^g in this assumption because peer effects sometimes can be negative (Lazear, 2001, Card and Giuliano, 2013, Carrell et al., 2018). If peers generate negative effects, they are substitutes to i in taking the action ω_{il}. On the contrary, peers are complements if their influences are positive (Blume et al., 2015). According to the utility maximization problem (1), any deviation from group Ω^g_{-i}'s actions, regardless of its direction, produces utility to i if $\phi^g < 0$ (substitution effect) or disutility to i if $\phi^g > 0$ (complementary effect). Because such a deviation may affect reputation, thus the individual i tends to conform to group social norms. In our context, for example, the substitution effect may result from the possibility that the absence may not be noticed. In contrast, the complementary effect occurs if the pressure from peers makes one follow group actions. Therefore, peer effects depend on the relative strength of the substitution and complementary effects.

In addition, the strength of peer relationships $|\phi^k| < |\phi^r| < |\phi^s|$ implies that a rational i should assign the weights to his or her peers following $0 < \alpha_{ik} < \alpha_{ir} < \alpha_{is}$. That is, closer peers are weighed more in trainees' decision-making. In addition, we have the weights summing up to 1, i.e., $\sum_{g=k,r,s} \sum_{j\in\Omega^g_{-i}} \alpha_{ig} = 1$. Since the group size becomes smaller for closer peers, it also suggests that $0 < \alpha_{ik} < \alpha_{ir} < \alpha_{is}$. An individual $j \neq i$, with characteristics $x_j \neq x_i$, has a different weighting scheme $\alpha_{jg} \neq \alpha_{ig}, g = k, r, s$. For example, females may differ from males in connections with peers from the same group. Since the YTEP is a job training program, trainees with different tenure statuses may also have asymmetric responsiveness to peer behaviors.

The F.O.C. of the utility maximization problem (1) w.r.t. ω_{il} indicates

$$px_i + \gamma x_l + z_{il} - \omega_{il} - \sum_{g=k,r,s} \phi^g \left(\omega_{il} - \alpha_{ig}\omega^g_{-il}\right) = 0. \tag{2}$$

We solve Equation (2) for ω_{il} as

$$\omega_{il} = \frac{px_i + \gamma x_l + z_{il}}{1 + \sum_{g=k,r,s} \phi^g} + \frac{\sum_{g=k,r,s}\left(\phi^g \alpha_{ig}\omega^g_{-il}\right)}{1 + \sum_{g=k,r,s} \phi^g}. \tag{3}$$

Equation (3) generates peer effect comparative statics:

$$\frac{\partial \omega_{il}}{\partial \omega^{g'}_{-il}} = \frac{\phi^{g'}\alpha_{ig'}}{1 + \sum_{g=k,r,s} \phi^g}, g' = k, r, s. \tag{4}$$

By $|\phi^k| < |\phi^r| < |\phi^s|$ and $0 < \alpha_{ik} < \alpha_{ir} < \alpha_{is}$, peer effects are ranked by

$$\left|\frac{\partial \omega_{il}}{\partial \omega^k_{-il}}\right| < \left|\frac{\partial \omega_{il}}{\partial \omega^r_{-il}}\right| < \left|\frac{\partial \omega_{il}}{\partial \omega^s_{-il}}\right|. \tag{5}$$

It indicates that i's action ω_{il} is affected the least by group Ω^k_{-i} (global peers), the second by group Ω^r_{-i} (local peers), and the most by group Ω^s_{-i} (close peers). Our model generates an insight that the size of peer effects depends on ϕ^g, the closeness of social relationships between Ω^g_{-i} and i, and on α_{ig}, the weighing scheme of i on his or her peers. In other words, closer peers generally exert larger peer effects. And individuals with different social backgrounds are affected asymmetrically by the same group of peers.

Let $\rho = \frac{p}{1+\sum_{g=k,r,s}\phi^g}$, $\delta = \frac{\gamma}{1+\sum_{g=k,r,s}\phi^g}$, $e_{il} = \frac{z_{il}}{1+\sum_{g=k,r,s}\phi^g}$, and $\theta^g = \frac{\phi^g \alpha_{ig'}}{1+\sum_{g=k,r,s}\phi^g}$, we rewrite Equation (3) as

$$\omega_{il} = \sum_{g=k,r,s} \theta^g \omega^g_{-il} + x_i\rho + x_l\delta + e_{il} \tag{6}$$

We use Equation (6) to guide our empirical specifications in Section 4. More specifically, we use various indicators available from data to measure i's action ω_{il} of participation. We calculate the related measures of peer participation ω^g_{-il}. x_i contains the demographic covariates. x_l captures the lecture characteristics. e_{il} is a random error.

3 Data

The data are from a large online teacher training program in China, known as the Young Teacher Empowerment Program (YTEP).[1] We use the data from the training year of 2019-2020, which consists of two general courses in Fall 2019 and twelve field courses in Spring 2020. All trainees are required to enroll in general courses. In contrast for field courses, trainees could self-select into any course based on the subjects. To have a larger sample size, we restrict our study to the general course *Career Development*. We choose this course rather than the other general course *Teacher Ethics* for two reasons: i) *Career Development* sees more active participation than *Teacher Ethics*; ii) *Career Development* has homework assignments while *Teacher Ethics* does not, allowing us to further investigate homework-based social interactions in Section 5.[2]

Online Lectures were held in the evening on an instructional platform. Figure 1 demonstrates the user interface of the platform. In particular, the participating trainees can observe a list of participants and nonparticipants (who enrolled in *Career Development*). The list refreshes whenever a trainee enters or exits the lecture room, and therefore, the trainee can observe the dynamic changes in peer participation. Each trainee has a unique identifier whose naming format is "*County + ID# + Full Name*". The list is sorted so that the participants from the same county are placed adjacently. Given this feature, a lecture participant could easily observe peers from the same county and recognize school colleagues by name, generating potential peer effects.

The data contain 8,627 trainees who are teachers from rural elementary and middle schools. These trainees come from 62 counties in 17 different provinces. The designated online instruction platform provides detailed records of synchronous participation minutes at the lecture level for every trainee.

The summary statistics of the participation and peer measures are shown in Table 1. We define participation as the number of lectures attended and the time spent on each lecture. We first use the number of lectures participated to measure participation defined by whether a trainee showed up in a lecture. We observe trainees' lecture attendance from two sources, survey self-report

[1] The YTEP program was initiated by the Youcheng China Social Entrepreneur Foundation. The YTEP details: `http://www.youcheng.org/news_detail.php?id=645`

[2] The results using the *Teacher Ethics* course data are similar.

data and platform data. More specifically, the routine program survey asked respondents to check a list of lectures that they attended, as well as to provide other information. The survey has two waves, one at the end of Fall 2019 (also the end of the *Career Development* course) and the other upon completion of the program in Spring 2020. The response rate is about 40% in each wave. We combine the two waves of respondents for a larger sample size. In addition, we observe everyone's accurate duration of attendance in minutes at the lecture level.

The course has 16 lectures in total. Based on the platform data, we calculate the number of lectures participated by a trainee if the duration of lecture attendance is greater than 0. Table 1 shows lecture participation based on both types of data. The two participation measures differ due to misreporting. For example, a typical trainee shows up in 12-13 lectures based on the survey response, which is about 80% of the entire course. However, based on the platform data, the average number of participated lectures is 10, which only accounts for 60% of the whole course. Figure 2 shows that the distribution of the number of lectures participated is more left-skewed in the survey data than in the platform data. Therefore, the auto-recorded data from the platform can reduce measurement errors in the self-reported data.

Figure 3 illustrates the distribution of the duration of attendance for all 8,627 trainees by lecture. The top line (solid triangle) displays the length by lecture. On average, a lecture lasts for about 95 minutes. The lower line (solid dot) plots the mean duration of attendance by lecture. For instance, an average trainee spends 25 minutes in lecture #6. The boxplot displays the 25th, 50th, and 75th percentiles of the duration of attendance for each lecture. For example, the box of lecture #9 shows that the 25th percentile of the duration of attendance is 0, the 50th percentile is 50 minutes, and the 75th percentile is 82 minutes. Over 75% of trainees do not show up in the first two lectures. After that, the overall attendance rate gradually improves and stays steady till lecture #16. This figure illustrates that participation varies by lecture. In addition, instructors usually differ across lectures of a course. Therefore, it is important to account for the cross-lecture heterogeneity in the individual-lecture analysis in the later section.

We identify trainees' peer networks by their county locations and school affiliations documented in the program registration. Supplementing the participation data with this offline connection information, we calculate the number of peers by the three groups: i) *global peers* attend the same lecture, ii) *local peers* attend the same lecture and come from the same county, iii) *close peers* attend the same lecture and come from the same school. We construct these three peer groups exclusively, that is, global peers account for the lecture participants other than those coming from the same county, and local peers include the same county participants other than those who work in the same school. No overlap between any two peer groups can reduce the correlations between peer regressors. Close peers have the strongest relationships because they are colleagues and work together at the same school every day. Moreover, they interact frequently, compete directly in tenure reviews, and care about reputation in the network. Local peers have the second closest relationships because they may occasionally see each other offline at some local meetings. Additionally, the program coordinators group their county participants on WeChat where they may communicate with each other.[3] Global peers have the weakest relationships because they only meet online and usually do not know each other offline. Based on Equation (5), we predict that close peers exert the largest peer effects, followed by local peers, and then global peers.

Based on the classification, we construct peer measures at the individual-course and individual-lecture levels, respectively. To proxy peer effects at the course level, we sum up the platform-recorded individual lecture attendance for peers from the three groups and then take the group average to calculate the average number of lectures participated for each peer group. Moreover, based on the peer appearance illustrated in Figure 1, we calculate the number of peers who showed up in the meeting room at each lecture and generate 138,032 observations at the individual-lecture level (total of 8,627 individuals by 16 synchronous lectures).

Table 1 also shows the summary statistics of peer measures with the number of peers calculated using the data from the platform because the survey has a smaller sample and cannot cover all peers who attended the lecture. This is another advantage of combining survey data with platform data. It also helps mitigate the selection issue in participating in a survey.

Table 2 presents some personal characteristics in the survey data. For example, males account for 20% of trainees who responded to the survey. Approximately

[3]WeChat is a popular Chinese instant messaging smartphone application, similar to WhatsApp.

30% of them are married with around 20% having at least one child. Roughly 70% of them are teachers college graduates. The trainees are mostly young teachers with about 2 years of teaching experience. Nearly 40% of them hold tenured positions and have already received indefinite contracts. As a result of the guaranteed employment, the incentive for those trainees to improve their skills may differ from those who have not been granted tenure yet, thus affecting the degree of peer influences. Additionally, 20% of the trainees report slow internet speed in their online participation.

4 Empirical Results

In this section, we estimate the peer effects at different levels using the data described above. We measure participation using both the program lectures attended and the duration of staying in a lecture. The first measure provides information about the peer effects in terms of encouraging attending lectures (the total number of lectures attended and whether attending a particular lecture), but it does not capture the intensity of the participation for each lecture. The second measure captures the time a participant stays in each lecture.

4.1 Peer Effects on Participation

Traditionally, we typically observe individual participation information at the aggregate course level, for example, the number of lectures participated. In the survey, trainees are asked to report their lecture attendance. The self-reported attendance may contain measurement errors. More importantly, the survey cannot provide information about peer participation because the sample is only a part of the trainees enrolled in the program. The advantage of our platform data is that it provides accurate information on the participation of each trainee and information about the trainee's complete peer network.

We first estimate the effect of peers on participating in lectures offered by the program. We adopt two definitions of lecture participation to take advantage of our data. First, a trainee is considered a participant in a lecture if the platform time of staying in the lecture is greater than 0, and second if the participant stayed over 30% of the whole lecture time. With these two measures of attendance, we construct the peer measures for each trainee by averaging the related participation measures (>0 minutes or >30% lecture time) of his or her global, local, and close peers.

In estimating the models of participation, we face the issue of no participation, because some trainees enrolled did not take part in any lecture and many of them missed some lectures during the program. Given the existence of zeros for the dependent variable, we apply the Tobit model in the estimation to take into account the difference between no participation and other amounts of participation. Another advantage of the Tobit model is that it can separately estimate the effect of peers on the probability of participation and the length of participation.

The estimated results are reported in Table 3. We first show the OLS estimates for comparison because OLS requires less restricted assumptions about the distribution of the error term. The estimated Tobit coefficients suggest that trainees participate in more lectures when their peers do so. If global, local, and close peers on average participate in 1 more lecture, an average individual attends 0.237, 0.310, and 0.553 more lectures, respectively. Close peers generate the largest effect, local peers the second, and global peers the smallest, consistent with our theoretical prediction. When comparing Column (2) with Column (1), the coefficient estimates do not differ much in the two models, with the OLS estimates marginally larger for the effects of global and local peers and smaller for the close peer effects. This result is not surprising as Figure 2 shows little censoring for participating in lectures.

Other covariates have little impact on course participation except for gender and years of teaching experience. For instance, the number of lectures participated by males is significantly smaller than females. The gender gap in participation is nearly 1 lecture on average. Senior teachers show up in more lectures than their junior counterparts holding other things constant, which may be attributed to their demand on updating knowledge and skills.

Furthermore, we define effective participation as staying in a lecture longer than 30% of its length. A lecture lasts for about 95 minutes shown in Figure 3, therefore, 30% is roughly 30 minutes. With the stricter measure on participation, global peer and close peer estimates become larger. In contrast, local peer estimate turns smaller. However, their differences are very small, showing that after a trainee entered a lecture, the length of peer stay does not generate much additional impact. Additionally, the estimates of other regressors are generally consistent with those in Column (2).

One concern for the above model is that we use the number of lectures attended by various peer groups to estimate their impacts on the number of lectures a trainee participated in. In this case, it requires that trainees can observe or form ex-ante expectations about peer participation. In other words, when deciding how many lectures to attend, one needs to know how many his or her peers will attend. This is a strong assumption in the online setting, especially for most global and some local peers, to generate potential peer effects. Therefore, to better connect with the mechanisms of peer influences, we take advantage of the platform data on the participation in every lecture for all trainees. More specifically, we estimate the effect of peer appearance by group on whether a trainee shows up and how long to stay in a lecture.

We first investigate peer effects on whether a trainee participates in a lecture. Based on Equation (6) in the theoretical framework, we assume that trainee i's attendance decision on lecture l is governed by the following linear equation:

$$D_{il} = \pi_0 + \sum_{g=k,r,s} \pi^g_{peer} N^g_{-il} + d_i + d_l + \mu_{il}, \tag{7}$$

where D_{il} is a dummy variable equal to 1 if i participated in lecture l, and 0 otherwise. N^g_{-il} is the total number of participants $-i$ other than i who attended the lth lecture for three exclusive groups of peers Ω^k_{-i}, Ω^r_{-i}, and Ω^s_{-i}. To estimate π^k_{peer}, π^r_{peer} and π^s_{peer}, we employ the fixed effect model by introducing personal dummies d_i and lecture dummies d_l. However, the control of fixed effects alone prevents us from uncovering covariates that potentially affect lecture participation. Hence, we adopt the specification that replaces the individual fixed effects d_i with a set of personal characteristics X_i for comparison. But there are two sources of endogeneity in the model specification with X_i: i) unobservables, e.g., trainees' abilities or motivations and instructors' teaching methods, that determine participation may meanwhile affect trainees' responses to peer appearance N^g_{-il}; ii) these unobserved confounders may be correlated with some covariates X_i, e.g., education or years of experience. Therefore, the control of fixed effects helps resolve the endogeneity issues resulting from unobserved heterogeneity (Hanushek et al., 2003, Lin, 2010).

A typical identification issue is endogenous peer group formation (Sacerdote, 2001, Zimmerman, 2003, Carrell et al., 2013, Lu and Anderson, 2015). In our case, the course is mandatory for all trainees, so the selection into the course is not a major concern. Moreover, individual fixed effects can control for time-invariant group common shocks and lecture fixed effects can deal with time-varying confounders that are common to peer groups. Another concern is that trainees are self-selected into the training program. In the YTEP, schools and local education administrations decide whom to enroll. This feature helps mitigate the self-selection problem. However, administrators from the school and/or the local administration may select teachers into the YTEP based on their characteristics unobservable to econometricians. In this case, the inclusion of individual fixed effects can control for the unobserved individual heterogeneity.

Furthermore, we employ the leave-one-out specification to construct our peer measures N^g_{-il}, to address the reflection problem between individual participation and peer participation (Manski, 1993). That is, an individual, him- or herself is excluded from the count of his or her peers (Angrist, 2014, Carrell et al., 2018). Additionally, the potential simultaneity bias in our setting should be minimal because the average size is over 4,000 for global peers in a lecture and about 200 for local peers. Given the large peer group sizes, the causal direction should be overwhelmed by the many-to-one peer effects on one trainee rather than the one-to-many reflection on his or her peers. It could be an issue for the estimated effects of participating close peers as the average size is about 5. Nonetheless, our estimated close peer effects can be seen as an upper bound. The literature alternatively uses lagged peer measures to solve simultaneity problems (Burke and Sass, 2013). In our context, two consecutive lectures are one week apart such that the intertemporal effects likely fade away.

We report the estimates of Equation (7) in Table 4. As we go to the individual-lecture level, the number of observations largely increases to around 100,000. We estimate a linear probability model with OLS to compare with Probit using the same specification with personal characteristics X_i. The coefficient estimates show positive peer effects on the probability of one attending a lecture. Specifically, the Probit average marginal effects for global, local, and close peers are 0.001, 0.003, and 0.079 respectively, much smaller than those in OLS. The results suggest if the number of peers who decide to participate in a lecture increases by 10 in global, local, and close peer groups, the probability of one attending that lecture as well increases by 0.1%, 0.3%, and 7.9% respectively. The estimated peer effects on the participation likelihood are larger when peers get closer.

The sign of personal covariate estimates is also largely in line with intuition.

For example, males are less likely to attend a lecture than females, leading to the gender difference in participation found at the course level. Because of family constraints, married trainees have a significantly lower probability to show up than their counterparts. However, those with children have a significantly larger propensity to attend a lecture. This counterintuitive estimate may result from the potential correlation between the chance of having children and the years of teaching experience. The graduates from teachers colleges have less incentive to go to the lectures, probably because they already had teacher training at school. Tenured teachers are less active in lecture participation than untenured ones, perhaps driven by their career motivations. We elaborate on the differences in participation by tenure status in Section 5. Conditional on teachers college background and tenure status, experienced teachers do not differ from inexperienced ones in the probability of attending a lecture. In addition, poor internet connectivity significantly deters online participation.

To correct the omitted variable bias due to the unobservables, we estimate the specification in Equation (7) with the fixed effect model. The change in the estimated peer effects from Column (2) to Column (3) is largely due to the control of other time-invariant unobservables. After including individual dummies, we get much larger estimates for local and close peer effects. On the contrary, the global peer estimate becomes negative. It suggests that the omitted variables introduce a negative bias to the local and close peer effects and a positive one to the global peer effects. The positive local and close peer effects are consistent with intuition. However, the global peer effects are negative. It implies that global peers play a disruptive role in a trainee's decision process. In other words, global peers are substitutes but local and close peers are complements in participation decisions.

Moreover, we apply the same model specification with individual and lecture fixed effects to the full sample. The results suggest if 10 more global peers decide to attend a lecture, one is 0.2% less likely to do so. This negative global peer effect is economically insignificant, although statistically significant. If these 10 more peers are county colleagues or school coworkers, the probability for one to attend a lecture increases by 2.3% and 31.3% respectively. Intuitively, trainees conform to the behaviors of the groups that they stay close to and do not care much about those remote from them. This is possibly driven by the fact that trainees may well coordinate with their local peers in county group chats and close peers privately, whereas they have limited means to communicate with global peers (outside their county) about whether to participate in an upcoming lecture. In addition, the absence triggers little embarrassment among unknown peers. But peer pressure drives one to participate in a lecture if more familiar colleagues do so. And the pressure becomes larger when peers get closer. These mechanisms generate larger effects for closer peers. From Column (3) to (4), we see a slight increase in all three group peer effects, indicating that the sole use of survey respondence negatively biases peer effect estimates in trainees' decisions on whether to join a lecture.

4.2 Peer Effects on Duration of Lecture Attendance

In reality, trainees cannot observe their peer presence in the lecture until they enter the online meeting room. In this case, the influence of observing peers is not apparent. To further explore peer effects, we incorporate direct channels about how observing peers' presence online influences one's own behavior. More specifically, after joining a lecture, a trainee can then observe the presence of various peers on the participation list of the instruction platform as shown in Figure 1, and moreover, the participation list refreshes whenever new entry and exit of participants occur. Therefore, the visual effect of observing peers will influence how long to stay in the lecture for all trainees.

Our data contain detailed records of the duration of attendance for every trainee in every lecture. It approaches big data as we observe accurate participation in minutes for each individual-lecture pair. Due to technical limitations, we do not observe at what point of time an individual enters and exits a lecture. Thus, we assume all participants enter a lecture at the beginning. We can then construct the number of peers who appear at the beginning of a lecture by global, local, and close peer groups. Our measures of peer presence use the initial stock of peers rather than the group mean of the contemporaneous duration of attendance, to deal with reflection problems. The causal direction should go from seeing how many peers are present upon entry into a lecture to deciding how long to stay there, not the other way around.

We identify the effect of observing how many peers are initially present in a virtual lecture on i's time spent in lecture l, assuming a variant of Equation (7):

$$y_{il} = \beta_0 + \sum_{g=k,r,s} \beta^g_{peer} N^g_{-il} + d_i + d_l + \varepsilon_{il} \tag{8}$$

where y_{il} measures how many minutes i stays in the lth synchronous lecture. Under the assumption that all trainees enter a lecture at $t = 0$, N^g_{-il} becomes the initial stock of peers who attend the lecture in group Ω^g_{-i}. Thus, β^g_{peer} measures the effect of observing the initial stock of participating peers on i's sequential decision on how long to stay in a lecture. As we cannot guarantee no one was late, the true number of peers who enter at $t = 0$ is $N^{g*}_{-il} = N^g_{-il} - N^g_{-il|t>0}$. The measurement error $N^g_{-il|t>0}$ would bias β^g_{peer} towards zero. We anticipate the attenuation bias is small because there exist video recordings for trainees to watch after lectures, which can help reduce tardy participants. Nevertheless, our estimate of β^g_{peer} can be seen as a lower bound. d_i and d_l maintain the same definitions as in Equation (7).

We estimate Equation (8) in Table 5. First, we apply OLS to the subsample of observations whose duration of lecture attendance is greater than 0. When we replace the demographic variables with individual fixed effects keeping others unchanged, the global peer effect turns negative, and local and close peer effects become significantly positive. If global peers showing up in a lecture increase by 10, one tends to stay for 0.005 minutes shorter. This negative effect is economically insignificant despite its statistical significance at 1% level. An increase of 10 participants from the same county or the same school results in one staying longer for 0.126 and 5.858 minutes respectively. Conditional on participation, global peers generate the smallest peer effects, local peers the second, and close peers the largest. This order is consistent with our prediction based on the intimacy of peer relationships. Like those in Table 4, global peers are substitutes, while local and close peers are complements in trainees' decisions about the time to spend in a lecture.

So far, the OLS results focus on the decisions on the duration of lecture attendance without considering no participation. However, the duration of attendance is often dependent on the whether-to-attend decisions through the correlation between the error terms (Gu et al., 2022). Hence, we need to take no participation into account in the estimations. To deal with the censoring issues due to no participation, we employ the Tobit model to re-estimate Equation (8) on the sample unconditional on participation.

The Tobit results with personal controls on the survey sample show that all three peer groups generate significant positive effects on the predicted own duration of attendance. Generally, individual characteristics have statistically significant effects on the duration of lecture attendance, with signs that are consistent with intuition. For instance, Male trainees stay shorter in a lecture than female ones, in line with the previous founding of gender differences in participation. Bonded by their family duties, trainees who are married or have at least one child spend less time in a lecture than their counterparts, although the differences are statistically insignificant. After accounting for no participation, trainees who graduate from teachers colleges are willing to stay longer in a lecture than their counterparts with no formal teacher training in colleges. Tenured teachers are more active in lecture participation than their untenured peers. The trainees who are more experienced in teaching are more willing to invest their time in the training. In addition, slow internet greatly reduces trainees' duration of attendance. When we control for more time-invariant individual characteristics through fixed effects, the estimated local and close peer effects increase by large while the estimated global peer effect becomes negative. This change in peer effect coefficients is consistent with that in Table 4, implying the original omitted variable bias is negative in the local and close peer effects and positive in the global peer effect.

Furthermore, we run Tobit on the full sample of individual-lecture observations controlling for individual and lecture fixed effects, which is our preferred specification. The results indicate if there is an increase of 10 in the number of global peers showing up in a lecture, a target trainee leaves a lecture 0.022 minutes earlier on average. This negative effect is economically insignificant despite its statistical significance at 1% level. In contrast, an increase of 10 peers from the same county or the same school present in a lecture leads to one's longer stay of 3.742 and 43.359 minutes respectively. The effect sizes in absolute terms are increasing with the closeness of peer relationships, consistent with Equation (5). Additionally, this is explained by the fact that the pressure may become larger as peers turn more intimate, further enlarging their effects on the duration of lecture attendance. The signs of peer effect estimates suggest that global peers generate a larger substitution effect relative to the complementary effect while local and close peers exhibit a larger complementary relative to the substitution effect in the decisions on the duration of lecture attendance. Intuitively, the substitution effect implies that early leave may not draw attention. On the contrary, the complementary effect suggests that a longer stay is prosocial when more peers are present.

Comparing Column (5) with (4), we find that using the survey sample alone tends to underestimate peer effects. Additionally, the Tobit model estimation generates larger peer effect coefficients than OLS. In other words, ignoring the effects of no participation would lead to underestimated peer effects on the duration of lecture attendance.

In Table 6, we estimate the gender subsamples to study peer effect heterogeneity. The literature has documented peer effect heterogeneity by gender (Lavy et al., 2012, Diette and Uwaifo Oyelere, 2014, Griffith and Rask, 2014). The results show that females benefit more from close and local peers and are less affected by global peers than their male counterparts, implying that females receive greater positive peer effects. As the ratio of female teachers is proportionally larger than that of male ones in primary and middle schools, greater peer effects for females may enhance the overall training outcomes by large. For both males and females, the close peer effects are the largest, the local peer effects are the second, and the global peer effects are the least.

5 Working Channels

5.1 Online Social Interactions

The literature has shown that active online social interactions can engage learners, promote participation, and improve performance (Picciano et al., 2002, Davies and Graff, 2005, Lee and Bonk, 2016). Moreover, online interactions may foster peer effects on learning outcomes and persistence (Bettinger et al., 2016). Therefore, we use survey data to explore potential channels that reflect social interactions among peers in the training program.

Firstly, trainees may talk privately with their friends in the program. Friend decisions are influential in making one's own (Card and Giuliano, 2013). These friends are likely to be their school colleagues, generating close peer effects. Secondly, trainees who enroll in the same course can communicate during lecture time in the chat room shown in Figure 1. Besides, there is a chat group on the instruction platform for them to discuss or share course-related matters. The communications and interactions among classmates are traditional channels to produce peer effects (Burke and Sass, 2013, Lu and Anderson, 2015).

In addition, homework may facilitate social interactions. In the YTEP, instructors may assign homework at the end of a lecture. For example, the trainees are asked to use a social network application to post a tweet with a few sentences and pictures to share their thoughts about what they have learned. Once a tweet is posted, all the participants in the same course can give a like or a comment, and the posting trainee can reply to comments. It works like Twitter. Additionally, instructors may ask trainees to write an essay and share it online with the other trainees, who can read it and give a like or a comment. It works like a Blog.

In Table 2, we summarize four variables representing social interactions discussed above: i) the weekly frequency of contacting friends in the training program, ii) the weekly frequency of communicating with peers in the lecture chat room and the course chat group, iii) the total number of likes and comments given and received in the tweet-like homework, iv) the total number of reads, likes and comments received in the blog-like homework. On average, trainees chat with their friends and communicate with their course peers about 4 times a week. For tweet-like posts, an average trainee receives over 100 times of likes or comments in total. In the blog-like homework, they receive about 4 times of reads, likes, or comments, much smaller than that of the tweet-like homework.

To look at the social interaction effects on course participation, we estimate their coefficients after controlling for individual characteristics in Table 7. We find that frequent contact with friends encourages one to actively participate in the online course. For example, an increase of 100 friend communications per week is associated with attending 2 more lectures. Interacting with peers 100 more times through tweet-like and blog-like is correlated with an increase of half-lecture and 3.5-lecture in course participation. However, the course group interactions generate slightly negative effects on participation, although statistically insignificant. The results are robust across different definitions of course participation, either the number of lectures effectively participated or the mean duration of lecture attendance.

The sign and size of estimated social interaction effects help explain the results we find in Section 4, especially the finding that peer effects are increasing with the closeness of peer relationships. The close peer effects result from friend interactions and most of the social media interactions that are normally between familiar peers. Seeing the presence of friends whom one interacts with the most generates the largest close peer effects. The remaining social media interactions may happen among local peers, because trainees may share their local stories or pictures in tweet-like posts and blog-like essays, resonating with peers geographically near them. Therefore, the local peers who share a common social and cultural background produce the second largest peer effects. The course group mostly consists of global peers who exhibit no significant interaction effects. It is generally consistent with our finding of economically insignificant global effects on lecture participation and duration of attendance.

5.2 Incentives by Tenure Status

In addition to social interactions, peer effects can originate from other channels. It is likely that tenured employees' job motivations are much different from their untenured colleagues (Berkowitz et al., 2017). Therefore, tenured teachers may differ from their untenured counterparts in the responsiveness to peer influences. We estimate the model using the subsamples divided by tenure status.[4]

Columns (3) and (4) of Table 6 show that tenured teachers are more responsive to their local peers, while special teachers are influenced by their close and global peers to a larger degree. Intuitively, special teachers from the same school may face competition in getting tenured and thus affect each other deeply. On the contrary, tenured teachers do not have this concern, and thus do not react that much, compared with special teachers, to their school colleagues. By contrasting the estimates of the same peer group, we find that substitution effects from global and local peers are relatively larger for special teachers than tenured ones, leading to a more negative global peer estimate and a smaller positive local peer estimate. Complementary effects are more dominating in school peer effects for special teachers and tenured ones, resulting in a larger close peer estimate. The latter suggests that competition pressure and reputation concerns among school colleagues are greater for special teachers than for tenured teachers.

In summary, we find that the closer the peers are, the larger the social interactions affect participation due to more interactions in the online environment. Furthermore, competition pressure rises when peers are closer. Those channels work to a different degree among various peer groups and thus generate increasing peer effects from global peers to local peers to close peers.

[4]For tenure status, we drop the observations who are substitute or unknown types of teachers, because they have very small sample size to be meaningful for comparison as shown in Table 2. "Special Teacher" refers to a position with an initial 3-year contract without tenure. Tenure can be granted with a high chance at the contract renewal. We only compare peer effect estimates using the subsamples of tenured and special teachers.

6 Conclusion

In this paper, we use a combination of survey data and platform data to study peer effects in online training participation. The unique data structure helps us overcome the challenges in estimating peer effects. More specifically, we are able to separate peer effects into three different levels based on the strength of social connections to investigate not only the peer effects on the attendance but also on the length of stay in a lecture. Moreover, we can mitigate the omitted variable bias using individual and lecture fixed effects. The survey data also allow us to identify working channels of peer influences.

We find some interesting results: i) the close peers (from the same school) have a larger impact than the local peers (from the same county), and both groups of peers have positive effects on the propensity of participating in a lecture and the duration of staying in the lecture; ii) the global peers (outside the county) instead have a negative but economically insignificant effect on participation. The results show that close and local peers generate complementary effects, but global peers generate substitution effects. Our results also indicate that females are influenced more by their peers compared to males, and that job security affects the responsiveness to the behavior of close peers. In addition, social interactions appear to be an important working mechanism for fostering peer effects in online education.

Our findings suggest potential policy implications for improving online education by facilitating various positive peer effects. To generate peer effects, online education platforms may improve visualization of peer presence in the online setting, such as highlighting close peers and local peers. In addition, the design of the online education program can integrate mechanisms that encourage peer interactions like creating various online social networks, to strengthen positive peer effects. Given the low cost of generating peer effects in the online environment, it is beneficial to integrate peer influence mechanisms in any online education program.

References

Alpert, W. T., K. A. Couch, and O. R. Harmon (2016). A randomized assessment of online learning. *American Economic Review 106*(5), 378–82.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics 30*, 98–108.

Berkowitz, D., H. Ma, and S. Nishioka (2017). Recasting the iron rice bowl: The reform of china's state-owned enterprises. *Review of Economics and Statistics 99*(4), 735–747.

Bettinger, E., J. Liu, and S. Loeb (2016). Connections matter: How interactive peers affect students in online college courses. *Journal of Policy Analysis and Management 35*(4), 932–954.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015). Linear social interactions models. *Journal of Political Economy 123*(2), 444–496.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Bramoullé, Y., R. Kranton, and M. D'amours (2014). Strategic interaction and networks. *American Economic Review 104*(3), 898–930.

Brown, B. W. and C. E. Liedholm (2002). Can web courses replace the classroom in principles of microeconomics? *American Economic Review 92*(2), 444–448.

Burke, M. A. and T. R. Sass (2013). Classroom peer effects and student achievement. *Journal of Labor Economics 31*(1), 51–82.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *Review of Economic Studies 76*(4), 1239–1267.

Card, D. and L. Giuliano (2013). Peer effects and multiple equilibria in the risky behavior of friends. *Review of Economics and Statistics 95*(4), 1130–1149.

Carrell, S. E., M. Hoekstra, and E. Kuka (2018). The long-run effects of disruptive peers. *American Economic Review 108*(11), 3377–3415.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica 81*(3), 855–882.

Davies, J. and M. Graff (2005). Performance in e-learning: online participation and student grades. *British Journal of Educational Technology 36*(4), 657–663.

Diette, T. M. and R. Uwaifo Oyelere (2014). Gender and race heterogeneity: The impact of students with limited english on native students' performance. *American Economic Review 104*(5), 412–17.

Figlio, D., M. Rush, and L. Yin (2013). Is it live or is it internet? experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics 31*(4), 763–784.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business and Economic Statistics 31*(3), 253–264.

Griffith, A. L. and K. N. Rask (2014). Peer effects in higher education: A look at heterogeneous impacts. *Economics of Education Review 39*, 65–77.

Gu, X., H. Li, Z. Lin, and X. Tang (2022). Peer effects with sample selection. *Working paper*.

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics 18*(5), 527–544.

Hrastinski, S. (2009). A theory of online learning as online participation. *Computers & Education 52*(1), 78–82.

Jain, T. and M. Kapoor (2015). The impact of study groups and roommates on academic performance. *Review of Economics and Statistics 97*(1), 44–54.

Lavy, V., O. Silva, and F. Weinhardt (2012). The good, the bad, and the average: Evidence on ability peer effects in schools. *Journal of Labor Economics 30*(2), 367–414.

Lazear, E. P. (2001). Educational production. *Quarterly Journal of Economics 116*(3), 777–803.

Lee, J. and C. J. Bonk (2016). Social network analysis of peer relationships and online interactions in a blended class using blogs. *The Internet and Higher Education 28*, 35–44.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics 28*(4), 825–860.

Lu, F. and M. L. Anderson (2015). Peer effects in microenvironments: The benefits of homogeneous classroom groups. *Journal of Labor Economics 33*(1), 91–122.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies 60*(3), 531–542.

Patacchini, E., E. Rainone, and Y. Zenou (2017). Heterogeneous peer effects in education. *Journal of Economic Behavior & Organization 134*, 190–227.

Picciano, A. G. et al. (2002). Beyond student perceptions: Issues of interaction, presence, and performance in an online course. *Journal of Asynchronous Learning Networks 6*(1), 21–40.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *Quarterly Journal of Economics 116*(2), 681–704.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics 85*(1), 9–23.

Tables

Table 1: Summary Statistics of Participation and Peer Measures

	Survey Data		Platform Data			
	Survey Self-report		Survey Matched		Platform Sample	
Variable	Mean	Obs	Mean	Obs	Mean	Obs
Participation Measures						
Individual-Course Level						
Lecture Participation Numbers	12.511	3,655	10.262	3,655	7.842	8,627
Individual-Lecture Level						
Lecture Participation (0/1)	0.782	58,480	0.641	58,480	0.490	138,032
Lecture Participation Minutes	NA	NA	43.482	58,480	31.851	138,032
Peer Measures						
Individual-Course Level						
Global Peers Average Lecture Numbers	NA	NA	7.831	5,351	7.854	8,617
Local Peers Average Lecture Numbers	NA	NA	8.655	5,335	7.899	8,560
Close Peers Average Lecture Numbers	NA	NA	8.889	4,480	7.650	7,350
Individual-Lecture Level						
Global Peers (conditional on participation, # in 10)	NA	NA	451.745	51,435	450.605	67,571
Local Peers (conditional on participation, # in 10)	NA	NA	22.026	51,428	20.444	67,457
Close Peers (conditional on participation, # in 10)	NA	NA	0.601	51,428	0.577	67,457

Notes:

1. In the survey data, respondents are asked to check a list of lectures that they attended. The survey-matched sample only includes those who answered the checklist questions. There are 16 lectures in the Career Development course.
2. The lecture participation is defined as 1 if the duration of lecture attendance is greater than 0.
3. The peer average lecture numbers are calculated by summing up the platform-recorded individual lecture attendance for peers and averaging over all peers in a related peer group. The corresponding peer measures are per lecture.
4. The three peer groups are exclusive, that is, global peers account for all participants per lecture other than those from the same county, and local peers include the participants from the same county but not from the same school.

Table 2: Summary Statistics of Variables – Survey Sample

Variable	Mean	Std	Min	Max	Obs
Individual Characteristics					
Gender (Male)	0.198	0.399	0	1	5,352
Married	0.295	0.456	0	1	5,357
Having Child	0.193	0.395	0	1	5,357
Teachers College	0.695	0.460	0	1	5,324
Years of Experience (yrs)	2.145	3.030	0	37	5,357
Tenured Teacher	0.388	0.487	0	1	5,357
Special Teacher	0.575	0.494	0	1	5,357
Substitute Teacher	0.005	0.071	0	1	5,357
Other Teacher	0.032	0.175	0	1	5,357
Fast Internet	0.236	0.424	0	1	5,357
Normal Internet	0.564	0.496	0	1	5,357
Slow Internet	0.200	0.400	0	1	5,357
Working Channel Variables					
Friend Interactions Per Week (#)	4.111	6.385	0	25	3,655
Course Group Interactions Per Week (#)	4.069	7.815	0	100	4,108
Tweet-like Post Interactions (#)	113.218	142.865	0	1000	4,043
Blog-like Essay Interactions (#)	3.874	13.409	0	100	8,495

Notes:

1. For individual characteristics, the sample combines the respondents in the two waves of survey for a larger sample size.
2. "Friend Interactions per Week" is defined as the average frequency per week one has interactions with their trainee friends.
3. "Course Group Interactions per Week" is defined as the average frequency per week one asks and answers questions in the lecture chatroom and the course chatgroup.
4. "Tweet-like Post Interactions" is defined as the total number of likes and comments one gives and receives on tweet-like homework posts.
5. "Blog-like Essay Interactions" is defined as the total number of reads, likes, and comments one receives on blog-like homework essays.
6. The first three working channel variables are self-reported by survey respondents in only one wave where the related questions were asked. The last working channel variable is retrieved from the platform database.
7. Variables with no designated units are dummies.

Table 3: Peer Effects on Number of Lectures Participated

	(1)	(2)	(3)
	y_i: Number of Lectures Participated		
	Survey Sample (>0 minutes) OLS	Survey Sample (>0 minutes) Tobit	Survey Sample (≥30%) Tobit
Global Peers Average	0.242^{***}	0.237^{***}	0.251^{***}
Lecture Numbers	(0.023)	(0.023)	(0.026)
Local Peers Average	0.314^{***}	0.310^{***}	0.292^{***}
Lecture Numbers	(0.026)	(0.026)	(0.027)
Close Peers Average	0.545^{***}	0.553^{***}	0.565^{***}
Lecture Numbers	(0.021)	(0.021)	(0.021)
Gender (Male)	-0.896^{***}	-0.904^{***}	-0.974^{***}
	(0.116)	(0.118)	(0.123)
Married	0.022	0.024	-0.041
	(0.146)	(0.149)	(0.159)
Having Child	-0.078	-0.076	-0.032
	(0.173)	(0.176)	(0.187)
Teachers College	0.143	0.140	0.176
	(0.094)	(0.095)	(0.100)
Tenured Teacher	-0.041	-0.039	0.017
	(0.092)	(0.094)	(0.098)
Years of Experience	0.096^{***}	0.096^{***}	0.089^{***}
	(0.021)	(0.021)	(0.021)
Slow Internet	-0.086	-0.087	-0.182
	(0.108)	(0.110)	(0.116)
Observations	4,439	4,439	4,439

Notes:

1. The dependent variable y_i is the number of lectures participated by trainee i in the career development course. In Columns (1) and (2), a trainee is considered a participant in a lecture if the duration of lecture attendance is greater than 0. In Column (3), a trainee is considered a participant if the duration of lecture attendance is at least 30% of the lecture length, otherwise is treated as a nonparticipant.
2. Peer measures are constructed using the two participation criteria, > 0 minutes in Columns (1) - (2) and ≥30% in Column (3), to calculate the average number of lectures participated by i's peers in a related peer group based on the complete network in the platform data.
3. Column (1) is estimated by OLS. Columns (2) and (3) are estimated by Tobit to account for no participation (zero number of lectures participated). The Tobit coefficients are reported.
4. Robust standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Table 4: Peer Effects on Participating in Lectures

	(1)	(2)	(3)	(4)
	D_{il}: Individual-Lecture Participation (0/1)			
	Survey Sample	Survey Sample	Survey Sample	Platform Sample
	OLS	Probit	Probit	Probit
Participating Global Peers (in 10)	0.013***	0.001***	-0.003***	-0.002***
	(0.000)	(0.000)	(0.001)	(0.000)
Participating Local Peers (in 10)	0.027***	0.003***	0.020***	0.023***
	(0.001)	(0.000)	(0.001)	(0.001)
Participating Close Peers (in 10)	0.510***	0.079***	0.263***	0.313***
	(0.018)	(0.004)	(0.009)	(0.007)
Gender (Male)	-0.085***	-0.122***		
	(0.004)	(0.004)		
Married	0.001	-0.053***		
	(0.005)	(0.005)		
Having Child	-0.004	0.014*		
	(0.006)	(0.006)		
Teachers College	0.036***	-0.078***		
	(0.003)	(0.003)		
Tenured Teacher	-0.053***	-0.096***		
	(0.003)	(0.003)		
Years of Experience	0.004***	-0.001		
	(0.001)	(0.001)		
Slow Internet	-0.032***	-0.083***		
	(0.004)	(0.004)		
Individual FE	N	N	Y	Y
Lecture FE	Y	Y	Y	Y
Observations	84,576	84,576	78,960	122,080

Notes:

1. The dependent variable D_{il} is a dummy variable equal to 1 if trainee i's duration of attendance in lecture l is greater than 0, and 0 otherwise.
2. Peer measures are constructed using the number of i's peers whose duration of lecture attendance is greater than 0 in a related peer group based on the complete network in the platform data.
3. The individual-lecture participation data are used for estimation. Columns (1) - (3) use the sample of survey respondents. Column (4) uses the sample of all trainees regardless of answering the survey or not.
4. Column (1) is estimated by OLS. Columns (2) - (4) are estimated by Probit and the average marginal effects are reported.
5. Robust standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Table 5: Peer Effects on Duration of Lecture Attendance

	(1)	(2)	(3)	(4)	(5)
	y_{il}: Duration of Lecture Attendance in Minutes				
	Sample Observations ($i\|y_{il} > 0$)		Sample Observations ($i\|y_{il} \geq 0$)		
	Survey	Platform	Survey	Survey	Platform
	OLS	OLS	Tobit	Tobit	Tobit
Participating Global Peers (in 10)	0.002	-0.005**	0.156***	-0.021***	-0.022***
	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)
Participating Local Peers (in 10)	0.003	0.126***	0.305***	3.405***	3.742***
	(0.004)	(0.031)	(0.007)	(0.002)	(0.002)
Participating Close Peers (in 10)	-0.470***	5.858***	5.159***	35.361***	43.359***
	(0.090)	(0.523)	(0.196)	(0.058)	(0.058)
Gender (Male)	-2.435***		-11.700***		
	(0.238)		(0.495)		
Married	-0.507		-0.117		
	(0.283)		(0.598)		
Having Child	-0.181		-0.533		
	(0.337)		(0.707)		
Teachers College	-0.094		4.745***		
	(0.194)		(0.411)		
Tenured Teacher	-0.094		-7.177***		
	(0.187)		(0.399)		
Years of Experience	-0.071*		0.525***		
	(0.035)		(0.074)		
Slow Internet	-1.730***		-4.944***		
	(0.229)		(0.478)		
Individual FE	N	Y	N	Y	Y
Lecture FE	Y	Y	Y	Y	Y
Observations	50,896	67,457	84,576	85,600	137,328

Notes:

1. The dependent variable y_{il} is trainee i's duration of attendance in minutes in lecture l.
2. Peer measures are constructed using the number of i's peers whose duration of lecture attendance is greater than 0 in a related peer group based on the complete network in the platform data.
3. Columns (1) and (2) use the individual-lecture data on the sample of trainees whose duration of lecture attendance is greater than 0. Columns (3) - (5) use the individual-lecture data on the sample of trainees unconditional on lecture participation. Among them, Columns (3) and (4) use the sample of survey respondents. Column (5) uses the sample of all trainees.
4. Columns (1) and (2) are estimated by OLS. Columns (3) - (5) are estimated by Tobit to account for no participation in a lecture and the Tobit coefficients are reported.
5. Robust standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Table 6: Heterogeneous Peer Effects on Duration of Lecture Attendance

	(1)	(2)	(3)	(4)
	y_{il}:Duration of Lecture Attendance in Minutes			
	Male	Female	Tenured Teacher	Special Teacher
Participating Global Peers (in 100)	-0.028***	-0.018***	-0.006***	-0.023***
	(0.000)	(0.000)	(0.000)	(0.000)
Participating Local Peers (in 100)	3.202***	3.465***	4.149***	2.600***
	(0.005)	(0.002)	(0.005)	(0.002)
Participating Close Peers (in 100)	28.537***	39.012***	27.916***	40.843***
	(0.106)	(0.072)	(0.112)	(0.068)
Individual FE	Y	Y	Y	Y
Lecture FE	Y	Y	Y	Y
Observations	16,832	68,192	33,248	49,200

Notes:

1. The dependent variable y_{il} is trainee i's duration of attendance in minutes in lecture l.
2. Peer measures are constructed using the number of i's peers whose duration of lecture attendance is greater than 0 in a related peer group based on the complete network in the platform data.
3. The subsamples of males and females are used in (1) and (2) respectively. The subsamples of tenured and special teachers are used in (3) and (4) respectively.
4. The models are estimated by Tobit to account for no participation in a lecture and the Tobit coefficients are reported.
5. Robust standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Table 7: Social Interaction Channels on Course Participation–Survey Sample

	(1) y_i :Number of Lectures Participated (≥30%)	(2) y_i : Mean Duration of Lecture Attendance
Friend Interactions	0.021*	0.064*
per Week	(0.010)	(0.031)
Course Group	-0.003	-0.002
Interactions per Week	(0.008)	(0.022)
Tweet-like Post	0.006***	0.010***
Interactions	(0.000)	(0.001)
Blog-like Essay	0.035***	0.113***
Interactions	(0.003)	(0.008)

Notes:

1. In Column (1), the dependent variable y_i is the number of lectures participated by trainee i in the career development. A trainee is considered a participant in a lecture if the duration of lecture attendance is at least 30% of the lecture length, otherwise is treated as a nonparticipant. In Column (2), the dependent variable y_i is trainee i's average duration of attendance over the 16 lectures.
2. The estimate in each cell is obtained from a separate regression of the dependent variable y_i on the corresponding social interactions with the same personal covariates as in Table 3. Column (1) is estimated by Tobit on the sample of survey respondents. Column (2) is estimated by OLS on the sample of survey respondents.
3. Robust standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figures

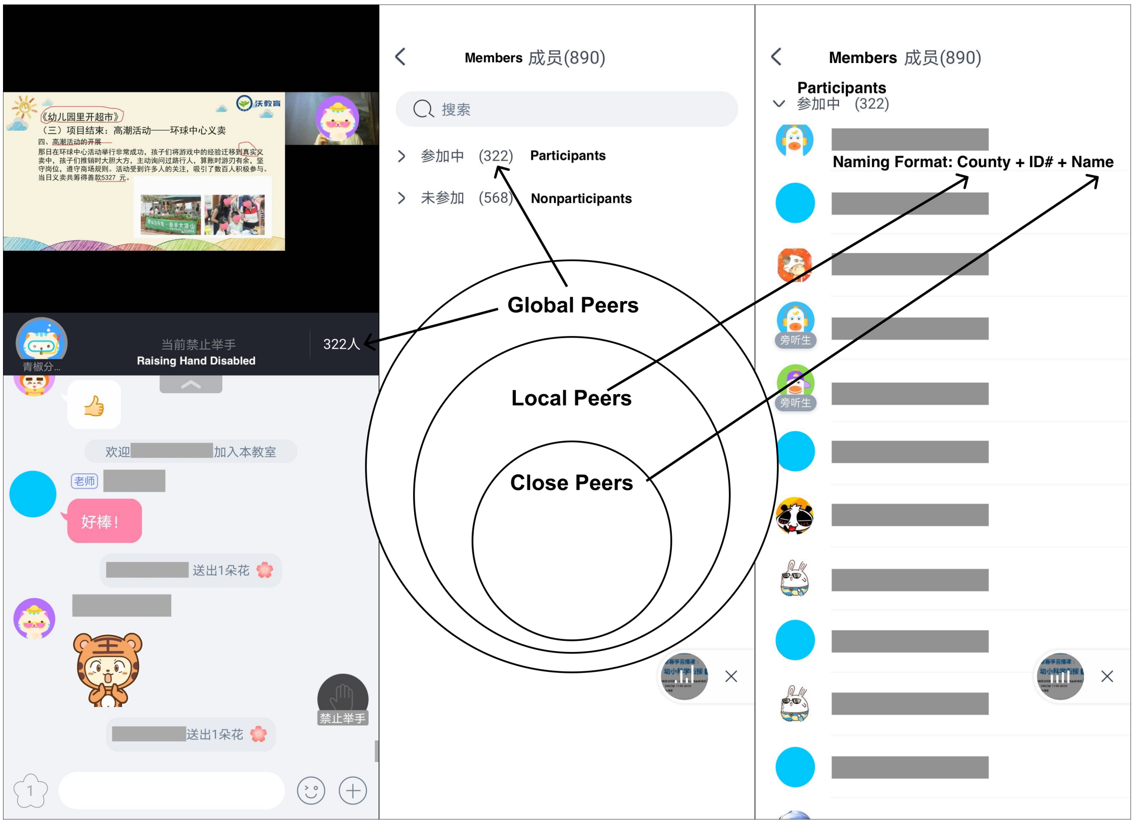

Figure 1: Platform Interface of Synchronous Lectures

Note:

1. The upper window of the left column is the live stream of the lecture presentation by the instructor.
2. Below that window, it is a chatroom where participants can communicate with the instructor, TAs, and other trainees by typing words and sending emojis.
3. Trainees can observe the list of participants and nonparticipants shown in the middle column. By clicking the list, one can see who is participating and who is not. The list of participants is refreshed wherever there is an entry or exit.
4. Each trainee has a unique identifier whose naming format is "*County + ID# + Name*". The list is sorted by the characters of the identifiers, therefore, the participants from the same county are placed adjacently. Given this feature, one can observe whether a peer is from the same county or the same school, and how many of them are present in the lecture.

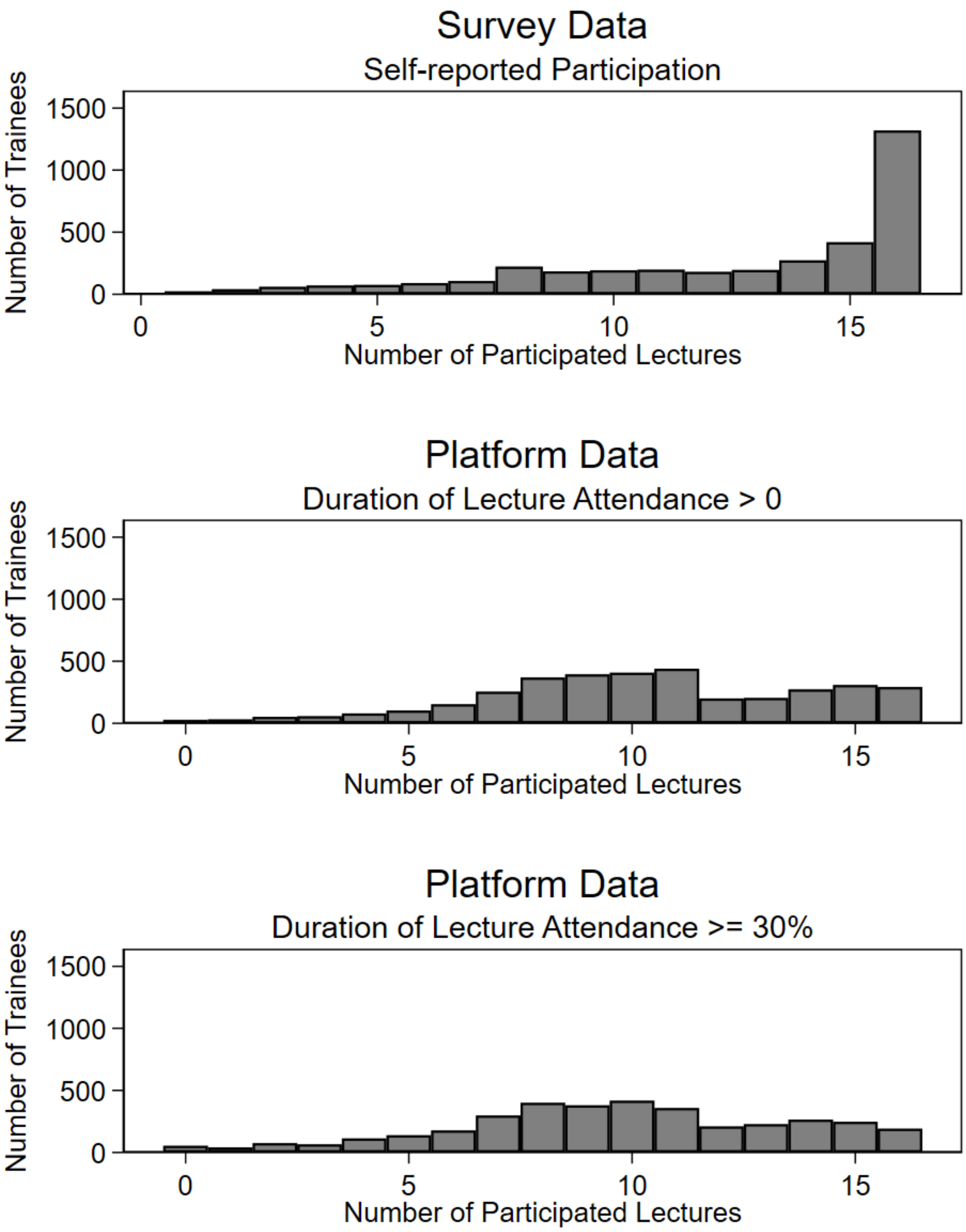

Figure 2: Comparison between Survey Data and Platform Data

Note:

1. The distributions include 3,655 trainees who self-report their lecture participation in the survey.
2. The number of lectures checked by survey respondents measures the self-reported participation in the survey data.
3. Their actual duration of attendance in lectures can be identified in the platform data. We calculate the accurate number of lectures participated based on the two criteria (duration >0 minutes and duration ≥30% of the lecture length) to define the participation in the platform data.

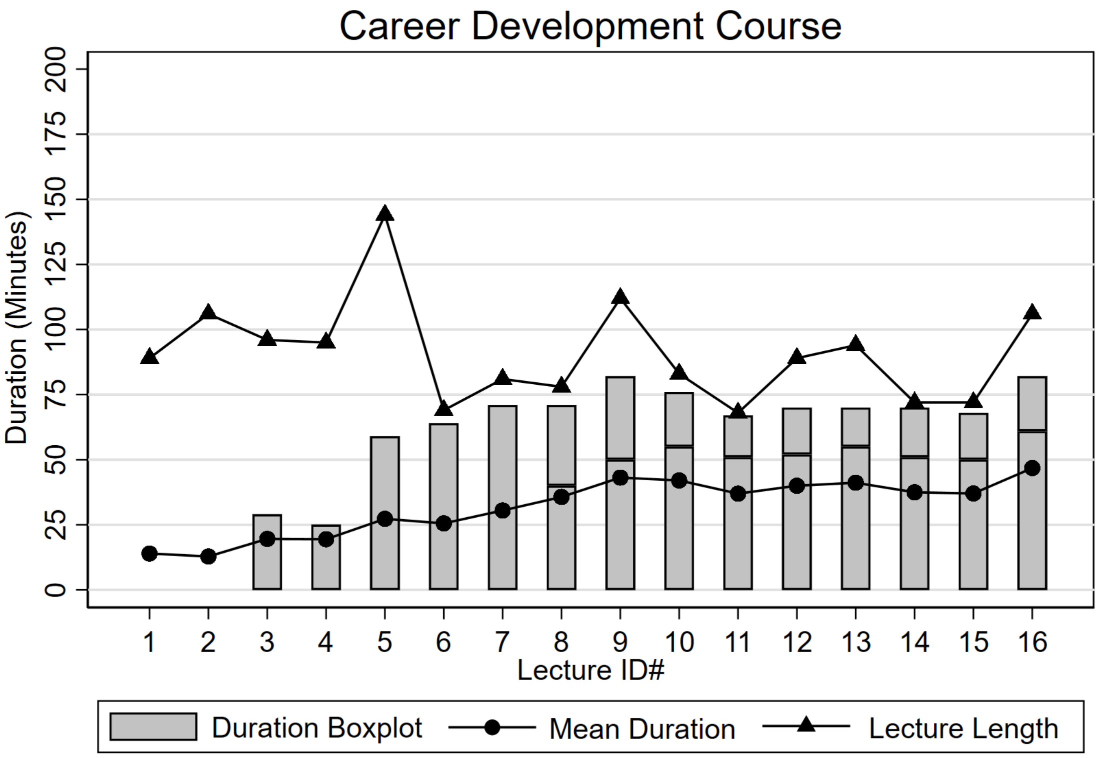

Figure 3: Duration of Lecture Attendance in Career Development Course

Note:

1. The figure is plotted based on 8,627 trainees' duration of attendance in 16 lectures.
2. The top line (solid triangle) displays the length by lecture. On average, a synchronous lecture lasts for about 95 minutes.
3. The lower line (solid dot) plots the mean duration of attendance by lecture.
4. The box displays the 25th, 50th, and 75th percentiles of the duration of attendance for each lecture. For example, the box of lecture #9 shows that the 25th percentile of the duration of attendance is 0, the 50th percentile is 50 minutes, and the 75th percentile is 82 minutes.
5. If the median line is missing in a box, it means that the 50th percentile of participation length is 0. For example, the box of lecture #7 has no median line in the box, which means the 50th percentile of the participation distribution in that lecture is 0.
6. Over 75% of the 8,627 trainees did not show up in the first two lectures, which degenerates the first two boxes. It does not mean that there is no variation in the duration of attendance in these two lectures. For example, the 90th percentile is 89 minutes in lecture #1 and 74 minutes in lecture #2. After then, the attendance rate gradually rises and stays steady till lecture #16.