

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Friess, Svenja; Rosendahl Huber, Laura

Conference Paper Breaking The Ice: Can Initially Active Peers Improve Platform Engagement And Persistence?

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2024: Upcoming Labor Market Challenges

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Friess, Svenja; Rosendahl Huber, Laura (2024) : Breaking The Ice: Can Initially Active Peers Improve Platform Engagement And Persistence?, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2024: Upcoming Labor Market Challenges, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at: https://hdl.handle.net/10419/302443

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

Breaking The Ice: Can Initially Active Peers Improve Platform Engagement

And Persistence?*

SVENJA FRIESS

Max Planck Institute for Innovation and Competition Department for Innovation and Entrepreneurship Research Marstallplatz 1 Munich, Germany, 80539 *e-mail: svenja.friess@ip.mpg.de*

LAURA ROSENDAHL HUBER

Erasmus University Rotterdam Rotterdam School of Management Burgemeester Oudlaan 50 Rotterdam, Netherlands, 3062 PA *e-mail: rosendahlhuber@rsm.nl*

Do not circulate without the authors' permission.

Last update: March 1, 2024

^{*}For valuable comments and suggestions, we thank Alessandra Allocca, Florian Englmaier, Dietmar Harhoff, Tobias Kretschmer, Nadzeya Laurentsyeva, Maren Mickeler, Timm Opitz, Helene Stradt, Rainer Widmann, and participants of the seminars at LMU Munich and the Max Planck Institute for Innovation and Competition. We thank the providers for data access and generous support in understanding the platform. Besides, we thank Haylee Ham and Yilun Xu for their invaluable support in extracting and processing the platform data. All remaining errors are our own.

Breaking The Ice: Can Initially Active Peers Improve Platform Engagement And Persistence?

ABSTRACT

In recent years, online knowledge exchange has flourished, yet struggles with low user engagement. This study investigates the role of early peer interactions in sustaining engagement on digital platforms. Analyzing novel data from 12,000+ professionals upskilling in an online business program, we exploit quasi-random variations in initial peer activity levels per cohort to estimate their impact on future engagement and platform persistence. Results reveal that a high initial share of active peers giving likes reduces platform persistence by 3%, influencing later commenting. Early active commenting peers exhibit a modest positive correlation with future engagement, while receiving early comments and likes significantly boosts future engagement and platform persistence. Employing cutting-edge Natural Language Processing techniques, we classify comment characteristics to shed light on underlying mechanisms. Our findings provide insights for digital platform designers to leverage early and directed peer interactions effectively, enhancing user experience and platform value.

Last update: February 29, 2024

Keywords: user engagement, digital platforms, peers, communication, upskilling, NLP

Breaking The Ice: Can Initially Active Peers Improve Platform Engagement And Persistence?

Digital advancements have revolutionized the exchange of knowledge. Ubiquitous online platforms and communities have expanded the reach of information dissemination and knowledge exchange by overcoming traditional constraints of time and location (Faraj, Krogh, et al. 2016; Rietveld and Schilling 2021). Besides, organizations increasingly rely on digital knowledge management tools to facilitate efficient information flows among members. Platforms like "Slack", "Microsoft SharePoint", or "Starmind" exemplify this trend. This type of information exchange is particularly important in the knowledge economy, where complex tasks prevail (Autor, Levy, and Murnane 2003).

A special type of knowledge exchange platform - online education platforms - has grown rapidly in recent decades. Accelerated by the COVID-19 pandemic, they have become an established pillar at all levels of (global) higher education (Gong, T. X. Liu, and Tang 2021; Li and Lalani 2020), offering flexibility, cost-effectiveness, and exposure to diverse insights from global peers.

In the wake of this development, major multinational corporations such as JP Morgan, Amazon, and Google have launched platform-based training to *upskill* their workforce with the latest business skills and tools (Bidwell, Briscoe, et al. 2013; Cappelli 2015; Deming and Noray 2018). As continuous human capital accumulation is essential to firm growth and innovation (Dragoni et al. 2009) and technical skills become outdated at an increasing pace (Deming and Noray 2018), firms strategically seek to upskill their global labor force to remain competitive (Illanes et al. 2018; Tamayo et al. 2023). For this purpose, digital training provides a scalable and cost-efficient approach.

All these knowledge exchange platforms share common interaction features to generate content, and to source and share information among users, like posting statements or questions and responding through comments and reactions, e.g., "likes". The underlying premise of these interactive elements is that shared knowledge and feedback from peers is expected to be meaningful (D. J. Zhang, Allon, and Van Mieghem 2017; Wang, X. Zhang, and Hann 2018). However, despite their prevalence and potential, most knowledge exchange platforms face critical challenges of sustaining user engagement and declining response rates over time (Ren, Harper, et al. 2012; Gallus

and Frey 2016; Baek and Shore 2020; Mickeler et al. 2023). Thus, it remains an ongoing effort to understand barriers to, and facilitators of, efficient online knowledge exchange and how digital interactions can be designed and governed to create value.

Until now, most of the literature on online knowledge exchange platforms has focused on interactions surrounding individuals' *own* questions and responding *behavior*, e.g., on determinants of knowledge sharing (Hwang, Singh, and Argote 2015; Wasko and Faraj 2005; Xu, Nian, and Cabral 2020; Faraj, Krogh, et al. 2016) and knowledge seeking (Mickeler et al. 2023; Borgatti and Cross 2003). Only Bettinger, J. Liu, and Loeb (2016) have begun to shed light on the impact of *peers' activities* on a user's outcomes in a virtual college education context. They find suggestive evidence that being directly nominated, i.e., referred to by one's peers improves course success for users who are relatively less interactive themselves. Beyond that, little is known about which activities and how the activities of one's peers may facilitate engagement and persistence on online platforms.

While a large literature in economics has shown that peers can have a positive impact in conventional education settings, i.e., a physical classroom (Sacerdote 2001; Zimmerman 2003; Calvó-Armengol, Patacchini, and Zenou 2009; Ammermueller and Pischke 2009; Golsteyn, Non, and Zölitz 2021; Feld and Zölitz 2022; Feld and Zölitz 2017). This evidence stems, however, from offline contexts where the precise mechanisms of how and under which conditions interactions with active peers are beneficial for engagement, persistence, and knowledge exchange remain unobserved.

This paper aims to fill this gap. We capture the notion of peers' activities and apply that to the interaction behavior online asking: *How does the presence of initially active peers impact users' future engagement and persistence on a digital platform?* By studying the timing, type, and content of online peer interactions, we contribute an in-depth understanding of how these factors influence users' subsequent platform behaviors.

We conducted a comprehensive study using a novel data set from an elite U.S. business school, encompassing over 12,000 working professionals enrolled in online business skills training courses. We capture 36 different cohorts over a five-year period from April 2017 to June 2021. We

exploit quasi-random variation in the initial share of active peers per cohort to estimate the impact of having a higher share of active peers and users' social engagement on users' future engagement and persistence.

Our findings indicate that being surrounded by an initial high share of active peers does not lead to more engagement or persistence on the platform. That is, we find that a higher share of peers giving likes significantly decreases platform persistence by about 3% on average and also slightly diminishes later commenting probability, while early actively commenting peers do not influence the likelihood of giving comments in the future or platform persistence. However, we also find that receiving directed comments and likes early on correlates with a robust ten-plus percentage point increase in future engagement, with the strongest impact observed for the same type of interaction. Additionally, early comments and likes received are associated with a 7% and 9% higher platform persistence, respectively.

By utilizing state-of-the-art Natural Language Processing (NLP) techniques, we developed a novel comment classification matrix to examine mechanisms. We capture the share of received (dis-)agreeing and (non-)elaborating comments from peers. Overall, the results of our study indicate that users engage with one another in diverse ways, displaying considerable heterogeneity in their communication styles. Specifically, elaborating and purely agreeing comments exhibit statistically significant positive correlations with users' future engagement levels and platform persistence. Notably, the combined presence of 'agreement & elaboration'-comments exhibit the most robust, positive association with all studied behavioral outcomes.

In sum, peers matter online - cohorts with active engagement ignite more interactions, yet receiving comments early is key. Designers of digital (learning) platforms are advised to better harvest social engagement to enhance their user experience and value generated by their platforms.

Our research allows us to contribute to several research domains. We contribute to the scholarship on user-generated content platforms (Kretschmer et al. 2022; Loh 2022; Claussen, Kretschmer, and Mayrhofer 2013; D. J. Zhang, Allon, and Van Mieghem 2017; Wasko and Faraj 2005; Bettinger, J. Liu, and Loeb 2016) by studying the timing, type, and content of early peer

interactions to gain an in-depth understanding of how these influence users' subsequent platform behaviors.

We also contribute to the literature on peer interactions (e.g., Sacerdote (2001), Ammermueller and Pischke (2009), Feld and Zölitz (2017), Feld and Zölitz (2022), and Bostwick and Weinberg (2022)) adding evidence from an online setting where our detailed interaction data and cohort setting allows us to measure the impact of early peer interactions on follow on users' behaviors and interactions over a longer period.

Our paper has implications for the scholarship on labor force training (Bidwell and Briscoe 2010; Cappelli 2015) due to our novel focus on a diverse, global sample of working professionals and executives in this context. To the best of our knowledge, this paper is among the first to advance the understanding of peer effects among working professionals in a virtual learning environment. Our focus on MBA training allows us to assess platform behaviors in a highly labor market-relevant setting.

Lastly, we add to a recent stream of works in social science using text data to generate novel measures for empirical analyses (Athey and Imbens 2019; Gentzkow, Kelly, and Taddy 2019)

The remainder of this paper is structured as follows. In Section 2, we shed light on the theoretical background and formulate our hypotheses. Section 3 describes the empirical setting, strategy, and data. Section 4 presents the results. Finally, Section 5 offers a discussion and conclusive remarks.

THEORY

The Role Of Early Peer Interactions For Future Engagement And Platform Persistence

Interaction and knowledge exchange are the core purposes of many online communities (Jones, Ravid, and Rafaeli 2004). Similarly, most online education platforms possess intentional features such as internal discussion boards for the same purpose. These discussion boards include familiar communication modes, such as posts, comments, and likes intended to encourage or even impose mandatory public posts or other forms of interaction among users. The underlying premise

of such built-in interaction features is the information shared with peers, and feedback given is meaningful (D. J. Zhang, Allon, and Van Mieghem 2017; Wang, X. Zhang, and Hann 2018). However, it is an open question whether and how early peer activities on knowledge exchange platforms can benefit individual users or provide a public (knowledge) good.

This paper focuses on how peer interactions *early* in new users' life cycles on an online knowledge exchange platform impact their future engagement behaviors and persistence on the platform. Specifically, we distinguish the following types of feasible and most common peer activities in the *initial* phase of users' platform life cycle:

- 1. *General Peer Activity:* Exposure to undirected peer activities, i.e., observing others' commenting and liking activities.
- 2. *Directed Peer Activity:* Receiving comments or likes from one's peers on one's posts and comments.

Starting with *General Peer Activity*, information overload is a common challenge in large online forums (Makos et al. 2013; Jones, Ravid, and Rafaeli 2004). This issue is particularly pertinent in settings where posting is mandatory or where online communities comprise a substantial number of (active) users. The resulting abundance of content makes it challenging for users to discern valuable insights from repetitive or low-quality contributions. Additionally, in many settings, users are professionals with full- or part-time employment, limiting their time on any platform. As such, the critical question of where users should focus their scarce attention arises. We propose that the activity, particularly *likes* endorsing posts or comments, could help identify promising content. Users may give likes to content they consider to be of high quality, thereby directing attention toward valuable contributions (Makos et al. 2013). Consequently, seeing that many peers are active early on and provide valuable content as well as like endorsements can motivate users to interact and contribute content to the platform in the future. Such motivators could be exchanging ideas, soliciting feedback, or aiming at "collecting likes" for one's posted content (Baek and Shore 2020). Hence, we hypothesize:

Hypothesis 1a. A high share of peers who initially distribute likes positively affects future engagement.

Similarly, observing that many peers distribute likes or comments in the initial phase on the platform could establish a social norm of high engagement for future periods and motivate participants to stay on the platform. We hypothesize:

Hypothesis 1b. A high share of peers who are initially active (with respect to distributing likes and comments) positively affects platform persistence.

Moving from the "passive" consumption of peers' activity on the platform to the peer activity *directed* to a user, they may *receive likes* and *comments*. We postulate that receiving likes from peers can boost motivation to continue. It may positively impact persistence on the platform through the channel of a sense of appreciation or reassurance in one's posts:

Hypothesis 2. Receiving like(s) in the initial phase of the platform life cycle positively affects platform persistence.

Besides, it has been established in the literature that reciprocity can be a key motivator for user contributions online (Ardichvili, Page, and Wentling 2003; Wasko and Faraj 2005). For instance, Chen, Baird, and Straub (2019) find that "positive votes" by peers may have a motivational effect on future engagement, in particular on knowledge contributions, using panel data from a technical Q&A-forum. Hence, receiving likes on one's posts early in a user's life cycle on a platform may create a positive feedback loop encouraging their future engagement. Specifically, hypothesize:

Hypothesis 3a. Receiving like(s) in the initial phase of the platform life cycle positively affects users' future engagement.

Moving beyond likes, we consider the role of receiving comments from peers early on. There are several facets to receiving comments. First, similar to receiving a like, a received comment can trigger a direct reaction or impact future engagement positively via the specific or general reciprocity channel (Surma 2016). Previous research has used data from an open-source software development community to analyze the role of community responses and member roles on platform persistence and found a positive impact of community member responses particularly on "users'" (i.e., members adopting software rather than modifying it) continued participation (C. Zhang, Hahn, and De 2013). Hence,

Hypothesis 3b. Receiving comments(s) in the initial phase of the platform life cycle positively affects giving comments in the future.

To delve deeper into the mechanisms driving the positive impact of peer activity on prospective engagement and platform persistence, we consider various types of comments as a moderating channel.

Comment Type As Moderator Of The Impact Of Receiving Comments On Future Platform Behavior

Taking an even closer look at the type of comments users provide, we distinguish the following types: Comments might either *agree* or *disagree* with an original post or previous comments. If users intend to solely state their (dis-)agreement with peers' content, they provide a *valuation*. Conversely, users may wish to provide a valuation of their peers' content and *elaborate* on why they (dis-)agree. First, considering comments that provide a *valuation* are either *purely agreeing* or not. Such purely agreeing comments in the initial phase on the platform may provide positive feedback and reassurance to the recipient. This early positive experience can have a motivating effect to reciprocate by engaging with peers' content in the future and persist longer on the platform. Using Facebook brand page data, Khobzi, Lau, and Cheung (2019) find that the sentiment of messages matters for users' engagement behavior (with brand posts). They find that more positively and negatively framed comments trigger user engagement. In our setting, positive comments correspond to "agreement". Hence, we hypothesize:

Hypothesis 4a. Receiving purely agreeing comments in the initial phase of the platform life cycle positively affects future engagement.

Hypothesis 4b. Receiving purely agreeing comments in the initial phase of the platform life cycle positively affects platform persistence.

Second, users may wish to not only provide a valuation of their peers' content but also *elaborate* on why they (dis-)agree. Looking at these longer, *elaborating comments*, might be particularly insightful as users might share new insights or explanations. We distinguish between *elaborating and agreeing* and *other* comments. Note that the latter category contains all comments that could not fit in the aforementioned three categories. These are, however, not "purely disagreeing"

in nature as such comments have not been found much on the platform. Previous work by Ziegele et al. (2018) investigates factors influencing the impact of the civility of comments on news websites on readers' willingness to comment by experimentally manipulating comment nature. Their findings suggest that the type of comments can dynamically shape user engagement in online discussions. Particularly, "deliberative" reader comments containing the discussion features such as questions or additional information (i.e., would be classified as "elaborating" in our framework) increased participants' willingness to reply to these comments primarily via cognitive involvement. In a similar vein, Kwon et al. (2019) studies the effects of different types of instructor comments on engagement in an online discussion in an educational setting. Their descriptive content analysis suggests that elaborating comments is positively associated with interactivity among learners. Hence, we hypothesize:

Hypothesis 5a. Receiving elaborating and agreeing comments in the initial phase of the platform life cycle positively affects platform persistence.

Hypothesis 5b. Receiving elaborating and agreeing comments in the initial phase of the platform life cycle positively affects future commenting behavior.

DATA AND EMPIRICAL STRATEGY

The empirical context of this study comprises the course progression and social interactions of a large, diverse population of working professionals of an elite U.S. business school's upskill training courses. We capture cohorts over a five-year period from April 2017 to June 2021, with most course offerings once per quarter. Unlike Massive Open and Online Courses (or MOOCs), these online business courses require prospective users to apply to the program, get accepted, and pay non-trivial tuition to enroll alongside a cohort of virtual peers.

The Empirical Setting

Platform, Course & Cohort Structure In particular, we assess the business school's online flagship program that consists of three business and economics courses. These three courses are taken simultaneously over 10-12 weeks, starting staggered one after another. This program's tuition

is above USD 2.000, and approximately one-third of enrolled users receive at least some corporate tuition reimbursement. Through our interviews with the program administrators, past users have indicated their motivations for taking the online courses, including having helped to bolster their CVs, improve their job performance, and enable them to join a network of like-minded peers.

The learning model for each course is designed around three core components, which include video lectures, case-based learning where users discuss and debate solutions to real-world business cases with their cohort peers, as well as social learning, where cohort members exchange ideas, offer input, seek out different viewpoints, and learn from one another's experiences and perspectives. This paper focuses primarily on the latter element of the courses. Upon course completion, users receive an online certificate from the business school. To earn a certificate of completion, users must complete each week's lessons by the weekly deadlines and earn an average quiz score at the end of each lesson of at least 50%.

This program is particularly suited to test our hypotheses for the following reasons. First, it is much more demanding than the other online training offered by the business school or comparable MOOCs. Because it consists of three courses that are taken in parallel over a 10-12 week period, the weekly time investment for the flagship program is estimated at 12-15 hours, as compared to individual courses that require the work of around 5-8 hours per week and for a shorter duration (6-8 weeks). To complete the course, users must take a three-hour final exam in person at one of the partnering testing facilities around the globe. These features ensure users have high incentives to engage with the course content and peers to learn and complete the program, including an official certificate.

Moreover, active participation in the social course elements is an explicit learning requirement within this program. It is incentivized such that they contribute to passing the program or improve the final grade, conditional on having passed the program. Hence, due to its intensity and the required active participation during the program, social interactions are an integral part of this program experience and, thus, a uniquely favorable setting to confirm our hypotheses.

Besides, interviews with the program administrators suggest that they aim to keep the cohorts

in a target size of 400 users (Mean = 384, SD = 73.6, Median = 376, N = 36 cohorts) to ensure a critical mass of users on the platform for interaction purposes. A cohort size of several hundred users is robust to time zone and work habit differences. It limits the likelihood that a user will not have any peers engaging with their contributions and ideas on the platform without becoming too large either. Due to the popularity of the flagship program, many course waves were larger than this target size. If a course offering has more than 600 users, the users are split into two separate cohorts. For the affected program waves, the administrators block randomized users into two cohorts according to users' country of residence if there were at least five users from the same country. Otherwise, users from the same country are grouped in the same cohort.

Users can only enroll in the course if they have passed the program's admission criteria, and about 50% of admitted users choose to enroll. Upon enrollment, users are assigned a cohort of peers who simultaneously start the course.

Our study context is uniquely suited to examine peer activity in online engagement for two reasons. First, our setting circumvents concerns of selection and common shocks, which are two issues possibly creating identification problems in previous peer effects studies (Manski 1993). Course syllabi are publicly posted, and video lectures are pre-recorded. These features remove heterogeneity in the delivery of course content that may arise due to changing instructors or spontaneous rearrangement of modules. More importantly, among the set of enrolled users, they apply independently and cannot select which (virtual) cohort "classroom" they are in, as cohorts are constructed based on the individuals who decided to apply to the same offering of the course, i.e., for the same "course wave" (Rosendahl Huber, N Lane, and Lakhani 2020).

Second, there are significant opportunities for peer-to-peer interaction, as the course content is structured around interactive learning via case studies and weekly discussions, where users are encouraged to discuss case and homework prompts with their cohort peers. The opportunities for peer-to-peer interaction are particularly useful for working professionals due to the knowledge and skills that experienced users bring to the course (Littlejohn et al. 2016).

Peer Interactions on the Platform While the courses are asynchronous in general, i.e., users are able to log in whenever they want—there is a certain degree of synchrony imposed by common (bi-)weekly deadlines and locked content. Each course module is unlocked at a fixed time, determining when users can start working on a module. Within a module, some content is locked/not visible until the user completes the previous content, meaning that one can look back at previous work but cannot skip ahead. There is also a shared module deadline by which all work in the module must be completed. Most modules are available for two weeks. This ensures that all users in a cohort work through the asynchronous material for a module in the same two to three-week period.

In addition to completing the quiz for each module, users are also expected to periodically post and respond to "shared reflections" on a common message board. These reflections typically ask users to reflect on some questions about the course material and offer their thoughts and opinions on the discussion board. Other users can then engage with the reflection by "liking" it or responding to it with their thoughts. In some cases, this results in a back-and-forth discussion between the original poster and other users within the comments of a reflection. We refer to this combination of a reflection and its associated comments as a thread. Both commenting and liking are entirely optional; a user could hypothetically complete the courses without submitting a single comment or liking a single response.

The platform also offers an activity feed that presents a news-feed-like presentation of course-related activities affecting the user on their program landing page, including an indication with a *link* if a shared reflection written by the user has been commented on or liked, a *pointer* to a peer help response given by the user that has received a response, platform announcements, and more. These discussion boards include familiar communication modes, such as *posts, comments*, and *likes* intended to encourage or even impose mandatory public posts or other forms of interaction with other users.

So-called " *shared reflections*" represent the key element of social interaction in this knowledge exchange platform. These elements are distributed consistently throughout each course, providing a consistent stream of interaction data that can be examined at multiple points within any

given course. Therefore, shared reflections are the focus of measurement around users' relative levels of on-platform social activity.

Sample Descriptives

The sample contains 12,687 users coming from diverse backgrounds and experiences with 52 industries (e.g., consulting, education, energy, and healthcare), 35 fields of study (e.g., accounting, computer science, engineering, psychology, sociology), and 129 countries (e.g., US, China, India, Australia, Brazil). Overall, 40% of users identified as women, and the age ranged between 20 to 76 years with a median age of 34 years (see Table A1). Unfortunately, their work experience and study field information are only available for a considerably smaller sample subset. Hence, the following proportions should be considered as only a crude approximation. The largest two groups of study fields are STEM and business-related fields, with 35% and 18% of users, respectively. The share of users with a background in any Social Science is about 13%. The residual share of users has studied a subject other than the three groups.

A large overall heterogeneity in their activity can be observed in users' interaction behaviors on the knowledge exchange platform. Table A2 displays summary statistics for the peer interactions on the platform. The median user receives 12 comments over their life cycle on the platform¹, conversely, the median of comments given is at 6. Likes, i.e., the faster and arguably cheaper type of peer interaction, are considerably more frequent, with a median of 56 given and 66 received likes, respectively. Hence, the aggregate interaction statistics suggest that the median user receives more interaction instances than they contribute.

Of particular interest to this paper are the interactions at the initial stage of users' life cycle on the platform. In the first phase (called "Module 1"), the share of users giving a minimum of one comment is 62% on average. There is, however, considerable variation between the different cohorts of these shares, ranging from 44% to 75%. Conversely, the share of users giving likes to their peers in the initial phase of their platform participation is higher, with an average of 81%. Similarly, users exhibit variation in this share with a range from 63% to 90% between cohorts. On the receiving end

¹In our setting, a life cycle is the same in every cohort of the course program.

of these interaction instances, 83% of users received at least one comment and 91% at least one like, respectively.

Natural Language Processing For Measure Construction.

We engineered computational measures of the content of users' comments, classifying a comment as agreeing or disagreeing with the original post and classification of a comment as either elaborating on a point or making a simple valuation statement. We had initially planned to include a measure of sentiment for each of the posts. However, as we prepared the labeled data, we noted that, since these posts are informational, they lacked a discernible sentiment and were overwhelmingly neutral.

Comment Classification Matrix. We classified comments as *agreeing* or *disagreeing* with the original post by using a DistilBERT model as the base. We then fine-tuned the sequence classification model on a subset of labeled data from the dataset. For training data, we labeled 82,000 comments as either agreeing or disagreeing with the original post, i.e., such that there are no neutral comments. After fine-tuning the model, we achieved an accuracy of 94%. The trained model was then used to classify all comments as either agreeing or disagreeing or disagreeing with the original post.

We developed a heuristic to classify a comment as either containing elaboration or not. We found that a simple heuristic of comment word count could be used to make the classification. We hand-labeled 906 comments as either elaborating on a new point or simply agreeing or disagreeing with previously established points. Using our heuristic, a comment would be classified as containing elaboration if it contained more than 10 words and not containing elaboration if it contained less than or equal to 10 words. With this heuristic applied to the labeled data, we were able to achieve a 91.5% accuracy rate.

Insert Figure 1 about here

Figure 1 depicts descriptive information on the aforementioned text classification measures. The overall share of agreeing comments among all comments is 66.8%, while the share of purely agreeing comments without elaboration (e.g., "Well said. ") is 15.4%. In particular, elaborating comments are the ones in which participants can share knowledge with their peers. We observe 78.3% of comments being classified as elaborating and a subset of 26.9% of comments as elaborating and disagreeing.

Empirical Strategy

Our goal is to estimate the effect of general and directed peer activity on users' subsequent engagement and persistence on the platform. The key variation we exploit is the share of active peers in the initial period that differs from cohort to cohort, which we argue is plausibly quasi-random. There is *uncertainty*, both on the part of the admissions office and on the part of potential users about the realized set of active peers in each cohort. In order for this quasi-random variation to have a *causal* interpretation, we have to assume that this share is also uncorrelated with other unobservable factors impacting the outcomes of interest at the cohort level (Bostwick and Weinberg 2022; Ammermueller and Pischke 2009).

Insert Figure 2 about here

Since there may be some endogeneity in terms of the timing when users decide to join the platform, e.g., users joining in winter may be different from users joining in spring, we only include so-called user "waves" where subscriptions were sufficiently high so that these users were *randomized* into two (or more) cohorts of similar size. This randomization further strengthens our identification strategy, because the assignment of more or less active peers across these cohorts is arguably exogenous. Together we estimate the following linear regression model:

$$Y_{ic} = \beta_0 + \beta_1 HighCommentGivers_c + \beta_2 HighLikeGivers_c + \beta_3 CommentsReceived_i$$

$$+\beta 4 LikesReceived_i + \beta_5 X_i + \beta_6 YearFE_c + \epsilon_{ic}$$
(1)

Here, Y_{ic} is the dependent variable of user *i* in cohort *c*. It is either their future engagement on the platform, precisely, a dummy indicating whether they gave any *Comment(s)* or *Like(s)* to their

peers in the later periods on the platform (Table A1) or their *Persistence* on the platform measured in the number of completed modules ranging from 1 to 17 (Table A2). We include four main explanatory variables on the right-hand side of the estimation equation. The first two explanatory variables capture the *general peer activity*, i.e., the exposure to the general *undirected* peer interaction activities on the platform. *HighCommentGivers_c* and *HighLikeGivers_c* are dummies indicating cohorts with an above-median number of users giving comments (>0.65% of users) and likes (>0.83 % of users) in the initial phase on the platform, respectively (Table A2). The latter two explanatory variables capture *directed* peer activities toward a user. These are dummy indicators for any *CommentsReceived* and *LikesReceived* in the initial platform period.

 X_{ic} is a vector of user controls, including a dummy for female sex (yes/no), age (in years), living in the US (yes/no), English as an official language at their location (yes/no), metrics of demographically similar peers at the cohort level (# of same gender, # similar age (+/- 2 years), # same country, and # same citizenship peers), procrastination (submission hours to quiz deadline), cohort size (N), and quiz score in the first submitted module as a vague proxy of their platform-related ability. We also include year-fixed effects (FE) to eliminate omitted variable bias caused by unobserved factors that evolve over time but are constant across users on the platform in a given year. Finally, ϵ_{ic} is the residual error term. Table A1 contains the summary statistics for the variables used in the main analyses.

RESULTS

Main Findings

Future Engagement. This paper asks how the presence of initially active peers impacts users' future engagement and persistence on a digital platform. Table 1 presents linear probability estimates of how the cohort composition of more or less active peers affects users' future engagement on the platform. Future engagement is measured by two dummy variables capturing *any comments given m2-7* (see Columns 1 and 2) and *any likes given m2-7* (see Columns 3 and 4) in later periods.

To begin, we consider the effect of *general peer activities* on users' prospective behavior on the platform. First, we test if a high share of peers who initially engage in commenting or liking has a positive effect on users' future engagement and platform persistence. The results show that being in a cohort of users with a high share² of comment-givers (i.e., above the median) increases an individual's likelihood of giving comments in the future by around 2-3 percentage points (pp) (Column 1). However, this result does not hold once we include year-fixed effects (Column 2). Furthermore, there seems to be no of effect being in a cohort with a high share of comment-givers on the propensity to distribute likes in the future. The effect is negligible with less than 0.5 pp and the switch of the sign of the coefficient from Column 3 to Column 4, i.e., without and with year-fixed effects included in the model, hints towards further instability of any directional effect.

With regards to general peer activity in terms of liking, we find that users in a cohort with an above-median share of peers distributing likes in the initial period are less likely to give comments in future periods (see Columns 1 and 2 of Table 1). This effect is 4.5 pp large and statistically significant at the 5% level, including a large set of individual user control variables. However, once year-FEs are controlled for, the effect decreases to 3 pp and lacks statistical significance at conventional levels. The high share of peers distributing likes during the initial period on the platform also does not seem to increase users' probability of giving likes in future periods. The estimated coefficients shown in Columns 3 and 4 are small and not statistically significant at any conventional level. Hence, overall we do not find support for our hypothesis that a high share of *undirected* peer activity affects users future platform engagement.

Next, we turn our attention to the early peer activities *directed* toward the users (see rows 3 and 4 in Table 1). The estimates indicate that receiving a comment in the initial period increases the propensity to give comments by 20 pp and likes by 12 pp, respectively. Conversely, receiving likes initially is associated with an increased propensity for commenting and liking in future periods of 10 pp and 14 pp, respectively. Hence, the effects are strongest by interaction type, i.e., comments in the first period are most strongly related to future comments, and vice versa for likes. All estimates

²As detailed in section 1, a "high share" indicates cohorts with an above-median number of users giving comments (>0.65% of users) and likes (>0.83 % of users) in the initial period on the platform, respectively. See Table A2.

of the early peer activity directed to users on their future platform engagement are statistically significant at the 1% level when including user-level controls and year FE. Thus, we find that *directed* user engagement at an early stage, is positively associated with individual users' future platform engagement, in terms of providing comments and giving likes to other users.

Insert Table 1 about here

Platform Persistence. Table 2 presents the estimates for platform persistence as the second outcome of interest. We start again by looking at undirected *general peer activity* as measured by the share of active comment- and like-givers in each cohort. We find that early actively commenting peers have a low, positive effect on platform persistence, albeit insignificantly. Furthermore, our findings indicate that an initial high share of active peers giving likes significantly decreases platform persistence by half a module or about 3% of overall persistence.

Turning to the *directed* peer activity (see rows 3 and 4 in Table 2), we find that receiving early comments and likes correlates with a 1 to 1.5 module increase in (or 7% and 9% higher) platform persistence, respectively. Taken together, these results suggest that while having a high share of general, undirected peer activity has no (or even a small negative effect) on platform persistence, early-stage *directed* peer activity can boost platform persistence significantly.

Insert Table 2 about here

Heterogeneity Analysis

The presented main results might vary for different types of users. Hence, we assess heterogeneity along the lines of the following user characteristics: median age, residency in the United States, being on the platform during the COVID-19 pandemic, user gender, and split future engagement into early versus later periods on the platform. All specifications discussed below include the same user-level controls as in our main specification and year-fixed effects. *Age.* The first variable for which we consider heterogeneity among users is their age. It could be that different age groups have different levels of familiarity with platform-based communication, i.e., younger users are more familiar with platform interfaces or navigating communication thereon (Venter 2017) or that career concerns are heterogeneous by age (Xu, Nian, and Cabral 2020).

Table A6 depicts the main estimates split by users' *median age* of 32. This split of the sample can serve as a rough proxy of users' career stages. Columns (1) to (3) report estimates for users below the median age and Columns (4) to (6) depict the equivalent for users above the median age. Looking at general peer activity, we see a clear differential effect by user age. Whereas having a high share of active peers does not seem to have any effect on younger users' platform engagement or persistence, we find that a high share of peers commenting in the first module is associated with a 3.9pp increase in the likelihood of giving comments in later modules for above median age users (significant at the 10% level). Furthermore, our results indicate that a high share of peers giving likes in 1st module has a significant negative effect on future commenting behavior (7 pp) and persistence (0.72 modules i.e. 4.2% of 7 modules) for users with an above-median age. These effects are statistically significant at the 5 and 1 percent level, respectively.

When regarding *directed* peer interactions (see rows 3 and 4), there are positive associations between receiving comments and likes on future engagement and platform persistence for both age groups. The effects appear larger for users above the median age.

US Residency. As 57% of users are residents of the US, we split the sample by users in and outside this large group to account for unobserved factors potentially more similar for US users than for others, i.e., communication culture, the popularity of certain social media, education platforms or (partial) time zone proximity.

Table A7 shows the main estimation results split by US residency status. Columns (1) to (3) report estimates for users residing in the US and Columns (4) to (6) for users who reside outside the US. Overall, the results for the two groups are rather similar. There are only two small differences. First, the results suggest that an initially high share of commenting peers is associated with a 4 pp higher likelihood of future commenting for US residents. This estimate is statistically significant at

the 5 percent level. Conversely, for non-US residents, the point estimate is positive as well and at 2 pp, however, not statistically significant. Second, the other estimate that is weakly significant, is the 0.47 modules reduction in average persistence of US residents in a cohort of active like-givers (see Column 3). However, also for non-US residents, this coefficient is negative, albeit not significant.

When turning to the impact of interaction instances directed to users from their peers in the initial period on the platform, the results in Table A7 indicate economically and statistically significant, positive correlations between the receipt of comments and likes early on future engagement as well as the persistence on the platform. While there are differences in magnitudes for US and non-US residents, the effect sizes are overall comparable.

COVID-19 Pandemic. In Table A8, we present the main estimation results split by cohorts pre- and during the COVID-19 period.³ The main motivation behind presenting this split is to account for a potentially different selection of users on the platform before or during the global COVID-19 pandemic. Columns (1) to (3) report estimates for users on the platform during the COVID-19 pandemic and Columns (4) to (6) for users who were on the platform pre-pandemic. Having a high share of commenting peers in the initial period on the platform significantly increases users' probability of giving comments in the future by 4.2 pp for users in the COVID-19 period. The corresponding point estimate is at 1.9 pp and statistically insignificant for users on the platform pre-Covid. Interestingly, a high share of commenting peers early on is associated with slightly lower persistence on the platform pre-Covid (see Column 6). While estimation results suggest a slight impact of a high, initial share of like-givers on prospective engagement and persistence, however, not statistically significant (Columns 4 - 6) pre-pandemic, the corresponding coefficient estimates are negligibly small during the COVID-19 period.

Turning from the initial general platform activity to peers' interaction instances received by users, the robust, results in Table A8 indicate a positive and highly significant relationship between receiving comments and likes early users' future platform engagement and persistence both, for

³We define user cohorts to be in the "COVID-19 period" starting from March 2020 because the WHO declared it COVID-19 a global pandemic during this month. *Source:* https://www.cdc.gov/museum/timeline/covid19. html, last retrieved on Feb 25, 2024.

users pre- and during COVID-19. The coefficient estimates are slightly larger for users during the pandemic throughout. In particular, receiving any comment(s) in the first period increases the likelihood of giving comments and likes in the future by 26.5 pp (Column 2) and 23.2 pp (Columns 3) during the pandemic, respectively, and by 24.2 pp (Column 4) and 19.1 pp (Column 5) pre-pandemic, respectively. In terms of platform persistence, the impact of receiving comments and likes in the initial period is approximately 0.5 and one module greater during the COVID-19 period, respectively (see Columns 3 and 6).

Gender. As a next heterogeneity check, we split the sample by gender as this is among the most salient demographic features and a large strand of research in social sciences presents robust evidence for behavioral gender differences. Table A9 shows the main estimation results for women in Columns (1) to (3) and for men in Columns (4) to (6).

Overall, the results suggest several notable gender differences. First, the presence of more comment-giving peers in the first period on the platform increases men's likelihood of actively commenting in future periods by 4.4 pp (Column 4). The coefficient estimates are significant at the 5%-level. Conversely, for women, the point estimate is 2.7 pp and not significant (Column 1). However, while also not statistically significant, only for women the impact of early active comment-givers on the propensity to distribute likes in the future and on persistence is negative (see Columns 2 and 3).

Second, the main negative effect of a high share of peers giving likes in the first period on future commenting and persisting appears to be slightly larger and only (weakly) significant for men. Third, while the regression estimates for the association between receiving comments and likes from one's peers in the initial platform period are positive and statistically significant at the 1%-level, the coefficient estimates of receiving comments seem to be slightly larger for women. The differences are economically significant, in particular for future engagement. Receiving comment(s) in the first period is associated with 30.3 pp and 22.6 pp for women and men, respectively (see Columns 1 and 4). Similarly, the propensity to distribute likes in future periods is 25.4 pp and 19.0 pp higher for women and men, respectively, upon the receipt of comment(s) from peers initially (see Columns 2

and 5).

Finally, there are also differences in the impact of receiving any like(s) by peer users in the first period pointing towards gender differences in reciprocating in online interactions. While the coefficient estimates of receiving likes by pears early suggest a slightly larger, positive impact on women's likelihood to give likes in future periods (see Columns 2 and 5), they are marginally larger for men considering future commenting as an outcome (see Columns 1 and 4).

Early and Later Periods on the Platform. Lastly, we test whether the potential impacts of peers' early platform activity on users' behavior has rather short or medium-term impacts. For this purpose, we split the future engagement outcomes in *early* versus *late* modules, i.e., periods on the platform. Table A10 presents results for future commenting and Table A11 for future liking behaviors.

Overall, the results do not indicate stark differential impacts for splitting the outcome variables by early versus later future periods on the platform, i.e., not pointing towards a rather short-lived impact of initial general or directed peer interactions on users' prospective engagement.

Mechanism: Comment Type

The main analysis suggests that receiving directed comments from peers at an early stage significantly increases users' future engagement and persistence on the platform. Against the backdrop of these empirical results, we investigate whether different types of comments matter for fostering engagement and persistence.

As detailed in section 1, we employ computational measures of users' comments' text, classifying each comment as *agreeing* or *disagreeing* with the original post and as either *elaborating* on a point or making a simple *valuation* statement. Consequently, we obtain four types of comments: (1) elaborating & agreeing comments, (2) elaborating & disagreeing comments, (3) purely agreeing comments, and (4) other comments. Note that the latter category contains all comments that could not fit in the aforementioned three categories. These are, however, not "purely disagreeing" in nature as such comments have not been found much on the platform. To analyze the separate impact of the

distinct comment types, we replaced the explanatory indicator variable Comment(s)Received - m1 in Equation 1 with dummies for each comment type in separate regressions.

In aggregate, our estimations indicate that the nature of comments received from peers during the initial stage of users' presence might matter for their subsequent behavior on the platform. Specifically, both elaborating types and purely agreeing comments exhibit statistically significant positive correlations with users' future engagement levels and platform persistence. Notably, the combined presence of agreement & elaboration within comments manifests the most robust, positive association with all three studied behavioral outcomes. Conversely, our findings indicate a negative correlation between other comment types, i.e., those neither elaborative nor agreeing, and users' future engagement and platform persistence. These findings underscore the differential impact of distinct comment types on users' propensity for continued engagement and interaction within the platform ecosystem.

Insert Figure 3 about here

Figure 3 presents coefficient plots derived from regressing the receipt of diverse comment types from peers during the initial platform period (m1) on the likelihood of engaging in future commenting activities across subsequent periods (m2-7), depicted along the horizontal axis. The first row serves as the benchmark, indicating that the receipt of any comment type during the initial period is associated with a substantial increase of 20.4 pp in the probability of engaging in future commenting (detailed in Column 2 of 1). Further analysis at the individual comment type level reveals nuanced effects on prospective commenting behavior. Particularly, receiving elaborative & agreeing comments during the initial period exhibits the most robust positive correlation, leading to a significant increase of 14.1 pp in the likelihood of future commenting. Similarly, the receipt of purely affirming comments, i.e., without elaboration, still yields a notable 11.7 pp increase in the probability of subsequent commenting. Conversely, the effect of receiving elaborative but *dis*agreeing comments yields a slightly lower yet substantial coefficient estimate of 10.2 pp, indicative of a comparable influence on future commenting behavior. In contrast, the receipt of

'other comments' - those lacking both elaboration and agreement - exhibits a contrasting effect, resulting in a decrease of 12.2 pp in the likelihood of engaging in future commenting activities.

Insert Figure 4 about here

Coefficient plots of the relationship between receiving various comment types during the initial period and the subsequent engagement in providing like endorsements are shown in Figure 4.

While the estimates generally manifest lower magnitudes compared to those observed for the association between comment types and future commenting behavior, they remain economically meaningful. Specifically, receiving agreeing & elaborating comments during one's initial platform life-cycle exhibits the strongest correlation with the propensity to offer likes to other users, with a coefficient estimate of 8.6 pp. Notably, the effects of receiving purely agreeing comments and agreeing comments with elaboration are also noteworthy, yielding comparable increases in the

likelihood of distributing likes by 6.6 pp and 5.3 pp, respectively. The results indicate a non-negligible, negative correlation between receiving "other" comments, i.e., not exhibiting the aforementioned features, and users' likelihood of liking others' posts and comments of a 7.6 pp decrease.

Insert Figure 5 about here

Figure 5 exhibits coefficient plots of regressing the receipt of different comment types by peers in the first period (m1) on platform persistence measured by the number of additional modules on the horizontal axis. In the first row, the benchmark coefficient estimating the effect of receiving any type of comment on persistence is 1.06 additional modules (see Column 2 in Table 2). Turning to the estimates for each comment type's effect, the receipt of agreeing & elaborating comments in the first period exhibits the strongest, positive correlate of 0.9 additional modules. When receiving comments that are "purely" agreeing without the elaborating component, the estimated increase in platform persistence is at 0.56. For disagreeing & elaborating comments, the equivalent coefficient estimate is of the same size. Finally, users who receive "other comments" in the first period on the

platform that offer neither any elaboration nor agreement to their original post tend to persist about 0.8 modules equals 4.7% of the maximum life cycle of 17 modules.

The full, corresponding estimation results are contained in the tables A3, A4, and A5 for the outcomes *CommentsGiven* – m2 – 7, *LikesGiven* – m2 – m7, and *Persistence*, respectively. They show that all aforementioned coefficients of interest discussed in this section are statistically significant at the 1% level in models including the same user-level controls and year-FEs as the main analyses.

DISCUSSION AND CONCLUSION

Nowadays, an indispensable share of knowledge sourcing and exchange in organizations, education settings, or related to private interests happens on digital platforms. However, the core issue many of such platforms share is the decay of user engagement over time. Most of the evidence from the literature on online platforms has focused on interactions surrounding users' *own behavior*, e.g., on determinants of knowledge sharing and sourcing. While there is robust evidence from economics that peers can positively matter in various offline settings ranging from the traditional classroom to consumption choices, little is known about how peers' initial activities may impact users' future behavior online. This paper asks if and how initially active peers can impact users' future engagement and persistence on a platform.

Overall, our empirical findings show that an initial high share of active peers giving likes significantly decreases platform persistence by about 3%, also slightly diminishing later commenting probability. Early actively commenting peers correlate with a low, positive likelihood of giving comments in the future and platform persistence, albeit insignificantly. Receiving early comments and likes correlates with a robust ten-plus percentage point increase in future engagement, with the strongest impact observed for the same type of interaction. Additionally, early comments and likes received are associated with a 7% and 9% higher platform persistence, respectively.

This study makes several contributions to the literature. First, our findings have implications for the design and structure of online education and digital communities. Studies on digital

communities have shown that only a small fraction of members typically contribute knowledge therein and that it can be difficult to retain members' engagement over time (Faraj and Johnson 2011; Ren, Kraut, and Kiesler 2007; Wasko and Faraj 2005; Mickeler et al. 2023; Claussen, Kretschmer, and Mayrhofer 2013). We contribute to this literature, by studying the timing, type, and content of online peer interactions to gain an in-depth understanding of how these factors influence subsequent user platform behavior.

Second, previous evidence from the peer effects literature strongly indicates that social interactions matter for improving learning outcomes. While having been widely studied in economics due to their impact on education and workplace behaviors, most previous work on peer effects has focused on various offline contexts, e.g., in conventional education in a physical classroom (Sacerdote 2001; Zimmerman 2003; Calvó-Armengol, Patacchini, and Zenou 2009; Ammermueller and Pischke 2009), but also with regards to agricultural technology adoption in developing countries (Foster and Rosenzweig 1995; Munshi 2004; Bandiera and Rasul 2006), entrepreneurship (Lerner and Malmendier 2013), and consumption behavior (Moretti 2011). Despite these studies offering ample evidence for peer effects to have an impact on behavioral choices and future (learning) outcomes, the precise mechanisms of how exactly and under which conditions interactions take place remain mostly unobserved in offline settings. In our study, the unique and detailed data on online peer interactions and outcomes of interest over a longer time allows us to shed some light on these effects and mechanisms.

Third, we advance management literature's understanding of content-related online communication behaviors by introducing a novel method of comment classification. For this purpose, we apply state-of-the-art Natural Language Processing (NLP) methods on comment texts to classify these into agreeing versus disagreeing versus elaborating on comments. This classification intends to bear new insights into communication patterns and their impact on successful digital knowledge transfers and learning outcomes. Hereby, we also add to a broader line of works in social science that makes use of novel text data methods (Gentzkow, Kelly, and Taddy 2019; Athey and Imbens 2019).

26

Fourth, our paper has implications for the scholarship on labor force training (Bidwell and Briscoe 2010; Cappelli 2015) due to our novel focus on a diverse, global sample of working professionals and executives in this context. To the best of our knowledge, this paper is among the first to advance the understanding of peer effects among working professionals in a virtual learning environment. Our focus on MBA training allows us to assess platform behaviors in a highly labor market-relevant setting. First, the courses' content in and of itself is a domain of knowledge that is highly employable. Second, the sampled courses are feasible alongside full-time occupation. Together, these features bear particularly high relevance for management scholars and practitioners as continuous human capital accumulation is essential to firm growth and innovation (Dragoni et al. 2009). Moreover, due to the increasing pace at which technical skills become outdated (Deming and Noray 2018), firms strategically seek to upskill their global labor force to remain competitive (Illanes et al. 2018). For this purpose, digital training provides a scalable and cost-efficient approach. These insights are likely applicable to intra-organizational knowledge platforms as well.

On a critical note, the utilized data source comprises a few limitations. First and foremost, data on a comprehensive set of demographic background information is only available for a subset of the sample limiting the internal validity of these measures. Second, some users drop out during the program. Although the observed attrition does not appear to occur systematically, we cannot rule out unobserved factors contributing to the drop-out of some users. Moreover, we may note that the effect sizes of the employed regression models are small in magnitude limiting their economic significance. Lastly, we provide purely correlational evidence on the directed peer activity measures, i.e., the *received* comments and likes. Hence, it remains for future work to exogenously vary factors that can foster or hinder beneficial, knowledge-sharing interactions in online contexts like the one this project investigates.

Gaining a better understanding of these heterogeneities in social interactions, motivations, and mechanisms through which beneficial information exchange can be fostered online can be a promising avenue for future work. We suggest that designing online environments to foster social interactions, in particular, encouragement and constructive knowledge sharing, among peers may be more effective in promoting continued engagement and platform persistence.

REFERENCES

- Ammermueller, Andreas and Jörn-Steffen Pischke (2009). "Peer effects in European primary schools: Evidence from the progress in international reading literacy study." In: *Journal of Labor Economics* 27.3, pp. 315–348.
- Ardichvili, Alexander, Vaughn Page, and Tim Wentling (2003). "Motivation and barriers to participation in virtual knowledge-sharing communities of practice." In: *Journal of knowledge management* 7.1, pp. 64–77.
- Athey, Susan and Guido W Imbens (2019). "Machine learning methods that economists should know about." In: *Annual Review of Economics* 11, pp. 685–725.
- Autor, David H, Frank Levy, and Richard J Murnane (2003). "The skill content of recent technological change: An empirical exploration." In: *The Quarterly Journal of Economics* 118.4, pp. 1279– 1333.
- Baek, Jiye and Jesse Shore (2020). "Forum size and content contribution per person: A field experiment." In: *Management Science* 66.12, pp. 5906–5924.
- Bandiera, Oriana and Imran Rasul (2006). "Social networks and technology adoption in northern Mozambique." In: *The Economic Journal* 116.514, pp. 869–902.
- Bettinger, Eric, Jing Liu, and Susanna Loeb (2016). "Connections matter: How interactive peers affect students in online college courses." In: *Journal of Policy Analysis and Management* 35.4, pp. 932–954.
- Bidwell, Matthew and Forrest Briscoe (2010). "The dynamics of interorganizational careers." In: *Organization Science* 21.5, pp. 1034–1053.
- Bidwell, Matthew, Forrest Briscoe, et al. (2013). "The employment relationship and inequality: How and why changes in employment practices are reshaping rewards in organizations." In: *Academy of Management Annals* 7.1, pp. 61–121.
- Borgatti, Stephen P and Rob Cross (2003). "A relational view of information seeking and learning in social networks." In: *Management Science* 49.4, pp. 432–445.
- Bostwick, Valerie K and Bruce A Weinberg (2022). "Nevertheless she persisted? Gender peer effects in doctoral STEM programs." In: *Journal of Labor Economics* 40.2, pp. 397–436.
- Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou (2009). "Peer effects and social networks in education." In: *The Review of Economic Studies* 76.4, pp. 1239–1267.
- Cappelli, Peter H (2015). "Skill gaps, skill shortages, and skill mismatches: Evidence and arguments for the United States." In: *ILR Review* 68.2, pp. 251–290.
- Chen, Langtao, Aaron Baird, and Detmar Straub (2019). "Why do participants continue to contribute? Evaluation of usefulness voting and commenting motivational affordances within an online knowledge community." In: *Decision Support Systems* 118, pp. 21–32.
- Claussen, Jörg, Tobias Kretschmer, and Philip Mayrhofer (2013). "The effects of rewarding user engagement: The case of Facebook apps." In: *Information Systems Research* 24.1, pp. 186–200.
- Deming, David J and Kadeem L Noray (2018). *STEM careers and the changing skill requirements of work*. Tech. rep. National Bureau of Economic Research.
- Dragoni, Lisa et al. (2009). "Understanding managerial development: Integrating developmental assignments, learning orientation, and access to developmental opportunities in predicting managerial competencies." In: *Academy of Management Journal* 52.4, pp. 731–743.
- Faraj, Samer and Steven L Johnson (2011). "Network exchange patterns in online communities." In: *Organization Science* 22.6, pp. 1464–1480.

- Faraj, Samer, Georg von Krogh, et al. (2016). "Special section introduction—Online community as space for knowledge flows." In: *Information Systems Research* 27.4, pp. 668–684.
- Feld, Jan and Ulf Zölitz (2017). "Understanding peer effects: On the nature, estimation, and channels of peer effects." In: *Journal of Labor Economics* 35.2, pp. 387–428.
- (2022). "The effect of higher-achieving peers on major choices and labor market outcomes." In: *Journal of Economic Behavior & Organization* 196, pp. 200–219.
- Foster, Andrew D and Mark R Rosenzweig (1995). "Learning by doing and learning from others: Human capital and technical change in agriculture." In: *Journal of Political Economy* 103.6, pp. 1176–1209.
- Gallus, Jana and Bruno S Frey (2016). "Awards: A strategic management perspective." In: *Strategic Management Journal* 37.8, pp. 1699–1714.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019). "Text as data." In: *Journal of Economic Literature* 57.3, pp. 535–574.
- Golsteyn, Bart HH, Arjan Non, and Ulf Zölitz (2021). "The impact of peer personality on academic achievement." In: *Journal of Political Economy* 129.4, pp. 1052–1099.
- Gong, Jie, Tracy Xiao Liu, and Jie Tang (2021). "How monetary incentives improve outcomes in MOOCs: Evidence from a field experiment." In: *Journal of Economic Behavior & Organization* 190, pp. 905–921.
- Hwang, Elina H, Param Vir Singh, and Linda Argote (2015). "Knowledge sharing in online communities: Learning to cross geographic and hierarchical boundaries." In: Organization Science 26.6, pp. 1593–1611.
- Illanes, Pablo et al. (2018). "Retraining and reskilling workers in the age of automation." In: *McKinsey Global Institute* 8.
- Jones, Quentin, Gilad Ravid, and Sheizaf Rafaeli (2004). "Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration." In: *Information systems research* 15.2, pp. 194–210.
- Khobzi, Hamid, Raymond YK Lau, and Terence CH Cheung (2019). "The outcome of online social interactions on Facebook pages: A study of user engagement behavior." In: *Internet Research* 29.1, pp. 2–23.
- Kretschmer, Tobias et al. (2022). "Platform ecosystems as meta-organizations: Implications for platform strategies." In: *Strategic Management Journal* 43.3, pp. 405–424.
- Kwon, Kyungbin et al. (2019). "Effects of different types of instructor comments in online discussions." In: *Distance Education* 40.2, pp. 226–242.
- Lerner, Josh and Ulrike Malmendier (2013). "With a little help from my (random) friends: Success and failure in post-business school entrepreneurship." In: *The Review of Financial Studies* 26.10, pp. 2411–2452.
- Li, Cathy and Farah Lalani (2020). "The COVID-19 pandemic has changed education forever." In: *World Economic Forum*. Vol. 29.
- Littlejohn, Allison et al. (2016). "Learning in MOOCs: Motivations and self-regulated learning in MOOCs." In: *The Internet and Higher Education* 29, pp. 40–48.
- Loh, Johannes (2022). "Selection, Consumption, and New Music Exploration in an Online Social Network: A Dyadic Approach." In.
- Makos, Alexandra et al. (2013). "Use of a "like" button in a collaborative online learning environment." In.

- Manski, Charles F (1993). "Identification of endogenous social effects: The reflection problem." In: *The Review of Economic Studies* 60.3, pp. 531–542.
- Mickeler, Maren et al. (2023). "Knowledge seeking and anonymity in digital work settings." In: *Strategic Management Journal*.
- Moretti, Enrico (2011). "Social learning and peer effects in consumption: Evidence from movie sales." In: *The Review of Economic Studies* 78.1, pp. 356–393.
- Munshi, Kaivan (2004). "Social learning in a heterogeneous population: Technology diffusion in the Indian Green Revolution." In: *Journal of Development Economics* 73.1, pp. 185–213.
- Ren, Yuqing, F Maxwell Harper, et al. (2012). "Building member attachment in online communities: Applying theories of group identity and interpersonal bonds." In: *MIS quarterly*, pp. 841–864.
- Ren, Yuqing, Robert Kraut, and Sara Kiesler (2007). "Applying common identity and bond theory to design of online communities." In: *Organization Studies* 28.3, pp. 377–408.
- Rietveld, Joost and Melissa A Schilling (2021). "Platform competition: A systematic and interdisciplinary review of the literature." In: *Journal of Management* 47.6, pp. 1528–1563.

Rosendahl Huber, Laura, Jacqueline N Lane, and Karim R Lakhani (2020). "Learning With People Like Me: The Role of Age-Similar Peers on Online Business Course Engagement." In.

Sacerdote, Bruce (2001). "Peer effects with random assignment: Results for Dartmouth roommates." In: *The Quarterly Journal of Economics* 116.2, pp. 681–704.

Surma, Jerzy (2016). "Social exchange in online social networks. The reciprocity phenomenon on Facebook." In: *Computer Communications* 73, pp. 342–346.

- Tamayo, Jorge et al. (2023). "Reskilling in the age of AI." In: Harvard Business Review 21.
- Venter, Elza (2017). "Bridging the communication gap between Generation Y and the Baby Boomer generation." In: *International journal of Adolescence and Youth* 22.4, pp. 497–507.
- Wang, Chong, Xiaoquan Zhang, and Il-Horn Hann (2018). "Socially nudged: A quasi-experimental study of friends' social influence in online product ratings." In: *Information Systems Research* 29.3, pp. 641–655.
- Wasko, Molly McLure and Samer Faraj (2005). "Why should I share? Examining social capital and knowledge contribution in electronic networks of practice." In: *MIS Quarterly*, pp. 35–57.
- Xu, Lei, Tingting Nian, and Luis Cabral (2020). "What makes geeks tick? a study of stack overflow careers." In: *Management Science* 66.2, pp. 587–604.
- Zhang, Chen, Jungpil Hahn, and Prabuddha De (2013). "Research note—Continued participation in online innovation communities: Does community response matter equally for everyone?" In: *Information Systems Research* 24.4, pp. 1112–1130.
- Zhang, Dennis J, Gad Allon, and Jan A Van Mieghem (2017). "Does social interaction improve learning outcomes? Evidence from field experiments on massive open online courses." In: *Manufacturing & Service Operations Management* 19.3, pp. 347–367.
- Ziegele, Marc et al. (2018). "The dynamics of online news discussions: Effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions." In: *Information, Communication & Society* 21.10, pp. 1419–1435.
- Zimmerman, David J (2003). "Peer effects in academic outcomes: Evidence from a natural experiment." In: *Review of Economics and Statistics* 85.1, pp. 9–23.

FIGURE 1 Given Comment Type Shares by Course



FIGURE 2 Histograms of the Share of Active Peers in Module 1 by Interaction Type



FIGURE 3 Future Commenting: Coefficient Plots by Comment Type







FIGURE 5 Platform Persistence - Coefficient Plots by Comment Type



	(1)	(2)	(3)	(4)
	Comments given	Comments given	Likes given	<i>Likes give</i> n
	m2-7 (1/0)	m2-7 (1/0)	m2-7 (1/0)	m2-7 (1/0)
High # comment givers m1 (1/0)	0.028*	0.019	0.004	-0.002
	[-0.003,0.058]	[-0.012,0.050]	[-0.019,0.027]	[-0.025,0.021]
	(0.016)	(0.016)	(0.012)	(0.012)
High # like givers m1 (1/0)	-0.045**	-0.030	0.013	0.023
	[-0.087,-0.004]	[-0.073,0.012]	[-0.019,0.045]	[-0.009,0.055]
	(0.021)	(0.022)	(0.016)	(0.016)
Comment(s) received m1 (1/0)	0.205***	0.204***	0.122***	0.121***
	[0.180,0.229]	[0.179,0.228]	[0.101,0.143]	[0.100,0.143]
	(0.013)	(0.013)	(0.011)	(0.011)
Like(s) received m1 (1/0)	0.107***	0.104***	0.147***	0.146***
	[0.067,0.146]	[0.065,0.143]	[0.109,0.185]	[0.108,0.184]
	(0.020)	(0.020)	(0.019)	(0.019)
Means Dep. Variable	0.658	0.658	0.847	0.847
Controls	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
Adjusted R ²	0.097	0.101	0.086	0.089
Ν	12,687	12,687	12,687	12,687

 TABLE 1

 Linear Probability Models Predicting Future Engagement

Notes. The dependent variable in Columns (1) and (2) *Comments given m2-7* equals 1 if any comments were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in Columns (3) and (4) *Likes given m2-m7* equals 1 if any likes were given in later modules 2 to 5/7 in the program and 0 otherwise. *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. *Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.*

	(1)	(2)
	Persistence	Persistence
<i>High # comment givers m1</i> (1/0)	0.089	0.084
	[-0.135,0.312]	[-0.139,0.308]
	(0.114)	(0.114)
High # like givers m1 (1/0)	-0.559***	-0.518***
	[-0.899,-0.219]	[-0.862,-0.175]
	(0.174)	(0.175)
Comment(s) received m1 (1/0)	1.058***	1.058***
	[0.840,1.275]	[0.840,1.275]
	(0.111)	(0.111)
Like(s) received m1 (1/0)	1.455***	1.460***
	[1.039,1.872]	[1.043,1.876]
	(0.212)	(0.212)
Maara Dan Warishla	15 (4 2	15 (42
Means Dep. variable	15.045	15.045
Adjusted K ⁻	0.137	0.137
Controls Veer EE	ies	ies
	10 12 697	12 697
IN	12,087	12,087

TABLE 2OLS Regressions Predicting Platform Persistence

Notes. The dependent variable in all Columns *Persistence* is the number of completed modules of the entire program (0-17). *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

APPENDIX A. TABLES

TABLE A1Summary Statistics

	Ν	Mean	Median	SD	Min	Max
Dependent Variables						
Comment(s) given in m2-7 $(1/0)$	12.687	0.66	1	0.47	0	1
Like(s) given in m2-7 $(1/0)$	12.687	0.85	1	0.36	Õ	1
Persistence (# modules)	12,687	15.64	17	3.64	1	17
Independent Variables						
A: Peer Interaction Variables						
High # comment givers (1/0)	12,687	0.53	1	0.50	0	1
High # like givers $(1/0)$	12,687	0.62	1	0.48	0	1
Comment(s) received m1 (1/0)	12,687	0.83	1	0.38	0	1
Like(s) received m1 (1/0)	12,687	0.94	1	0.23	0	1
B: Control Variables						
Female (%)	12,687	0.40	0	0.49	0	1
Age (yrs)	12,687	33.98	32	8.49	20	76
US resident (1/0)	12,687	0.57	1	0.50	0	1
Official language English (1/0)	12,687	0.62	1	0.49	0	1
# same age peers (+/-2 yrs)	12,687	87.53	80	54.79	1	290
# peers from same country	12,687	134.44	170	119.54	1	392
# peers w/ same gender	12,687	197.19	195	53.13	87	293
# peers w/same citizenship	12,687	90	60	87.15	1	303
Cohort size (N)	12,687	384.11	376	73.60	271	566
Submission before deadline (h)	12,687	43.77	18.63	74.15	0	3,022.47
Quiz score in fin. accounting m1	12,687	80.45	85	16.62	0	100

N	Mean	Median	SD	Min	Max
12,687	0.53	1	0.50	0	1
12,687	0.64	0.65	0.07	0.44	0.75
12,687	0.62	1	0.48	0	1
12,687	0.81	0.83	0.06	0.63	0.90
12,687	22.15	12	34.65	0	762
12,687	21.8	6	51.24	0	1,008
12,687	0.66	1	0.47	0	1
12,687	0.83	1	0.38	0	1
12,687	130.47	66	202.71	0	3,031
12,687	131.79	56	263.48	0	13,531
12,687	112.81	43	225.98	0	9,542
12,687	0.94	1	0.23	0	1
	N 12,687 12,687 12,687 12,687 12,687 12,687 12,687 12,687 12,687 12,687 12,687 12,687	N Mean 12,687 0.53 12,687 0.64 12,687 0.62 12,687 0.81 12,687 22.15 12,687 21.8 12,687 0.66 12,687 0.83 12,687 130.47 12,687 131.79 12,687 12.81 12,687 0.94	N Mean Median 12,687 0.53 1 12,687 0.64 0.65 12,687 0.62 1 12,687 0.81 0.83 12,687 22.15 12 12,687 21.8 6 12,687 0.66 1 12,687 0.83 1 12,687 0.83 1 12,687 130.47 66 12,687 131.79 56 12,687 12.81 43 12,687 0.94 1	N Mean Median SD 12,687 0.53 1 0.50 12,687 0.64 0.65 0.07 12,687 0.62 1 0.48 12,687 0.81 0.83 0.06 12,687 22.15 12 34.65 12,687 21.8 6 51.24 12,687 0.66 1 0.47 12,687 0.83 1 0.38 12,687 130.47 66 202.71 12,687 130.47 56 263.48 12,687 112.81 43 225.98 12,687 0.94 1 0.23	N Mean Median SD Min 12,687 0.53 1 0.50 0 12,687 0.64 0.65 0.07 0.44 12,687 0.62 1 0.48 0 12,687 0.81 0.83 0.06 0.63 12,687 22.15 12 34.65 0 12,687 21.8 6 51.24 0 12,687 0.66 1 0.47 0 12,687 0.83 1 0.38 0 12,687 130.47 66 202.71 0 12,687 131.79 56 263.48 0 12,687 112.81 43 225.98 0 12,687 0.94 1 0.23 0

TABLE A2 Summary Statistics of Peer Interaction Variables

	(1) Comments given	(2) Comments given	(3) Comments given	(4) Comments given	(5) Comments given
	(0/1) /-7111	(0/1) /-7111	(0/1) /-7111	(0/1) /-7111	(0/1) /-7111
High # comment givers m1 (1/0)	0.019 [-0.012,0.050]	0.017 [-0.014,0.049]	0.021 [-0.010,0.052]	0.021 [-0.010,0.053]	0.019 [-0.012,0.050]
High # like givers m1 (1/0)	(0.016) -0.030 F 0.073 0.0121	(0.016) -0.018 1.0.021 0.0241	(0.016) -0.028 5.0.020 0.111	(0.016) -0.034 5.0.077.0.0001	(0.016) -0.027
Comment(s) received m1 (1/0)	[210:0;210:0-] (0.022) 0.204*** [0.179,0.228]	(0.022)	(0.022)	(0.022)	(0.022)
Purely agree comment(s) received m1 (1/0)	(0.013)	0.117*** [0.099,0.135]			
Agree & elaborate comment(s) received m1 (1/0)		(600.0)	0.141^{***} [0.121,0.161]		
Disagree & elaborate comment(s) received m1 (1/0)			(0.010)	0.102^{***} $[0.084, 0.120]$	
Other comment(s) received (1/0)				(0000)	-0.122*** [-0.142,-0.102]
Like(s) received m1 (1/0)	$\begin{array}{c} 0.104^{***}\\ [0.065, 0.143]\\ (0.020) \end{array}$	0.179*** [0.142,0.217] (0.019)	0.133*** [0.095,0.172] (0.020)	0.167*** [0.129,0.204] (0.019)	(0.010) 0.190*** [0.153,0.227] (0.019)
Means Dep. Variable Controls	0.658 Ves	0.658 Yes	0.658 Yes	0.658 Yes	0.658 Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Adjusted R2 N	0.101 12,687	0.091 12,687	0.095 12,687	0.089 12,687	0.089 12,687

	(1) Likes given m2-7 (1/0)	(2) Likes given m2-7 (1/0)	(3) <i>Likes given</i> m2-7 (1/0)	(4) <i>Likes given</i> m2-7 (1/0)	(5) Likes given m2-7 (1/0)
High # comment givers (1/0)	-0.002 [-0.025,0.021] (0.012)	-0.003 [-0.025,0.020] (0.012)	-0.001 [-0.024,0.022]	-0.000 [-0.023,0.022]	-0.002 [-0.025,0.020]
High # like givers (1/0)	0.0212) 0.023 [-0.009,0.055] 0.016)	(0.012) 0.030* [-0.002,0.062] (0.016)	(0.012) 0.025 [-0.007, 0.057]	(0.012) 0.021 [-0.011,0.053]	(0.012) 0.025 [-0.007, 0.057]
Comment(s) received m1 (1/0)	0.121^{***} 0.120,0.143 0.011				
Pure agree comment(s) received m1 (1/0)	(110.0)	0.066*** [0.054,0.079] (0.006)			
Agree & elaborate comment(s) received m1 (1/0)			0.086^{***} [0.071,0.102] (0.008)		
Disagree & elaborate comment(s) received mI (1/0)				0.053*** [0.040,0.066] (0.007)	
Other comment(s) received m1 (1/0)					-0.076*** [-0.089,-0.063] (0.007)
Like(s) received m1 (1/0)	0.146*** [0.108,0.184] (0.019)	0.191*** [0.154,0.228] (0.019)	0.162*** [0.124,0.200] (0.019)	0.186*** [0.149,0.223] (0.019)	0.196^{+**} 0.160,0.233 (0.019)
Means Dep. Variable Controls	0.847 Yes	0.847 Yes	0.847 Yes	0.847 Yes	0.847 Yes
Year FE Adjusted R2 N	Yes 0.089 12,687	Yes 0.082 12,687	Yes 0.086 12,687	Yes 0.080 12,687	Yes 0.082 12,687

TABLE A4 Comment Classification 2

	Comment Cla	assification 3			
	(1)	(2)	(3)	(4)	(5)
	Persistence	Persistence	Persistence	Persistence	Persistence
High # comment givers (1/0)	0.084	0.079	0.088	0.096	0.076
	[-0.139, 0.308] (0.114)	[-0.146,0.303] (0.114)	[-0.135, 0.312] (0.114)	[-0.128,0.319] (0.114)	[-0.148, 0.299] (0.114)
High # like givers (1/0)	-0.518***	-0.461***	-0.506***	-0.541***	-0.501***
	[-0.862, -0.175] (0.175)	[-0.806,-0.115] (0.176)	[-0.849,-0.163] (0.175)	[-0.886,-0.196] (0.176)	[-0.845,-0.157] (0.176)
Comment(s) received m1 (1/0)	$\begin{array}{c} 1.058^{***} \\ [0.840, 1.275] \\ (0.111) \end{array}$				
Pure agree comment(s) received m1 (1/0)		0.560^{***} [0.440,0.680] (0.061)			
Agree & elaborate comment(s) received m1 (1/0)			0.899*** [0.742,1.056] (0.080)		
Disagree & elaborate comment(s) received m1 (1/0)				0.555^{***} [0.421,0.689] (0.068)	
Other comment(s) received m1 (1/0)					-0.796*** [-0.923,-0.669] (0.065)
Like(s) received m1 (1/0)	1.460*** [1.043,1.876]	1.860*** [1.456,2.264]	1.523*** [1.111,1.936]	1.774^{***} [1.367,2.180]	1.882*** [1.479,2.284]
	(0.212)	(0.700)	(0.211)	(707.0)	(607.0)
Means Dep. Variable	15.643	15.643	15.643	15.643	15.643
Controls	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.137	0.131	0.137	0.132	0.134
Z	12,687	12,687	12,687	12,687	12,687

TABLE A5 omment Classification

	Ag	ge below Median		Υð	te above Median	
	(1)	(2)	(3)	(4)	(5)	(9)
	<i>Comments m</i> 2-7 (1/0)	Likes m2-7 (1/0)	Persistence (1/0)	<i>Comments m</i> 2-7 (1/0)	Likes m2-7 (1/0)	Persistence (1/0)
High # comment givers m1 (1/0)	0.019	-0.030	-0.094	0.039*	-0.000	0.111
1	[-0.018, 0.056]	[-0.069, 0.009]	[-0.381, 0.194]	[-0.006, 0.085]	[-0.048, 0.048]	[-0.308, 0.530]
	(0.019)	(0.020)	(0.147)	(0.023)	(0.024)	(0.214)
High # like givers m1 (1/0)	0.032	0.014	0.014	-0.070^{**}	-0.008	-0.721***
	[-0.034, 0.099]	[-0.054, 0.083]	[-0.537, 0.564]	[-0.124, -0.016]	[-0.064, 0.048]	[-1.219,-0.222]
	(0.034)	(0.035)	(0.281)	(0.027)	(0.029)	(0.254)
Comment(s) received m1 (1/0)	0.238^{***}	0.191^{***}	1.428^{***}	0.265^{***}	0.224^{***}	2.109^{***}
	[0.203, 0.272]	[0.161, 0.222]	[1.093, 1.762]	[0.233, 0.296]	[0.195, 0.252]	[1.764, 2.454]
	(0.018)	(0.016)	(0.171)	(0.016)	(0.015)	(0.176)
Like(s) received m1 (1/0)	0.213^{***}	0.150^{***}	3.247^{***}	0.188^{***}	0.167^{***}	4.113^{***}
	[0.164, 0.262]	[0.110, 0.190]	[2.608, 3.886]	[0.138, 0.238]	[0.127, 0.207]	[3.410, 4.815]
	(0.025)	(0.020)	(0.326)	(0.026)	(0.020)	(0.358)
Means Dep. Variable	0.658	0.847	15.643	0.658	0.847	15.643
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.074	0.043	0.094	0.080	0.052	0.120
Ν	6,061	6,061	6,061	7,266	7,266	7,266
<i>Notes.</i> The dependent variable otherwise. The dependent varia otherwise. The dependent varia otherwise. The dependent varia equals 1 if the number of peers peers giving likes in module 1 if <i>t</i> <i>Like(s) received m1</i> equals 1 if <i>t</i> English as an official language (quiz deadline), cohort size, quiz in parentheses. * $p < 0.10, **p$	in columns (1) and (4) <i>Co</i> able in columns (2) and (ble in columns (3) and (6) giving comments in modu giving comments in modu all above the median of all any likes from peers were (yes/no), $\#$ of same gendei c score in first submitted n c score in first submitted n of 0.05, *** p < 0.01.	mments given $m2^{-7}$ (5) Likes given $m2^{-1}$ Persistence is the m ule 1 is above the mot l cohorts Comment(received and 0 other r, similar age (+/- 2 nodule (m1 in financ	equals 1 if any com <i>m7</i> equals 1 if any amber of completed edian of all cohorts <i>s) received m1</i> equa rwise. Control varia revise. Same countr years), same countr	ments were given in later likes were given in later modules of the entire pro- and 0 otherwise. $High \#$, ls 1 if any comments fro bles include gender (male y & same citizenship pee- se). 95% confidence inter	modules 2 to 5/7 in modules 2 to 5/7 in gram (0-17). <i>High #</i> <i>ike givers m1</i> equal m peers were receiv m peers were receiv fremale), age, living ss, procrastination (vals in brackets. Ro	the program and 0 the program and 0 <i>comment givers m1</i> s 1 if the number of ed and 0 otherwise. g in the US (yes/no), submission hours to bust standard errors

TABLE A6 Robustness 1: Split by Median Age

TABLE A7	stness 2: Split by US Residency
	Robustne

		US residents		Ž	on-US residents	
	(1)	(2)	(3)	(4)	(5)	(9)
	Comments m2-/ (1/0)	Likes m2-/ (1/U)	Persistence (1/0)	Comments m2-/ (1/0)	Likes m2-/ (1/0)	Persistence (1/0)
High # comment givers (1/0)	0.040^{**}	-0.010	-0.156	0.020	-0.027	-0.000
	[0.004, 0.075]	[-0.047, 0.027]	[-0.467, 0.156]	[-0.026, 0.067]	[-0.077, 0.023]	[-0.363, 0.362]
	(0.018)	(0.019)	(0.159)	(0.024)	(0.026)	(0.185)
High # like givers (1/0)	-0.023	0.021	-0.474*	-0.045	-0.014	-0.206
	[-0.077, 0.031]	[-0.035, 0.077]	[-0.983, 0.035]	[-0.109, 0.018]	[-0.080, 0.051]	[-0.709, 0.296]
	(0.027)	(0.029)	(0.260)	(0.032)	(0.033)	(0.256)
Comment(s) received m1 (1/0)	0.247^{***}	0.195^{***}	1.938^{***}	0.259^{***}	0.228^{***}	1.614^{***}
	[0.216, 0.278]	[0.168, 0.223]	[1.602, 2.273]	[0.224, 0.294]	[0.196, 0.259]	[1.266, 1.961]
	(0.016)	(0.014)	(0.171)	(0.018)	(0.016)	(0.177)
Like(s) received m1 (1/0)	0.199^{***}	0.147^{***}	3.844^{***}	0.207^{***}	0.173^{***}	3.322^{***}
	[0.154, 0.245]	[0.110, 0.184]	[3.206, 4.482]	[0.153, 0.260]	[0.129, 0.217]	[2.606, 4.038]
	(0.023)	(0.019)	(0.325)	(0.027)	(0.022)	(0.365)
Means Dep. Variable	0.658	0.847	15.643	0.658	0.847	15.643
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.069	0.039	0.109	0.091	0.064	0.105
Z	7,598	7,598	7,598	5,729	5,729	5,729
Notes. The dependent variable otherwise The dependent va	le in columns (1) and (4) C ariable in columns (2) and	Comments given m2- 1 (5) Likes given m2	7 equals 1 if any con 2-m7 equals 1 if any	mments were given in late	rt modules 2 to 5/7 in modules 2 to 5/7 in	n the program and 0

otherwise. The dependent variable in columns (3) and (6) Persistence is the number of completed modules of the entire program (0-17). High # comment givers m1 equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. High # like givers m1 equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts Comment(s) received m1 equals 1 if any comments from peers were received and 0 otherwise. Like(s) received m1 equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. , progr ά 2,2 * p < 0.10, ** p < 0.05, *** p < 0.01.

		OVID-19 Period		Pre-	COVID-19 Period	
	(1) Comments m2-7 (1/0)	(2) Likes m2-7 (1/0)	(3) Persistence (1/0)	(4) <i>Comments m2-7 (1/0)</i>	(5) Likes m2-7 (1/0)	(6) Persistence (1/0)
(0/1) [m s	0.042**	-0.014	0.031	0.019	0.010	-0.472
	[0.010, 0.074]	[-0.048, 0.019]	[-0.228, 0.290]	[-0.043, 0.082]	[-0.055, 0.075]	[-1.052, 0.107]
	(0.016)	(0.017)	(0.132)	(0.032)	(0.033)	(0.296)
(1/0)	0.000	0.000	0.000	-0.025	-0.013	-0.136
	[0.000, 0.000]	[0.000, 0.000]	[0.000, 0.000]	[-0.080, 0.030]	[-0.069, 0.043]	[-0.627, 0.355]
	()	()	(·)	(0.028)	(0.029)	(0.250)
mI (1/0)	0.265^{***}	0.232^{***}	2.103^{**}	0.244^{***}	0.191^{***}	1.541^{***}
	[0.230, 0.300]	[0.200, 0.263]	[1.721, 2.484]	[0.213, 0.275]	[0.163, 0.219]	[1.228, 1.855]
	(0.018)	(0.016)	(0.195)	(0.016)	(0.014)	(0.160)
1/0)	0.204^{***}	0.166^{***}	4.105^{***}	0.204^{***}	0.157^{***}	3.316^{***}
	[0.148, 0.260]	[0.121, 0.212]	[3.315, 4.896]	[0.159, 0.248]	[0.121, 0.194]	[2.723, 3.910]
	(0.028)	(0.023)	(0.403)	(0.023)	(0.018)	(0.303)
1)	0.658	0.847	15.643	0.658	0.847	15.643
	Yes	Yes	Yes	Yes	Yes	Yes
	Yes	Yes	Yes	Yes	Yes	Yes
	0.063	0.040	0.117	0.081	0.050	0.095
	7,271	7,271	7,271	6,056	6,056	6,056

TABLE A8 Robustness 3: Split by COVID19 Affectedness otherwise. The dependent variable in columns (2) and (5) *Likes given m2-m7* equals 1 if any likes were given in later modules 2 to 5/7 in the program and 0 equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. High # like givers m1 equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts Comment(s) received mI equals 1 if any comments from peers were received and 0 otherwise. Like(s) received m1 equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. otherwise. The dependent variable in columns (3) and (6) *Persistence* is the number of completed modules of the entire program (0-17). *High # comment givers m1*

		Women			Men	
	(1)	(2)	(3)	(4)	(5)	(9) (100
	Comments m2-7 (1/0)	Likes m2-7 (1/0)	Persistence (1/0)	Comments m2-7 (1/0)	Likes m2-7 (1/0)	Persistence (1/0)
High # comment givers m1 (1/0)	0.027	-0.018	-0.171	0.044^{**}	0.001	0.014
	[-0.016, 0.069]	[-0.063, 0.027]	[-0.524, 0.182]	[0.005, 0.083]	[-0.040, 0.041]	[-0.317, 0.345]
	(0.022)	(0.023)	(0.180)	(0.020)	(0.021)	(0.169)
High # like givers m1 (1/0)	-0.022	0.032	-0.331	-0.047*	-0.035	-0.644**
)	[-0.087, 0.043]	[-0.037, 0.100]	[-0.920, 0.259]	[-0.102, 0.009]	[-0.092, 0.021]	[-1.142, -0.146]
	(0.033)	(0.035)	(0.301)	(0.028)	(0.029)	(0.254)
Comment(s) received m1 (1/0)	0.303^{***}	0.253^{***}	2.162^{***}	0.226^{***}	0.190^{***}	1.610^{***}
	[0.264, 0.342]	[0.219, 0.287]	[1.722, 2.601]	[0.197, 0.256]	[0.163, 0.217]	[1.315, 1.905]
	(0.020)	(0.017)	(0.224)	(0.015)	(0.014)	(0.151)
Like(s) received m1 (1/0)	0.197^{***}	0.177^{***}	3.541^{***}	0.211^{***}	0.149^{***}	3.666^{***}
	[0.139, 0.255]	[0.132, 0.223]	[2.729, 4.353]	[0.167, 0.256]	[0.112, 0.185]	[3.067, 4.265]
	(0.029)	(0.023)	(0.414)	(0.023)	(0.019)	(0.306)
Means Dep. Variable	0.658	0.847	15.643	0.658	0.847	15.643
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.087	0.056	0.109	0.071	0.042	0.106
Z	5,188	5,188	5,188	7,841	7,841	7,841
Notes. The dependent variable	in columns (1) and (4) C_{c}	omments given m2-7	equals 1 if any con	ments were given in late	r modules 2 to 5/7 ii	a the program and 0
otherwise. The dependent vari	iable in columns (2) and	(5) Likes given m2-i	m7 equals 1 if any	likes were given in later	modules 2 to 5/7 in	the program and 0
otherwise. The dependent varia	ble in columns (3) and (6)	Persistence is the m	umber of completed	modules of the entire pro	gram (0-17). High #	comment givers m1
equals 1 if the number of peers	giving comments in mod	ule 1 is above the me	edian of all cohorts	and 0 otherwise. High #	like givers mI equal	s 1 if the number of
peers giving likes in module 1	is above the median of al	l cohorts Comment(.	s) received m1 equa	ls 1 if any comments fro	m peers were receiv	ed and 0 otherwise.
Like(s) received m1 equals 1 if	any likes from peers were	received and 0 othe	erwise. Control vari	ables include (male/fema	ıle), age, English as	an official language
(yes/no), residency in the US (y	yes/no), # of same gender	, similar age (+/- 2 y	ears), same country	& same citizenship pee	rs, procrastination (s	submission hours to

quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE A9 Robustness 4: Split by Gender

	(1) Comments given early: m2-3 (1/0)	(2) Comments given early: m2-3 (1/0)	(3) Comments given late: m4-7 (1/0)	(4) Comments given late: m4-7 (1/0)
High # comment givers (1/0)	0.016	0.020	-0.009	-0.004
ingn # comment givers (170)	[-0.016.0.048]	[-0.010.0.050]	[-0.042.0.023]	[-0.035, 0.027]
	(0.016)	(0.015)	(0.017)	(0.016)
High # like givers (1/0)	-0.025	0.003	-0.013	0.010
	[-0.068.0.019]	[-0.030.0.036]	[-0.057.0.031]	[-0.024.0.043]
	(0.022)	(0.017)	(0.023)	(0.017)
Comment(s) received m1 (1/0)	0.209***	0.212***	0.191***	0.193***
	[0.185,0.234]	[0.188,0.237]	[0.168,0.214]	[0.169,0.216]
	(0.013)	(0.013)	(0.012)	(0.012)
Like(s) received m1 (1/0)	0.083***	0.084***	0.083***	0.083***
	[0.045,0.122]	[0.045,0.122]	[0.049,0.117]	[0.049,0.117]
	(0.020)	(0.020)	(0.017)	(0.017)
Means Dep. Variable	0.616	0.616	0.472	0.472
Controls	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
Adjusted R ²	0.095	0.094	0.096	0.095
Ν	12,687	12,687	12,687	12,687

TABLE A10 Robustness 5a: Future Commenting Engagement - split by Early vs. Late Modules

Notes. The dependent variable in columns (1) and (2) *Comments given m2-3* equals 1 if any comments were given in later early modules 2 to 3 in the program and 0 otherwise. The dependent variable in columns (3) and (4) *Comments given m4-7* equals 1 if any comments were given in later modules 4 to 5/7 in the program and 0 otherwise. *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

	(1) Likes given early: m2-3 (1/0)	(2) Likes given early: m2-3 (1/0)	(3) Likes given late: m4-7 (1/0)	(4) Likes given late: m4-7 (1/0)
High # comment givers (1/0)	0.004	0.004	-0.001	0.001
	[-0.020,0.028]	[-0.019,0.027]	[-0.029,0.027]	[-0.025,0.028]
	(0.012)	(0.012)	(0.014)	(0.013)
High # like givers (1/0)	0.019	0.034***	-0.005	0.019
	[-0.015,0.053]	[0.008, 0.060]	[-0.045,0.034]	[-0.010,0.049]
	(0.017)	(0.013)	(0.020)	(0.015)
Comment(s) received m1 (1/0)	0.139***	0.140***	0.156***	0.158***
	[0.116,0.161]	[0.118,0.163]	[0.132,0.180]	[0.134,0.182]
	(0.011)	(0.011)	(0.012)	(0.012)
Like(s) received m1 (1/0)	0.153***	0.154***	0.149***	0.151***
	[0.114,0.191]	[0.115,0.192]	[0.110,0.188]	[0.112,0.191]
	(0.020)	(0.020)	(0.020)	(0.020)
Means Dep. Variable	0.822	0.822	0.741	0.741
Controls	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
Adjusted R2	0.095	0.094	0.104	0.103
Ν	12,687	12,687	12,687	12,687

TABLE A11Robustness 5b: Future Liking Engagement - split by Early vs. Late Modules

Notes. The dependent variable in columns (1) and (2) *Likes given m2-3* equals 1 if any likes were given in later early modules 2 to 3 in the program and 0 otherwise. The dependent variable in columns (3) and (4) *Likes given m4-7* equals 1 if any likes were given in later modules 4 to 5/7 in the program and 0 otherwise. *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

APPENDIX B. FIGURES



FIGURE A1 Interaction frequencies by type and module