Working Paper 21-072

Learning From People Like Me: The Impact of Demographic Diversity on Invisible Knowledge Consumption and Performance in Online Knowledge Exchange Platforms

Laura Rosendahl Huber
Jaqueline N. Lane
Haylee Ham
Karim R. Lakhani

Harvard Business School

Learning From People Like Me: The Impact of Demographic Diversity on Invisible Knowledge Consumption and Performance in Online Knowledge Exchange Platforms

Laura Rosendahl Huber
Rotterdam School of Management

Jacqueline N. Lane
Harvard Business School

Haylee Ham
Mission Lane

Karim R. Lakhani
Harvard Business School

Working Paper 21-072

Funding for this research was provided in part by Harvard Business School.

Learning From People Like Me: The Impact of Demographic Diversity on Invisible Knowledge Consumption and Performance in Online Knowledge Exchange Platforms

Laura Rosendahl Huber[1,4], *Jacqueline N. Lane*[2,4], *Haylee Ham*[3], *& Karim R. Lakhani*[2,4]

[1]Rotterdam School of Management, Erasmus University
[2]Harvard Business School, Harvard University
[3]Mission Lane
[4]Digital, Data, & Design Institute (D^3) Institute at Harvard

Abstract

The proliferation of online knowledge-sharing and online learning platforms with broad accessibility to participants across the globe has instigated a shift in how people acquire knowledge and new skills. These platforms bring together a demographically diverse pool of individuals, creating opportunities for individuals to tap into diverse experiences, skills, and cultures to enhance their learning experiences and improve performance. However, a persistent question is to what extent opportunities to engage and consume content from demographically diverse peers are beneficial to individuals' performance on the platform. To address this question, we study 2.5 million online interactions among 12,000 working professionals enrolled in an online business training program offered by a large elite U.S. University over a six-year period. Our empirical strategy exploits a natural experiment setting in which participants are randomly assigned to cohorts of peers as well as fine-grained data on individuals' knowledge consumption patterns on the course platform. Our results show that being assigned to a cohort with more demographically similar peers causally increases performance and course completion. Then turning to the impact of individuals' knowledge consumption patterns on the platform, we find that the positive effect of demographically similar peers on performance is mediated by individuals' tendency to consume content from individuals who are like them in terms of age, gender, and nationality. Our study makes several important contributions to the literature on online knowledge sharing and has important implications for the effective design of knowledge exchange and learning platforms.

Keywords: online platforms, knowledge sharing, diversity, business courses

1. Introduction

Geographic flexibility in where people live, work, and study (Choudhury, 2022; Gonsalves, 2020; Yang et al., 2022), has fueled a dramatic increase in online platforms for asynchronous knowledge exchange both within and across organizations (Argote, Guo, Park, & Hahl, 2022; Haas, Criscuolo, & George, 2015; Hwang, Singh, & Argote, 2015; Neeley & Leonardi, 2018). In parallel with these developments in the work environment, online and hybrid learning has witnessed tremendous growth in popularity (Baek & Shore, 2020; Eesley & Wu, 2020). Asynchronous knowledge exchange and online learning platforms provide notable advantages by facilitating a wide-ranging, demographically varied community where everyone can participate, regardless of time or location constraints (Faraj, von Krogh, Monteiro, & Lakhani, 2016; Kretschmer, Leiponen, Schilling, & Vasudeva, 2022; McIntyre & Srinivasan, 2017). As these types of platforms are crucial to efficient knowledge exchange and learning in any modern organization (Haas & Hansen, 2007; Mickeler, Khashabi, Kleine, & Kretschmer, 2023), it is important to understand how engagement with content on these platforms shapes individual performance. In this paper, we examine how invisible knowledge consumption—defined as the consumption of digital content without explicit or noticeable interaction from the user—affects an individual's performance on online learning platforms.

Prior research has highlighted the importance of information sharing and advice on knowledge creation (Dahlander, O'Mahony, & Gann, 2016; Jeppesen & Lakhani, 2010) and performance (Chatterji, Delecourt, Hasan, & Koning, 2019; Hasan & Koning, 2019). This research, often focusing on face-to-face interactions, has also emphasized the potential benefits of gaining different perspectives both inside and outside the workplace (Ibarra, 1992; Kleinbaum, Stuart, & Tushman, 2013; Lazarsfeld & Merton, 1954; Reagans, 2011). However, these diverse perspectives may not always be relevant to an individual's task performance; instead, they may create unnecessary distractions. This is particularly evident in online settings, where the proliferation of content generates an unprecedented volume and diversity of information that compete for people's attention (Haas et al., 2015; Hansen & Haas, 2001).

Whereas scholars of online knowledge exchange and transfer have devoted increasing attention to interactions between people in online communities (Faraj et al., 2016; Haas et al., 2015) in terms of *visible* interactions as measured by comments, questions or responses posted on online knowledge sharing platforms (e.g., Hwang et al., 2015; Pu, Liu, Chen, Qiu, & Cheng, 2022; Reagans, Singh, & Krishnan, 2015), relatively less attention has been directed to understanding the implications of individuals' *invisible consumption of content* on online platforms, which play a key role in how knowledge is diffused (Cranefield, Yoong, & Huff, 2015; Lai & Chen, 2014). Most platforms are designed to allow content consumption through viewing or reading others' posts, in a manner that tends to be invisible or anonymous (Preece, Nonnecke, & Andrews, 2004; Sun, Rau, & Ma, 2014). For example, platforms, such as Reddit, Quora, Wikipedia, and YouTube allow users to browse and read content without revealing their identity to the author or other users.

Although many platforms enable people to search for and read online knowledge content without ever leaving a visible trace, we know very little about how this type of behavior is influenced by the demographic composition of the peers encountered on the platform, and its implications on an individual's learning and performance. On one hand, demographic diversity increases the number of available perspectives that an individual can attend to, which may improve their understanding of the content or topic of discussion (Roche, Oettl, & Catalini, 2022; Tortoriello, McEvily, & Krackhardt, 2015). On the other hand, if an individual seeks out a vast array of perspectives from both demographically similar and dissimilar peers, this may increase their cognitive load of processing disparate information inputs that are potentially irrelevant to their needs or understanding. This attentional strategy may be problematic if it comes at the expense of more in-depth exposure to specific concepts or points of view (Piezunka & Dahlander, 2015; Rhee & Leonardi, 2018; Sullivan, 2010). In this paper, we seek to understand the extent to which invisible knowledge consumption patterns with demographically similar or dissimilar peers in online platforms are supportive of online learning and performance.

To address this question, we study the individual task performance and invisible knowledge consumption among 12,000 working professionals enrolled in a demographically diverse, 10-17 week online business training program offered by a large elite U.S. University over a six-year period. The courses are asynchronous and facilitated by pre-recorded videos. A key aspect of the course is the use of discussion forum posts as a mandatory component of the course activities. Although engaging with peers' content invisibly or visibly (i.e., through a comment) is not mandated in the course, most individuals choose to engage invisibly with others' posts on the platform as they proceed through the course. Our data provides us with the unique opportunity to study individuals' invisible knowledge consumption patterns by tracking which posts people read, as well as link these behaviors to their individual course outcomes.

Our empirical strategy identifies the causal effect of demographic diversity in terms of the number of demographically similar peers in the cohort on course performance. To estimate a causal effect, we leverage a natural experiment in which the program administrators randomly assigned participants to two equal-sized cohorts when the course enrollment exceeded five hundred participants for a given instance of the course. The random assignment of participants to cohorts creates idiosyncratic variation in the composition of the cohorts along observable demographic characteristics, such as age, gender, and country of citizenship. To test how diversity in invisible knowledge consumption relates to performance, we have information on individual performance outcomes on the platform in terms of weekly quiz scores and course completion. Complementing the demographic data and course performance data on the participants in each cohort of the online program, we have dyad-level interaction data on the participants' invisible knowledge consumption patterns as measured by the number of outgoing views to content posted by their peers.

Our findings show a positive causal effect of demographic similarity between online peers on individual task performance on the platform. Second, we find that this positive effect is mediated by demographic similarity in invisible knowledge consumption of content posted by demographically similar online peers. Thus, although online knowledge-sharing platforms often exhibit demographic diversity, our findings reveal that focusing one's attention on demographically similar peers is more effective for individual performance outcomes. In our post-hoc analyses, we explore the relationship between the number of demographically similar peers per cohort and online invisible knowledge consumption further. These analyses reveal that as the number of demographically similar people in their cohort increases, individuals attend to *less* content overall, but to *more* content from demographically similar peers. Thus, providing further support for our theoretical argument about more focused invisible knowledge consumption.

We make several important contributions to the literature. First, we contribute to the literature on knowledge diversity and performance (e.g., Jeppesen & Lakhani, 2010; Piezunka & Dahlander, 2015; Reagans & McEvily, 2003) by showing that when it comes to asynchronous online knowledge exchange and online learning having peers with *similar* demographic characteristics is beneficial for performance on these platforms. In particular, we find that the performance effect is driven by the invisible consumption of knowledge or content provided by these demographically similar peers. This means that demographic similarity matters even when people only interact through small snippets of text. As such, similarity in background serves as a useful mental shortcut to guide people to relevant information and avoid information overload.

Second, we contribute to the literature on online knowledge sharing (e.g., Haas et al., 2015; Hwang et al., 2015; Pu et al., 2022) by studying the role of invisible knowledge consumption on individuals' performance outcomes, which for the most part has been relatively less studied compared to *visible* social interactions between online or offline peers. Our results show that in the presence of large amounts of asynchronously available content the anonymous nature of invisible knowledge consumption creates allows people to view and explore content from people from diverse backgrounds. However, our results also indicate that this may come at the expense of more focused and in-depth knowledge accumulation, with negative consequences for performance outcomes on the platform.

Finally, given the prevalence and increased popularity of asynchronous online knowledge exchange and online learning platforms, our findings have important implications for our understanding and the optimal design of such platforms (McIntyre & Srinivasan, 2017; Mickeler et al., 2023). While previous research has often focused on increasing engagement and thereby the available content on knowledge exchange platforms (Afuah, 2013), our results paint a more nuanced picture. Our findings show that the demographic similarity of an individual's knowledge consumption is critical for performance outcomes. This finding has implications for platform design, as it suggests that organizational efforts to bring demographically similar peers together can be an effective way to improving engagement and persistence in online knowledge platforms. These organizational design choices are relatively simple compared to platform changes that tend to require more resources and effort to implement.

2. Theory and Hypotheses

Building on the notion that social interaction is integral to the effectiveness of online knowledge exchange and online learning (Bettinger, Liu, & Loeb, 2016; Faraj et al., 2016), we consider how the degree of demographic diversity in online knowledge exchange platforms shapes individual behaviors and individual task performance. We focus on demographic diversity in terms of age, national, and gender diversity, because these characteristics are increasingly important in a globalized business world as well as the most visible features on online organizational knowledge exchange platforms.[1] Although demographic diversity is prevalent in most contemporary workforces (Ancona & Caldwell, 1992; Cummings, 2004), the extent to which people have the opportunity to engage and interact with demographically diverse peers has increased tremendously due to the recent trend of working and learning remotely (Choudhury, 2022; Choudhury, Foroughi, & Larson, 2021; Yang et al., 2022). Hence, it is crucial to understand how working professionals with diverse demographic backgrounds interact and learn together in a virtual environment.

2.1. Demographic Diversity and Individual Task Performance on Online Knowledge Exchange Platforms

Online knowledge exchange platforms draw people from a variety of backgrounds. There is abundant research on the importance and benefits of demographic heterogeneity and breadth of knowledge to enhance learning capabilities (Reagans & Zuckerman, 2001) and to improve performance (Leiponen & Helfat, 2010; Wagner, Hoisl, & Thoma, 2014). The underlying premise is that people from diverse backgrounds have different experiences and values and that interactions between diverse individuals lead to knowledge transfer and novel insights (Reagans & McEvily, 2003; Tortoriello et al., 2015). Being exposed to diverse sources of information increases the chances for novel insights and reduces the likelihood of receiving redundant information (Dahlander et al., 2016; Jeppesen & Lakhani, 2010; Piezunka & Dahlander, 2015).

[1] This information is available from profile pictures, names and location information that are commonly available on organizational knowledge exchange platforms and online learning environments.

In the context of online knowledge exchange platforms, with a broad range of global participants, demographic diversity is essential to exposure to diverse sources of information. In offline settings, studies indicate that demographically diverse peers are more likely to bring different perspectives, ideas and opinions into the discussions, and have the potential to increase performance (Cummings, 2004; Hasan & Bagde, 2013; Reagans & McEvily, 2003). For example, being exposed to different perspectives and ideas encourages people to think more critically about their own ideas (Hasan & Koning, 2019; Tortoriello et al., 2015). It could also improve their problem-solving skills because they have now seen different ways to approach the same problem which they may be able to apply themselves in the future (Rhee & Leonardi, 2018). Finally, being exposed to information and knowledge provided by people from different cultures and different backgrounds could lead to more openness and understanding of diversity in future interactions, which could benefit their overall performance.

However, diversity along demographic dimensions runs counter to individuals' preferences to affiliate with similar others (Kleinbaum et al., 2013; Reagans, 2011). The extent to which pairs of individuals share a social or demographic characteristic, such as race, gender, age, or nationality is an important predictor of interaction and relationship formation. A large literature on homophily shows that people are attracted to and prefer to form bonds with others who share similar demographic characteristics (Lazarsfeld & Merton, 1954; McPherson, Smith-Lovin, & Cook, 2001), such that contact between similar people occur at a higher frequency than between dissimilar people (McPherson et al., 2001).

In online knowledge exchange platforms, participants may also exhibit a preference to engage with others who share similar demographic characteristics, a phenomenon grounded in the principles of homophily. For instance, age-similar individuals, who are likely to be at comparable life stages and to have undergone similar socialization experiences, may be more motivated to exchange information with each other (Kilduff, Angelmar, & Mehra, 2000; Tsui, Egan, & O'Reilly III, 1992). This is particularly relevant in platforms catering to working professionals, where hierarchical differences may emerge between younger and older members. While older peers might serve as mentors, relationships between junior and senior members can often be more hierarchical, potentially impeding effective knowledge exchange (Reagans, 2011; Smith, McPherson, & Smith-Lovin, 2014). Furthermore, participants may gravitate towards content created by individuals of a similar age, even when their own engagement with the content is not visible to others. This preference may stem from the belief that content from peers of a similar age is likely to be more pertinent to their own life circumstances and career stage, and hence, more performance enhancing.

Gender also plays a significant role in these online interactions. Research indicates a prevalent trend of gender homophily in workplace settings, where individuals, including women and other minorities, tend to seek out and engage with others like themselves (Ibarra, 1992, 1997; Khattab, van Knippenberg, Pieterse, & Hernandez, 2020). Furthermore, in online settings, it could be that being female is one of the most salient characteristics. Prior research has shown that being a minority group sharing a demographic characteristic could make this characteristic more salient to an individual within the larger group (Kleinbaum et al., 2013; Reagans, 2011). For example, if the topics and content covered on the platform are perceived as belonging to a more masculine domain, this could increase preferences for interactions with other women, draw attention to, and foster confidence through engagement with this type of content (Brown & Diekman, 2010; Carli, Alawa, Lee, Zhao, & Kim, 2016). Finally, gender differences in written communication style (DeJesus, Umscheid, & Gelman, 2021; Exley & Kessler, 2022), could make content by same-gender peers more recognizable and thereby more useful for task performance.

Lastly, national cultures is another influential factor, as it shapes values, orientations and priorities—which guide the meanings that people attach to aspects of the world around themselves, such as social norms and appropriate behaviors (Gelfand et al., 2011), their underlying assumptions (Earley & Gibson, 2002) and interpretations of information and inputs (Dahlin, Weingart, & Hinds, 2005). Hence, people with the same national origin also tend to be attracted to one another because they are more likely to share values, beliefs, orientations, and attitudes (Earley & Mosakowski, 2000; Earley & Gibson, 2002). That said, national culture not only shapes how people think but also how they communicate. People from the same national background are more likely to use similar terms, symbols, and stylistic cues in their language (Dahlin et al., 2005; Hofstede, 1994). For example, national cultures differ in their use of nonverbal and contextual cues in communicating and interpreting messages. These cross-national communication difficulties may be amplified in online knowledge exchange platforms, because the written nature of communication collapses verbal and nonverbal cues and facial gestures into text—which can make it even harder for people from dissimilar national cultures to communicate and develop shared meanings (Cramton, 2001). These difficulties associated with communicating across national boundaries may limit the amount and/or the effectiveness of asynchronous engagement and information exchange. In other words, content from individuals with a similar national background is likely to be easier for people to understand and interpret, as it is framed within a familiar context that may reflect shared values, experiences, and perspectives.

Taken together, these arguments suggest that shared understanding is more likely to form among demographically similar peers, due to the commonalities in their experiences and socialization. These shared values are important, because there is a need for common understanding among peers as a foundation to learn and exchange knowledge - i.e., peers need to "speak the same language" (Cramton, 2001; Fiol, 1994; Neeley, Hinds, & Cramton, 2012). Although demographically dissimilar peers may bring greater diversity in perspectives and ideas to the discussions, and hence increase opportunities for learning, diverse platform members may not be sufficiently motivated to learn from one another due to their salient demographic differences (Dahlin et al., 2005). These differences may become even more salient in virtual settings where all interactions are reduced to written communications, thereby suppressing verbal (e.g., tone, emotion) and non-verbal cues (e.g., eye gaze, body movements, gestures, facial expressions) that help with relationship formation and consensus building (Gibson & Gibbs, 2006; Wiesenfeld, Raghuram, & Garud, 1999). Accordingly, we predict that demographic similarity in online knowledge exchange platforms is beneficial for individual task performance. Thus, we hypothesize:

Hypothesis 1 (H1). Increasing demographic similarity with respect to age (H1a), gender (H1b) and nationality (H1c) in an online knowledge exchange platform has a positive effect on individual task performance.

2.2. Demographic Diversity, Invisible Knowledge Consumption, and Attention Allocation

The arguments presented in the previous section about the benefits of demographically similar peers in online knowledge exchange platforms pertain to commonalities in terms of values, knowledge, and written expression. Due to the asynchronous nature of online knowledge exchange platforms, an essential way that commonalities between peers are conveyed is through individuals' consumption of online content, often via discussion posts, which are stored and retrievable at any point in time (Treem, Leonardi, & van den Hooff, 2020). The prevalence of posted content that remains persistent on the platform long after the initial act of communication (Leonardi, 2014; Treem & Leonardi, 2013) creates opportunities for invisible knowledge consumption with the content posted by others on the platform. That said, individuals can choose how they attend to the knowledge posted by their peers, choosing a divided or focused attentional allocation strategy (Kahneman, 1973), with respect to engaging with a greater number of demographically dissimilar or demographically similar peers. As individuals are constrained in their time, and the attention they can pay to any given piece of information is limited, they need to decide which information to read and what to ignore (Haas et al., 2015; Lang, 2000; Ocasio, 2011).

On one hand, individuals may choose to divide their attention to consume knowledge from different demographic groups. Individuals may be inclined to deploy a divided attentional strategy particularly when the platform facilitates *invisible knowledge consumption* of their peers' content. Invisible knowledge consumption refers to the anonymous actions people undertake on an online platform, such as actively browsing through, reading, and learning from existing posts or threads on a certain topic. People's motivations for such *invisible* engagement with their peers are likely to be conceptually distinct from *visible* engagement, such as commenting, posing, or responding to questions, which tends to be governed by external incentives and social considerations, such as reputation costs, self-presentation concerns and social motivations (Argote et al., 2022; Hwang et al., 2015; Mickeler et al., 2023). When such constraints are absent, the anonymized nature of invisible knowledge consumption creates opportunities for individuals to gain access to more demographically diverse individuals, without needing to visibly expose themselves to others, and cross intergroup boundaries.

Although engaging with demographically dissimilar peers increases people's exposure to alternative perspectives, it also means that learners need to symmetrize their attentional resources across many demographic groups (Rhee & Leonardi, 2018), whose members may possess differentiated viewpoints that are potentially in conflict with one another. The act of processing and evaluating online content or information posted by individuals with different backgrounds has been shown to be more cognitively demanding (Haas et al., 2015; Wasko & Faraj, 2005). Furthermore, although demographic diversity in invisible knowledge consumption allows individuals to access a broader range of stimuli, it comes at the expense of in-depth processing and repeated exposure to a specific concept or point of view (Piezunka & Dahlander, 2015; Rhee & Leonardi, 2018; Sullivan, 2010). Another downside of invisible knowledge consumption with demographically dissimilar peers is that it can lead to information overload, due to the exposure to an array of disparate and potentially disjointed sources of information.

On the other hand, individuals may deploy a focused attention strategy, when choosing to engage invisibly with more demographically similar peers. A focused attention strategy involves the cognitive processing of fewer information sources (Argote & Levine, 2020; Kahneman, 1973; Ocasio, 1997). A focused attention strategy with respect to engaging with demographically similar peers can have two important implications on an individual's task performance. First, an individual may be more likely to avoid information overload by focusing their attention on processing the informational inputs of demographically similar peers. This could reduce the amount of irrelevant information that is cognitively demanding, but not necessarily conducive to individual task performance. Second, a focused attention strategy may improve an individual's ability to absorb new knowledge. As people are more likely to learn from one another when they share some common knowledge and experiences together (Bower & Hilgard, 1981; Carlile, 2004), this suggests that individuals may be cognitively limited in absorbing or processing information from demographic dissimilar peers, due to their lack of knowledge overlap (Piezunka & Dahlander, 2015; Sullivan, 2010; Wasko & Faraj, 2005).

As engaging with demographically diverse peers is likely to increase the amount of information from disparate sources, with less overlapping knowledge (Page, 2019), a divided attention allocation strategy could negatively affect their performance. In contrast, by focusing on demographically similar peers, an individual may gain exposure to similar information multiple times, which may increase their likelihood of absorbing it into their existing knowledge base. Taken together, these arguments suggest that the positive effect of demographic similarity on individual task performance is mediated by an individual's invisible consumption of knowledge from a greater number of demographically similar peers. We, therefore, formulate the following second hypothesis:

H2: The positive effect of demographic similarity with respect to age, gender, and nationality on individual task performance in an online knowledge exchange platform is mediated by the invisible knowledge consumption of content posted by demographically similar peers.

3. Research Setting

To address our research questions and test our hypotheses, we investigate an online business fundamentals program offered by an elite U.S. business school over a nearly six-year period from January 2017 to September 2022. The program consists of three courses (economics for managers, business analytics and financial accounting) designed to offer participants an understanding of business concepts to contribute to business discussions and decision-making. The three courses are taken simultaneously over 10-17 weeks and are typically offered 11 times per year (other than December). Participants enroll for a variety of reasons, such as to improve their job performance, enhance their career opportunities, and prepare for an MBA program. The tuition fee of this program is around USD 2,500, and approximately one-third of enrollees receive at least some corporate tuition reimbursement. Each offering of the program has roughly 400-450 students (Mean = 442 SD = 113, N = 73, median = 426). If the enrollment size exceeds 500 participants, then the participants in the course offering (called a "wave") are randomly split into two or more separate cohorts. If there are five or more participants from the same country, participants are block randomized by country of *residence* into these cohorts. Otherwise, students from the same country are placed in the same cohort. The random assignment of participants to different cohorts within the same course offering, provides us with an identification strategy to test the causal impact of demographic similarity in cohort composition on the participants' course performance and their invisible knowledge consumption patterns. We focus our analysis on the 18-course waves and 38 cohorts for which enrollment exceeded the target cohort size and participants were randomly assigned to two or more cohorts.

3.1. Program Details

The program is entirely online and asynchronous. It is delivered through a propriety course platform that features videos of faculty and guest business experts, business case discussions, "cold calls" to keep participants engaged, and interactive course elements. It is self-paced with weekly deadlines and requires a weekly time investment of between 8-15 hours over 10-17 weeks. There are a total of 17 lesson plans called "modules" over the duration of the program. Each module has a quiz that needs to be completed by the due date. Each course module is unlocked at a fixed time, which synchronizes when participants can start working on a module. Within a module, some content is locked/not visible until the participant completes the previous content, meaning that one can look back at previous work, but cannot skip ahead. To complete the course, participants must take a three-hour final exam in person at one of the partnering testing facilities around the globe. Upon passing the program, students receive a certificate that they can post on their LinkedIn profiles.

Peer learning via interactive course elements on the platform is an important aspect of the program. Participants are required to make discussion posts on the learning elements and are incentivized via extra credit[2] to comment on their peers' discussion posts.

3.2. Invisible Knowledge Consumption on the Platform

We focus on the participants' engagement with the interactive course elements, which allows us to investigate their invisible knowledge consumption of their peers' posted content. Figure 1 shows an example of the mosaic of peer responses to a typical course element that participants would see in the Economics course of the program. The platform randomly assigns each participant twelve responses to view.[3] This feature ensures that the distribution of views remains consistent across cohorts, irrespective of their size of enrollment. Each mosaic includes a picture of the learner, from which their gender, age, and ethnicity can be inferred. A participant can click on a mosaic to *view* its contents and decide whether to type a *comment* in response to the post. Whereas a participant's view of another learner's post is

[2] Participants can pass the program with honors or high honors, and active participation on the platform is one way to earn extra credit that can result in passing with honors or high honors.
[3] Although participants are randomly assigned 12 shared reflections, there is a search bar that participants can use to find additional posts. Moreover, the platform does not record which mosaics a participant was randomly assigned.

invisible and anonymous to others, a comment is visible and identifiable to everyone else on the platform. Hence, a view corresponds to our theoretical construct of *invisible knowledge consumption*.

4. Data and Variables

4.1. Data

Our complete dataset covers a nearly six-year period between January 2017 - September 2022 and includes 73 cohorts and 24,252 learners. Of these 73 cohorts, our empirical strategy leverages 38 cohorts, corresponding to 43% of the data (or 11,861 learners) for which the target enrollment size exceeded 500 students. For these cohorts, as described in Section 3, the participants were block randomized by country of *residence* into two equal-sized cohorts (other than May 2020 when participants were block randomized into four cohorts; see Table A1 for the years and months corresponding to multiple cohorts).

Our dataset consists of three sources of data on each participant. First, we have demographic data on each participant (such as gender, age, and country of citizenship) collected from their application to the program. Second, we have performance data on each participant, which consists of their quiz score for each module and course in the program, for a total of 17 quiz scores. Accompanying the quiz score data, we also have each student's course completion status, which indicates whether they passed the program. Third, we have timestamped communication data on the number of views given by each participant i in cohort c to learner j in cohort c.

To test H1, we construct an individual level dataset to investigate how the demographic characteristics of the cohort is causally related to course performance. Recall that since individuals are randomly assigned to cohorts, the demographic composition of each cohort is exogenously determined. Second, we test H2 by estimating to what extent this effect is mediated by people's invisible knowledge consumption (i.e., more or less demographic similarity in their viewing behaviors).

4.2. Main Analysis (Program Performance)

4.2.1. Dependent Variables

We use participant i's *Average quiz score* across the 17 modules in the program as the main measure of performance in the program. As an alternative measure, we also measure whether a learner i completed the course using the dependent variable, *Program completion.*

4.2.2. Independent Variables

We measure demographic similarity in cohort composition using three sources of homophily that have been found to affect social interactions and relationship formation: age, gender, and country (Ibarra, 1997; Kleinbaum et al., 2013; Lazarsfeld & Merton, 1954; McPherson et al., 2001; Reagans, 2011). We measure *Similar age peers* as the number of similar-age peers in the cohort. More specifically, for each individual we calculate the number of peers in their cohort that are similar in age which we define being within a +/- 2 year range. We also use several alternative measures of age similarity, such as the number of peers that are exactly the same age and the number of peers within a +/- 5 year range. We measure *Same gender peers* as the number of peers in the cohort that have the same self-reported gender as based on their application form. Lastly, we measure *Same country peers* as the total number of peers that have the same self-reported country of primary citizenship. We use the participant's country of citizenship rather than their country of residence to investigate demographic similarity in nationality, which is more closely related to their culture, background, and experience.

4.2.3. Mediator Variables

We capture demographic similarity of viewing behavior at the individual level by computing participant i's total number of outgoing views to demographically similar peers. We create three measures of demographic similarity in viewing behavior, corresponding to the number of participant i's total views given to peers of a similar age (±2 years), same gender, and same country of primary citizenship. Our resulting measures are as follows: $\#\, similar\ age\ views\ (\pm 2\ years)$; $\#\, same\ gender\ views$; $\#\, same\ country\ views\ (citizenship)$.

We also examine an alternative threshold of views to similar aged peers, namely if peers are exactly the same age and within ± 5 years of each other. Lastly, as robustness, we create an alternative mediation measure corresponding to the *proportion* of similar views, which is measured as participant i's total views given to demographically similar peers divided by the total number of outgoing views to all peers during the program.

4.2.4. Controls

We add a number of participant-level variables as controls. These include a participant's gender, age and country of citizenship. For age we use the self-reported age of the individual at the start of the program. Based on this, we create a categorical variable dividing participants into three equal-sized groups: age 18 to 25 years, age 26 to 31 years, and age >31 years. For country of citizenship, we create a categorical variable with the following categories: United States (45.12%), European Union (7.75%), Developed Countries (excl. USA and EU)[4] (8.63%), BRICS (Brazil, Russia, India, China and South Africa) (18.84%), and Other (19.67%). Furthermore, we add controls for whether the participant is a *manager* (dummy variable equal to 1 if the participant is a mid-level manager or above, and 0 otherwise) and whether the participant is from an *English-speaking* country for which English is a *de jure* official language based on their country of primary citizenship (dummy variable equal to 1 if their primary country of citizenship is English-speaking and 0 otherwise). We further add *invisible knowledge consumption*, which is measured by the sum of all outgoing views in all 17 modules as an additional control in some of our analyses. Lastly, we use year-month dummies to account for whether the participants were enrolled in the same course offering (but randomly assigned to different cohorts) as well as cohort size to account for the difference in total number of students between the cohorts.

4.2.5. Empirical Approach

To estimate the causal effect of demographic cohort composition on course performance, we use the number of demographically similar peers who were exogenously assigned to the same cohort. Due to the block random assignment of participants by country of residence into cohorts (among course waves exceeding 500 participants), the number of similar-aged peers and same-gender peers are exogenously determined, while the number of same-country of residence peers in each cohort is placebo test for participants whose country of citizenship overlaps with their country of residence. To understand how variations in demographic similarity affects individual course performance (H1), we estimate several versions of the following linear regression model:

$$Average\ Quiz\ Score_{ict} = \beta_0 + \beta_1 \#\ of\ similar\ peers_{ict} + \delta X_i + \theta_c + \xi_t + \varepsilon_{ict}. \quad (1)$$

We first estimate the effect of each demographic characteristics (age (*H1a*), gender (*H1b*) and nationality (country of citizenship) (*H1c*)) separately, before including all three in the same model and

[4] This category contains the following countries: Canada, United Kingdom, New Zealand, Australia, Japan and Israel.

subsequently adding additional control variables. In Equation (1), X_i corresponds to participant covariates, θ_c represents the control for cohort size, and ξ_t corresponds to year-month fixed effects. As the number of similar peers in a cohort is exogenously determined, this means that the estimate for β_1 can be interpreted causally.

To test if the positive performance effect of being exposed to a larger number of demographically similar peers on an online knowledge exchange platform is mediated by the invisible knowledge consumption of content from demographically similar peers (Hypothesis 2), we use the Sobel-Goodman test of mediation. To this end, we estimate the following equations consecutively:

$$\#\ of\ views\ to\ similar\ peers_{ict} = \beta_0 + \beta_2(\#\ of\ similar\ peers_{ict}) + \delta X_i + \theta_c + \xi_t + \varepsilon_{ict} \quad (2)$$

$$Average\ Quiz\ Score_{ict} = \gamma_0 + \gamma_1(\#\ of\ views\ to\ similar\ peers_{ict}) +$$

$$\gamma_2(\#\ of\ similar\ peers_{ict}) + \delta X_i + \theta_c + \xi_t + \varepsilon_{ict}. \quad (3)$$

We estimate both equations (2 and 3) for each of the three mediators, i.e., for each measure of demographic similarity (age, gender, country) separately. Equation (2) regresses each mediator, i.e., total number of views to similar aged peers, same gender peers, and same country peers, on the number of demographically similar peers in their cohort *c* at time *t* in the corresponding dimension (i.e., age, gender and country of citizenship respectively). In equation (3), the dependent variable (i.e., Average quiz score) is regressed on the independent variable (number of demographically similar peers) and the mediator, again for each of the mediator separately. In both equations the same control variables are included as in Equation (1), with the exception of *invisible knowledge consumption*, due to its high correlation with the mediators. If a mediation effect exists, a significant β_2 would indicate there is an effect of the independent variable on the mediator, whereas γ_1 is expected to be significant if there is an indirect effect through/via the mediator on the dependent variable.

5. Results

We begin in Table 1 by showing the covariate balance checks for the 18 course waves for which participants were exogenously assigned into two or more cohorts (N = 11,861). The two tailed t-tests show that there are no observable differences between these 38 cohorts. Table 2 shows a correlation matrix and summary statistics of the variables used in the main analyses and indicates that there is a low degree of correlation between total views and cohort size due to the platform's random assignment of posts (i.e., shared reflections) to participants. This confirms that distribution of outgoing views remains consistent across cohorts, irrespective of their size of enrollment. We report our results in subsections: we first test Hypotheses H1a-H1c (demographic similarity and course performance) in Section 5.1, then turn to testing the mediation effect of invisible knowledge consumption (Hypothesis H2) in Section 5.2, and finally reporting the results of our post-hoc exploratory analysis in Section 5.3.

[Tables 1 and 2 about here]

5.1. Results: Demographic Similarity and Course Performance (H1)

Hypothesis 1 theorized that demographic similarity among cohort peers would have a positive effect on individual course performance. Table 3 presents the OLS regression results for the average quiz score on the three demographic similarity measures we theorize: age (H1a), gender (H1b) and nationality (H1c). We begin with the simplest model, which includes the main variables of interest, *# of similar age peers* (Model 1), *# of same gender peers* (Model 2) and *# of same country peers* (Model 3). The results show that having 10 more similar-aged or same gender peers increases average quiz score by 0.4 (Model 1: 0.0422, $p < 0.001$) and 0.3 points (Model 2: 0.0300, $p < 0.001$) respectively, whereas having more peers with the same country of citizenship seems to have a marginally significant negative effect on course performance (Model 3: -0.0049, $p < 0.05$). Recall that because the number of similar peers in a cohort is exogenously determined, the estimated coefficients can be interpreted as causal relationships. Moreover, Model 4 includes all three demographic similarity measures and Model 5 adds individual level controls, including the participant's total invisible knowledge consumption, age, gender, country of citizenship, as well as whether an individual has a managerial role and if English is an official language in their country of primary citizenship. Model 5 shows that once individual level controls are included the size of the coefficients for similar age (0.0240, $p < 0.001$) and same gender (0.0164, $p < 0.001$) remain significant but decrease slightly, and the effect for same country peers is not significant (0.0040, *ns*). Model 6 adds year-month fixed effects and controls for cohort size and shows that the coefficients for all three demographic similarity measures remain stable and significant for both age and gender (Model 6: 0.0340, $p < 0.001$; 0.0252, $p < 0.01$, respectively). Hence, we find support for H1a and H1b but not for H1c.

[Tables 3 about here]

5.2. The Mediating Effect of Invisible Knowledge Consumption (H2)

Next, we test H2, which theorized that the positive effect of having more demographically similar peers in a cohort on performance is mediated by individuals' invisible knowledge consumption patterns with more demographically similar peers. The results are presented in Table 4. As can be seen in Models 1-3, the number of demographically similar peers in the cohort has a positive and significant impact on the invisible knowledge consumption of content by these individuals as measured by the total number of views to content by peers that share that demographic characteristic (Similar age: 1.132, $p < 0.001$; Same gender: 1.059, $p < 0.001$; Same country: 0.710, $p < 0.001$). In Models 4-6, the dependent variable (Average quiz score) is regressed on both the mediator as well as the independent variable. The results show that once the mediators are included, the estimated coefficients for the independent variable, i.e., number of demographically similar peers, are no longer significant for similar age (-0.00375, *ns*) and same country (-0.0104, *ns*), and negative and significant for same gender (-0.0177, $p < 0.05$). At the same time, the estimated coefficients on the mediators are positive and significant for all three mediators, i.e., # similar age views (0.0351, $p < 0.001$), # same gender views (0.0172, $p < 0.001$) and # same country views (0.0288, $p < 0.001$). To formally test the mediation effect, we calculate the indirect effect of each of the three mediators using the product of coefficient approach (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). We then use bootstrapped standard errors and bias-corrected as well as percentile confidence intervals to test if the indirect effect is indeed significant (Preacher & Hayes, 2008; Zhao, Lynch, & Chen, 2010). The results are presented in Table 5 and show that the indirect effect for all three mediators is significant at the 5% significance level, thus providing support for H2.[5]

[Tables 4 and 5 about here]

5.3. Post-hoc Analysis: Peer Similarity and Invisible Knowledge Consumption

[5] These results are also robust to the alternative dependent variable, *Course completion*, as well as our alternative mediation measure corresponding to the *proportion* of similar views. For our alternative mediator, due to a lower correlation between the three mediators, we are also able to use a seemingly unrelated regressions (SURs) approach and test all mediators in one model (Preacher & Hayes, 2008). The results confirm the findings presented above and are available from the authors upon request.

To gain a better understanding of how demographic peer similarity in cohort composition influences invisible knowledge consumption, we conduct two types of post-hoc analyses. First, we want to explore if total invisible knowledge consumption (i.e., the total number of outgoing views to other people's content) changes if there are more or less demographically similar peers in the cohort. We estimate, for each of our demographic characteristics separately, the differential effect of peer similarity on the total # of outgoing views (Table 6, Models 1-3). The results of these analyses show that for older (age > 25 years) and female course participants their invisible knowledge consumption *decreases* as the number of similar age peers and same gender peers in the cohort increases. We do not observe any clear patterns in knowledge consumption by country of citizenship. Second, we want to see if this decrease in invisible knowledge consumption is indeed related to a more focused attention strategy, as argued in Section 2.2. To this end, we estimate the differential effect of cohort peer similarity on the number of views to demographically similar peers, conditional on the total number of outgoing views (Table 6, Models 4-6). These results show that, *conditional on total invisible knowledge consumption*, increasing the number of demographically similar peers in the cohort leads to an increase in the knowledge consumption of content by demographically similar peers. An effect that is particularly strong for older course participants (age > 25 years). Figure 2 shows the corresponding marginsplots by age: Figure 2a) clearly shows the decrease in total # of outgoing views for the two older age groups as the number of similar age peers increases, while 2b) shows the opposite trend for these two age groups when it comes to the number of views to similar age peers. Together we take this as suggestive evidence for a more focused attention strategy by older learners on the platform (i.e., fewer overall views, but more targeted at similar peers).

[Table 6 and Figure 2 about here]

6. Discussion and Conclusion

The proliferation of online knowledge-sharing and online learning platforms with broad accessibility to participants across the globe has instigated a shift in how people acquire knowledge and new skills (Bettinger, Fox, Loeb, & Taylor, 2015; Haas et al., 2015). These platforms bring together a demographically diverse pool of individuals which creates the potential for individuals to tap into a broad and diverse set of experiences, skills, and cultures to enhance their learning experiences and improve performance. Moreover, when it comes to asynchronous online knowledge exchange, content that people post remains visible and retrievable long after the initial communication took place, leading to an unprecedented volume and unprecedented diversity of available information. This new mode of knowledge exchange and learning can be enriching, but also creates challenges to finding relevant information and avoid cognitive overload. We aim to understand how the demographic composition of peers that people encounter on online knowledge exchange platforms and people's engagement with content posted by these peers causally affects performance on such platforms.

To understand the extent to which being exposed to demographically (dis)similar others in an online knowledge exchange is beneficial for performance, we highlight the importance of *invisible knowledge consumption*, which refers to the anonymous actions people undertake on an online platform, such as reading and learning from existing posts or threads. Much of the prior literature has mainly focused on visible actions taken on online knowledge exchange platforms, such as posting questions and responses (Argote et al., 2022; Hwang et al., 2015). Instead, we focus on *invisible* behavior, because much of what people do online consists of invisibly reading and learning from online available content, yet we know very little about how this behavior may affect knowledge flows and performance (Cranefield et al., 2015; Lai & Chen, 2014). We argue that while the anonymous nature of invisible knowledge consumption offers opportunities to access a broad range of perspectives and stimuli, it may come at the expense of more in-depth exposure to specific concepts of points of view (Piezunka & Dahlander, 2015; Rhee & Leonardi, 2018; Sullivan, 2010).

Our results show that increasing demographic similarity in cohort composition, in terms of age and gender has a positive impact on performance, while we do not find such an effect for similarity in nationality. Moreover, we find that the effect of similarity on performance for all three demographic characteristics is mediated by individuals' tendencies to attend to the content of their demographically similar peers. The findings from our post-hoc analyses provide further support for this notion. The results from these analyses showed that as the as the demographic composition of the cohort shifts towards more similar peers, especially older learners and females become more focused in their viewing behavior. Taken together we find that it is beneficial to expose people to more demographically similar peers, to allow them to focus their attention and to consume more content that was provided by these similar peers.

With these findings we contribute to the existing literature in several ways. First, we advance the literature on knowledge diversity and performance (Argote & Miron-Spektor, 2011; Jeppesen & Lakhani, 2010; Reagans & Zuckerman, 2001). Despite the potential benefits of diversity in perspectives and ideas for knowledge creation and learning, we show that, in the context of asynchronous online knowledge exchange, similarity in terms of demographic characteristics between participants on the platform leads to improved performance outcomes. Previous research has already demonstrated the importance of common ground for successful online knowledge transfer (Wasko & Faraj, 2005). Our results indeed indicate that invisible consumption of content posted by peers with similar demographic characteristics plays an important role in that regard.

Our study also contributes to the literature on online knowledge sharing by highlighting the importance of *invisible knowledge consumption* (Haas et al., 2015; Hwang et al., 2015; Mickeler et al., 2023). Our results show that this type of engagement plays an important role in online knowledge exchange yet seems to be different from its more frequently studied counterpart, i.e., visible knowledge contributions. This is particularly relevant since prior research has shown that visible knowledge contributions in online communities is sparse, i.e., only a small fraction of people ever comment, which has been argued to limit the efficacy of such online platforms (Baek & Shore, 2020; Faraj et al., 2016; Wasko & Faraj, 2005).

Finally, our paper also has important implications for the literature on online learning and the effective design of online learning platforms (Bettinger et al., 2016; Eesley & Wu, 2020; Steffens, 2015). Numerous studies have shown that sustaining participant engagement is a major point of weakness for online learning platforms, potentially due to the difficulty in establishing interpersonal bonds among large cohorts of learners with diverse motivations and backgrounds (Baek & Shore, 2020; Breslow et al., 2013; Zhang, Allon, & Van Mieghem, 2017). Our study is among the first to investigate how demographic similarity influences invisible knowledge consumption, and how these interactions contribute to course completion and performance. The results from our study indicate that in online learning settings, having content to engage with invisibly may be more important than having the option to ask questions or solicit peer help. Thus, to improve online learning outcomes, a potentially effective platform design strategy could be to make reflections and active participation mandatory. This will ensure that there is enough available content for any person to invisibly engage with to improve their learning and course performance.

As with any study, our paper is invariably subject to limitations that open-up several interesting avenues for future research. First, we study a specific course, where people pay a substantial tuition fee and are therefore more likely to be engaged and interested in reading other people's posts. Hence, one important open question is to understand if our findings also translate to other online knowledge exchange platforms or online learning programs that are free of charge and that suffer from lower engagement levels and lower completion rates. Second, our setting benefits from the unique feature that all participants are required to write reflections at regular intervals, leading to enough content for others to invisibly engage with. This feature is not representative for public online knowledge exchange platforms, such as Quora or Stack Overflow, where the creation of relevant content relies on voluntary contributions of the community members. This means that content is likely to depend on a variety of factors, such as the expertise of the community members and the topic or content of the prompt or question. This said, our study shows that it is important to consider visible and invisible online engagement in tandem. For future research it would be important to investigate the optimal balance between these two and to understand how much visible content is required for invisible engagement to be effective. Furthermore, we note that focusing on social attributes as cues for information relevance might only be an effective strategy in the short term and in a setting where the objectives are very clearly defined, i.e., passing the next quiz and ultimately the course. One drawback of using demographic similarity as a heuristic for engaging with others is that it may disadvantage underrepresented groups at the periphery who have fewer peers to engage with. This suggests that a fruitful avenue for future work is to investigate how and why demographic similarity drives short-term performance and to identify other shared attributes that could foster similarity and improved engagement between learners. Finally, although we found that demographic similarity improved short-term outcomes, we do not have information on long(er) term benefits of demographic diversity in invisible engagement. An important step for future research is to investigate whether demographic diversity in invisible engagement, could lead to a positive impact in the long term. One could imagine that the concepts learned in the course, as well as being exposed to different perspectives, examples and ideas, are likely to have long run pay-offs (e.g., in terms of critical thinking, and career opportunities). Indeed, emerging evidence from sectoral training programs is beginning to show that participation in skills training programs leads to career and earnings benefits over the long-term (Katz, Roth, Hendra, & Schaberg, 2022). It is just as critical to understand how participation in online business course training affects long-term career outcomes.

References

Afuah, A. (2013). Are network effects really all about size? The role of structure and conduct. *Strategic Management Journal*, *34*, 257–273.

Ancona, D. G., & Caldwell, D. F. (1992). Demography and design: Predictors of new product team performance. *Organization Science*, *3*, 321–341.

Argote, L., Guo, J., Park, S.-S., & Hahl, O. (2022). The Mechanisms and Components of Knowledge Transfer: The Virtual Special Issue on Knowledge Transfer Within Organizations. *Organization Science*, *33*, 1232–1249.

Argote, L., & Levine, J. M. (2020). *The Oxford Handbook of Group and Organizational Learning*. Oxford University Press.

Argote, L., & Miron-Spektor, E. (2011). Organizational Learning: From Experience to Knowledge. *Organization Science*, *22*, 1123–1137.

Baek, J., & Shore, J. (2020). Forum Size and Content Contribution per Person: A Field Experiment. *Management Science*, *66*, 5906–5924.

Bettinger, E., Fox, L., Loeb, S., & Taylor, E. (2015). Changing distributions: How online college classes alter student and professor performance. *American Economic Review*.

Bettinger, E., Liu, J., & Loeb, S. (2016). Connections Matter: How Interactive Peers Affect Students in Online College Courses. *Journal of Policy Analysis and Management*, *35*, 932–954.

Bower, G. H., & Hilgard, E. R. (1981). *Theories of learning*. Prentice-Hall.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying Learning in the Worldwide Classroom Research into edX's First MOOC. *Research & Practice in Assessment*, *8*, 13–25.

Brown, E. R., & Diekman, A. B. (2010). What Will I Be? Exploring Gender Differences in Near and Distant Possible Selves. *Sex Roles*, *63*, 568–579.

Carli, L. L., Alawa, L., Lee, Y., Zhao, B., & Kim, E. (2016). Stereotypes About Gender and Science: Women ≠ Scientists. *Psychology of Women Quarterly*, *40*, 244–260.

Carlile, P. R. (2004). Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organization Science*, *15*, 555–568.

Chatterji, A., Delecourt, S., Hasan, S., & Koning, R. (2019). When does advice impact startup performance? *Strategic Management Journal*, *40*, 331–356.

Choudhury, P. (Raj). (2022). Geographic Mobility, Immobility, and Geographic Flexibility: A Review and Agenda for Research on the Changing Geography of Work. *Academy of Management Annals*, *16*, 258–296.

Choudhury, P. (Raj), Foroughi, C., & Larson, B. (2021). Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management Journal*, *42*, 655–683.

Cramton, C. D. (2001). The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science*, *12*, 346–371.

Cranefield, J., Yoong, P., & Huff, S. (2015). Rethinking Lurking: Invisible Leading and Following in a Knowledge Transfer Ecosystem. *Journal of the Association for Information Systems*, *16*. https://doi.org/10.17705/1jais.00394

Cummings, J. N. (2004). Work Groups, Structural Diversity, and Knowledge Sharing in a Global Organization. *Management Science*, *50*, 352–364.

Dahlander, L., O'Mahony, S., & Gann, D. M. (2016). One foot in, one foot out: How does individuals' external search breadth affect innovation outcomes? *Strategic Management Journal*, *37*, 280–302.

Dahlin, K. B., Weingart, L. R., & Hinds, P. J. (2005). Team diversity and information use. *Academy of Management Journal*, *48*, 1107–1123.

DeJesus, J. M., Umscheid, V. A., & Gelman, S. A. (2021). When Gender Matters in Scientific Communication: The Role of Generic Language. *Sex Roles*, *85*, 577–586.

Earley, C. P., & Mosakowski, E. (2000). Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal*, *43*, 26–49.

Earley, P. C., & Gibson, C. B. (2002). *Multinational work teams: A new perspective*. Routledge.

Eesley, C., & Wu, L. (2020). For Startups, Adaptability and Mentor Network Diversity Can be Pivotal: Evidence from a Randomized Experiment on a MOOC Platform. *Management Information Systems Quarterly*, *44*, 661–697.

Exley, C. L., & Kessler, J. B. (2022). The Gender Gap in Self-Promotion*. *The Quarterly Journal of Economics*, *137*, 1345–1381.

Faraj, S., von Krogh, G., Monteiro, E., & Lakhani, K. R. (2016). Special Section Introduction—Online Community as Space for Knowledge Flows. *Information Systems Research*, *27*, 668–684.

Fiol, C. M. (1994). Consensus, diversity, and learning in organizations. *Organization Science*, *5*, 403–420.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., … Arnadottir, J. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*, 1100–1104.

Gibson, C. B., & Gibbs, J. L. (2006). Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. *Administrative Science Quarterly*, *51*, 451–495.

Gonsalves, L. (2020). From face time to flex time: The role of physical space in worker temporal flexibility. *Administrative Science Quarterly*, *65*, 1058–1091.

Haas, M. R., Criscuolo, P., & George, G. (2015). Which Problems to Solve? Online Knowledge Sharing and Attention Allocation in Organizations. *Academy of Management Journal*, *58*, 680–711.

Haas, M. R., & Hansen, M. T. (2007). Different knowledge, different benefits: Toward a productivity perspective on knowledge sharing in organizations. *Strategic Management Journal*, *28*, 1133–1153.

Hansen, M. T., & Haas, M. R. (2001). Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. *Administrative Science Quarterly*, *46*, 1–28.

Hasan, S., & Bagde, S. (2013). The Mechanics of Social Capital and Academic Performance in an Indian College. *American Sociological Review*, *78*, 1009–1032.

Hasan, S., & Koning, R. (2019). Prior ties and the limits of peer effects on startup team performance. *Strategic Management Journal*, *40*, 1394–1416.

Hofstede, G. (1994). The business of international business is culture. *International Business Review*, *3*, 1–14.

Hwang, E. H., Singh, P. V., & Argote, L. (2015). Knowledge Sharing in Online Communities: Learning to Cross Geographic and Hierarchical Boundaries. *Organization Science*, *26*, 1593–1611.

Ibarra, H. (1992). Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm. *Administrative Science Quarterly*, *37*, 422–447.

Ibarra, H. (1997). Paving an Alternative Route: Gender Differences in Managerial Networks. *Social Psychology Quarterly*, *60*, 91–102.

Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and Problem-Solving Effectiveness in Broadcast Search. *Organization Science*, *21*, 1016–1033.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, N.J: Prentice-Hall.

Katz, L. F., Roth, J., Hendra, R., & Schaberg, K. (2022). Why do sectoral employment programs work? Lessons from WorkAdvance. *Journal of Labor Economics*, *40*, S249–S291.

Khattab, J., van Knippenberg, D., Pieterse, A. N., & Hernandez, M. (2020). A Network Utilization Perspective on the Leadership Advancement of Minorities. *Academy of Management Review*, *45*, 109–129.

Kilduff, M., Angelmar, R., & Mehra, A. (2000). Top management-team diversity and firm performance: Examining the role of cognitions. *Organization Science*, *11*, 21–34.

Kleinbaum, A. M., Stuart, T. E., & Tushman, M. L. (2013). Discretion Within Constraint: Homophily and Structure in a Formal Organization. *Organization Science*, *24*, 1316–1336.

Kretschmer, T., Leiponen, A., Schilling, M., & Vasudeva, G. (2022). Platform ecosystems as meta-organizations: Implications for platform strategies. *Strategic Management Journal*, *43*, 405–424.

Lai, H.-M., & Chen, T. T. (2014). Knowledge sharing in interest online communities: A comparison of posters and lurkers. *Computers in Human Behavior*, *35*, 295–306.

Lang, A. (2000). The Limited Capacity Model of Mediated Message Processing. *Journal of Communication*, *50*, 46–70.

Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and Control in Modern Society*, *18*, 18–66.

Leiponen, A., & Helfat, C. E. (2010). Innovation objectives, knowledge sources, and the benefits of breadth. *Strategic Management Journal*, *31*, 224–236.

Leonardi, P. M. (2014). Social Media, Knowledge Sharing, and Innovation: Toward a Theory of Communication Visibility. *Information Systems Research*, *25*, 796–816.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A Comparison of Methods to Test Mediation and Other Intervening Variable Effects. *Psychological Methods*, *7*, 83.

McIntyre, D. P., & Srinivasan, A. (2017). Networks, platforms, and strategy: Emerging views and next steps. *Strategic Management Journal*, *38*, 141–160.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444.

Mickeler, M., Khashabi, P., Kleine, M., & Kretschmer, T. (2023). Knowledge seeking and anonymity in digital work settings. *Strategic Management Journal*, *n/a*. https://doi.org/10.1002/smj.3504

Neeley, T. B., Hinds, P. J., & Cramton, C. D. (2012). The (un) hidden turmoil of language in global collaboration. *Organizational Dynamics*, *41*, 236–244.

Neeley, T. B., & Leonardi, P. M. (2018). Enacting knowledge strategy through social media: Passable trust and the paradox of nonwork interactions. *Strategic Management Journal*, *39*, 922–946.

Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, *18*, 187–206.

Ocasio, W. (2011). Attention to Attention. *Organization Science*, *22*, 1286–1296.

Page, S. E. (2019). *The diversity bonus: How great teams pay off in the knowledge economy*. Princeton University Press.

Piezunka, H., & Dahlander, L. (2015). Distant Search, Narrow Attention: How Crowding Alters Organizations' Filtering of Suggestions in Crowdsourcing. *Academy of Management Journal*, *58*, 856–880.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891.

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior*, *20*, 201–223.

Pu, J., Liu, Y., Chen, Y., Qiu, L., & Cheng, H. K. (2022). What Questions Are You Inclined to Answer? Effects of Hierarchy in Corporate Q&A Communities. *Information Systems Research*, *33*, 244–264.

Reagans, R. (2011). Close encounters: Analyzing how social similarity and propinquity contribute to strong network connections. *Organization Science*, *22*, 835–849.

Reagans, R., & McEvily, B. (2003). Network Structure and Knowledge Transfer: The Effects of Cohesion and Range. *Administrative Science Quarterly*, *48*, 240–267.

Reagans, R., Singh, P. V., & Krishnan, R. (2015). Forgotten Third Parties: Analyzing the Contingent Association Between Unshared Third Parties, Knowledge Overlap, and Knowledge Transfer Relationships with Outsiders. *Organization Science*, *26*, 1400–1414.

Reagans, R., & Zuckerman, E. W. (2001). Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams. *Organization Science*, *12*, 502–517.

Rhee, L., & Leonardi, P. M. (2018). Which pathway to good ideas? An attention-based view of innovation in social networks. *Strategic Management Journal*, *39*, 1188–1215.

Roche, M. P., Oettl, A., & Catalini, C. (2022, June). *(Co-)Working in Close Proximity: Knowledge Spillovers and Social Interactions* [Working Paper]. National Bureau of Economic Research.

Smith, J. A., McPherson, M., & Smith-Lovin, L. (2014). Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. *American Sociological Review*, *79*, 432–456.

Steffens, K. (2015). Competences, Learning Theories and MOOCs: Recent Developments in Lifelong Learning. *European Journal of Education*, *50*, 41–59.

Sullivan, B. N. (2010). Competition and Beyond: Problems and Attention Allocation in the Organizational Rulemaking Process. *Organization Science*, *21*, 432–450.

Sun, N., Rau, P. P.-L., & Ma, L. (2014). Understanding lurkers in online communities: A literature review. *Computers in Human Behavior*, *38*, 110–117.

Tortoriello, M., McEvily, B., & Krackhardt, D. (2015). Being a Catalyst of Innovation: The Role of Knowledge Diversity and Network Closure. *Organization Science*, *26*, 423–438.

Treem, J. W., & Leonardi, P. M. (2013). Social Media Use in Organizations: Exploring the Affordances of Visibility, Editability, Persistence, and Association. *Annals of the International Communication Association*, *36*, 143–189.

Treem, J. W., Leonardi, P. M., & van den Hooff, B. (2020). Computer-Mediated Communication in the Age of Communication Visibility. *Journal of Computer-Mediated Communication*, *25*, 44–59.

Tsui, A. S., Egan, T. D., & O'Reilly III, C. A. (1992). Being different: Relational demography and organizational attachment. *Administrative Science Quarterly*, 549–579.

Wagner, S., Hoisl, K., & Thoma, G. (2014). Overcoming localization of knowledge—The role of professional service firms. *Strategic Management Journal*, *35*, 1671–1688.

Wasko, M. M., & Faraj, S. (2005). Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly*, *29*, 35–57.

Wiesenfeld, B. M., Raghuram, S., & Garud, R. (1999). Communication patterns as determinants of organizational identification in a virtual organization. *Organization Science*, *10*, 777–790.

Yang, L., Holtz, D., Jaffe, S., Suri, S., Sinha, S., Weston, J., … Hecht, B. (2022). The effects of remote work on collaboration among information workers. *Nature Human Behaviour*, *6*, 43–54.

Zhang, D. J., Allon, G., & Van Mieghem, J. A. (2017). Does Social Interaction Improve Learning Outcomes? Evidence from Field Experiments on Massive Open Online Courses. *Manufacturing & Service Operations Management*, *19*, 347–367.

Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*, 197–206.

Table 1. Balance Checks for Randomization into Cohorts (N = 18 Waves, N = 38 cohorts)

Cohort characteristics	Cohort 1	Cohort 2	p-value (two tailed t-test)
Cohort size	373.905 (69.397)	368.135 (69.311)	$p = 0.763$
Female	0.391 (0.049)	0.381 (0.052)	$p = 0.489$
U.S. citizen	0.410 (0.101)	0.403 (0.089)	$p = 0.722$
Mean age as of course start	30.248 (2.737)	30.107 (2.572)	$p = 0.847$
# similar age peers (±5 yrs.)	155.258 (44.488)	154.281 (42.683)	$p = 0.936$
# similar age peers ((±2 yrs.)	80.975 (29.574)	80.333 (27.752)	$p = 0.937$
# same gender peers	180.863 (41.418)	178.909 (39.796)	$p = 0.862$
# same country peers (citizenship)	87.843 (32.987)	84.119 (33.345)	$p = 0.683$
# same location peers	123.719 (51.068)	120.720 (47.431)	$p = 0.826$

Note: 17 waves have two cohorts, and 1 wave, May 2020 wave has four cohorts. See Table A1 for details.

Table 2. Summary Statistics and Correlation Between Main Individual Level Variables

	Variable	Mean	Std. dev.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12
1	# outgoing views	588.186	837.727	0	12414	1.000											
2	# similar age views	120.093	194.026	0	3389	0.860	1.000										
3	# same gender views	291.274	431.739	0	6420	0.958	0.815	1.000									
4	# same country views	126.911	247.782	0	4470	0.656	0.604	0.619	1.000								
5	Manager	0.464	0.499	0	1	0.032	-0.059	0.042	-0.042	1.000							
6	English speaking	0.716	0.451	0	1	-0.021	-0.019	-0.022	0.292	-0.053	1.000						
7	Age	29.711	8.255	18	71	0.067	-0.147	0.078	0.024	0.466	0.024	1.000					
8	Female	0.389	0.488	0	1	0.030	0.044	-0.081	0.031	-0.095	-0.027	-0.074	1.000				
9	US citizen	0.483	0.500	0	1	-0.032	-0.015	-0.039	0.396	-0.112	0.609	-0.006	0.017	1.000			
10	# similar age peers	85.637	54.510	1	290	-0.036	0.215	-0.041	0.027	-0.371	0.014	-0.663	0.083	0.070	1.000		
11	# same gender peers	194.088	56.652	2	293	-0.017	-0.001	0.092	0.000	-0.029	0.008	-0.048	-0.513	-0.003	0.304	1.000	
12	# same country peers	88.905	89.922	1	303	-0.035	0.006	-0.036	0.402	-0.176	0.581	-0.095	0.024	0.873	0.267	0.201	1.000
13	cohort size	384.651	75.184	275	567	0.003	0.046	0.003	0.047	-0.129	0.022	-0.134	0.053	0.047	0.515	0.641	0.341

Note: Significant for $|\rho|>0.021$ at $p < 0.05$

Table 3. OLS Regressions of Course Performance on Demographic Similarity Between Cohort Peers

	Dependent Variable: Average Quiz Score					
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
# similar age (±2 years)	0.0422***			0.0413***	0.0240***	0.0340***
	(0.00415)			(0.00454)	(0.00586)	(0.00621)
# same gender		0.0300***		0.0227***	0.0164***	0.0252***
		(0.00396)		(0.00422)	(0.00544)	(0.00823)
# same country			-0.00491**	-0.0145***	0.00404	0.00536
			(0.00250)	(0.00263)	(0.00634)	(0.00717)
# outgoing views					0.00926***	0.00917***
					(0.000334)	(0.000335)
Age (omitted: age 18-25)						
age 26-31					-0.0107	0.452
					(0.577)	(0.582)
age > 31					-2.598***	-1.924***
					(0.716)	(0.729)
Female					-1.462***	-1.040
					(0.555)	(0.701)
Country *(omitted: Other)*						
USA					-3.153**	-3.288**
					(1.283)	(1.365)
EU					3.627***	3.681***
					(0.832)	(0.827)
Developed countries (excl. EU & USA)					-0.417	-0.550
					(1.012)	(1.007)
BRICS					1.867**	1.938**
					(0.784)	(0.785)
Manager					-2.168***	-2.098***
					(0.497)	(0.504)
English-speaking					1.788**	2.216***
					(0.754)	(0.755)
Cohort size						-0.0199
						(0.0376)
Constant	69.78***	67.59***	73.83***	66.76***	64.50***	67.14***
	(0.437)	(0.803)	(0.305)	(0.812)	(1.368)	(11.59)
Year-Month FE	N	N	N	N	N	Y
Observations	11,861	11,861	11,861	11,861	11,861	11,861
R-squared	0.009	0.005	0.000	0.013	0.121	0.135

Notes: Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4. Mediation Analysis using Sobel-Goodman test of mediation

VARIABLES	# similar age views (±2 years)	# same gender views	# same country views (citizenship)	Dependent Variable: Average Quiz Score		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
# similar age views (±2 years)				0.0351***		
				(0.00166)		
# same gender views					0.0172***	
					(0.000658)	
# same country views (citizenship)						0.0288***
						(0.00145)
# similar age (±2 years)	1.132***			-0.00375		
	(0.0556)			(0.00648)		
# same gender		1.059***			-0.0177**	
		(0.102)			(0.00839)	
# same country			0.710***			-0.0104
			(0.0778)			(0.00722)
Age (omitted: age 18-25)						
26 to 31 years	39.77***	31.96***	19.07***	-0.285	-0.408	-0.339
	(4.807)	(10.28)	(5.779)	(0.589)	(0.561)	(0.568)
> 31 years	33.47***	62.41***	25.89***	-1.983***	-4.367***	-3.926***
	(5.300)	(10.70)	(5.641)	(0.737)	(0.608)	(0.612)
Female	12.35***	0.703	15.33***	-2.587***	-2.210***	-2.523***
	(3.684)	(9.989)	(4.350)	(0.438)	(0.708)	(0.441)
Cohort size	-0.349	-2.087***	-0.643*	-0.0119	0.0145	-0.00417
	(0.295)	(0.722)	(0.343)	(0.0378)	(0.0380)	(0.0381)
Constant	121.0	680.5***	219.9**	71.62***	66.99***	72.90***
	(92.06)	(225.5)	(107.0)	(11.75)	(11.70)	(11.85)
Individual-level controls	Y	Y	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y	Y	Y
Observations	11,862	11,824	11,862	11,861	11,823	11,861
R-squared	0.069	0.032	0.182	0.109	0.125	0.104

Notes: This Table employs the Sobel-Goodman test of mediation. Models 1-3 regress the mediators on the independent variable and all control variables. Models 4-6 regress the dependent variable on the mediator as well as the independent variable and all control variables. Individual-level controls not shown are: Country of primary citizenship, Manager, and English-speaking. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5. Mediation Test of Indirect Effects

VARIABLES	Indirect effects Observed Coefficients	Bootstrapped Std. Err.	95% Conf. Interval		
# similar age views ($\pm$2 years)	0.0397	0.00200	0.0358	0.0436	(BC)
			0.0359	0.0437	(P)
# same gender views	0.0182	0.00187	0.0145	0.0218	(BC)
			0.0145	0.0218	(P)
# same country views (citizenship)	0.0203	0.00199	0.0164	0.0242	(BC)
			0.0164	0.0242	(P)

Notes: All indirect effects are estimated using Sobel-Goodman Mediation Tests; bootstrap (50,000 reps), observed coefficients with bootstrap standard errors, bias-corrected (BC) and percentile (P) 95% confidence intervals.

Table 6. Moderation of number of demographically similar peers on viewing behavior

	DV: Total # of Outgoing Views			DV: # of Views to Similar Peers		
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
# similar age ($\pm$2 years)	0.363			0.904***		
	(0.255)			(0.0344)		
# same gender		-0.270			1.517***	
		(0.276)			(0.126)	
# same country			5.897			4.458**
			(10.54)			(1.818)
age 26-31 X # similar age	-1.715**			0.505***		
	(0.796)			(0.0748)		
age >31 X # similar age	-1.031			1.031***		
	(0.638)			(0.0751)		
Female X # same gender		-0.840**			0.0889	
		(0.339)			(0.0634)	
Country (omitted: Other)						
USA X # same country			-5.904			-3.777**
			(10.53)			(1.816)
EU X # same country			7.089			-1.617
			(17.13)			(2.955)
Developed countries (excl. EU & USA) X # same country			-2.063			-2.655
			(12.23)			(2.057)
BRICS X # same country			-3.765			-2.065
			(10.51)			(1.806)
# outgoing views				0.202***	0.496***	0.201***
				(0.00428)	(0.00569)	(0.00813)
Constant	952.5**	1,033**	1,088**	-21.08	168.0***	16.70
	(426.1)	(423.2)	(423.2)	(39.33)	(61.65)	(71.67)
Year-Month FE	N	N	N	N	N	Y
Observations	11,862	11,862	11,862	11,862	11,824	11,862
R-squared	0.019	0.019	0.019	0.819	0.941	0.638

Notes: Controls not shown are: Female, Age, Country, Manager, English-speaking, and Cohort Size. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1: Screenshot of Peer Activity in Module Reflection Example

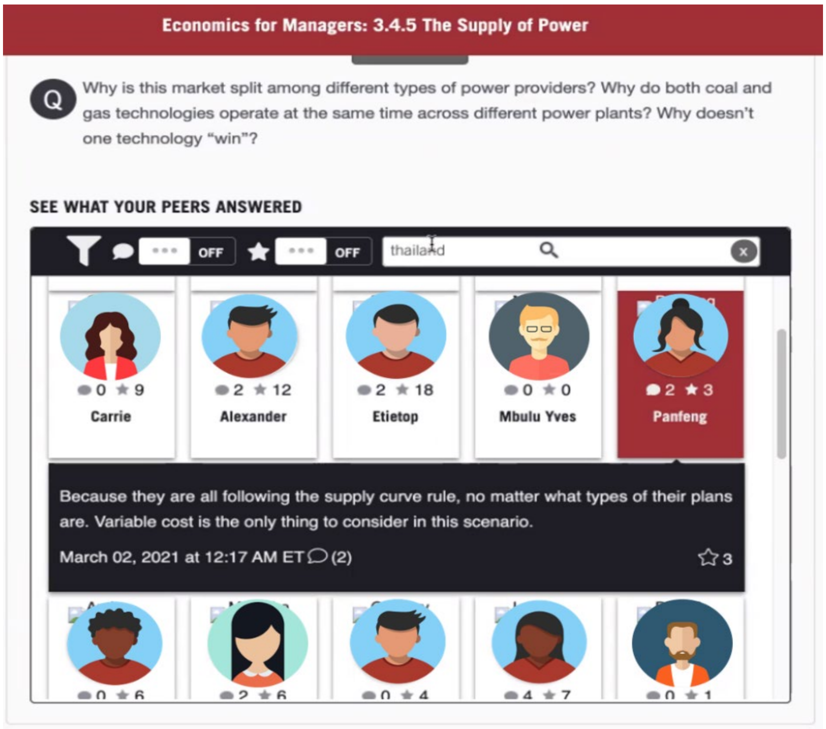
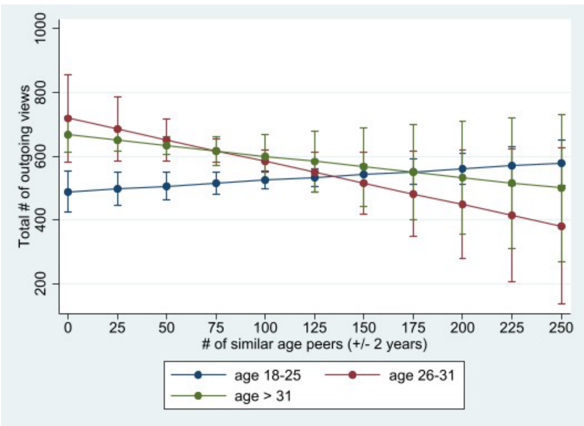
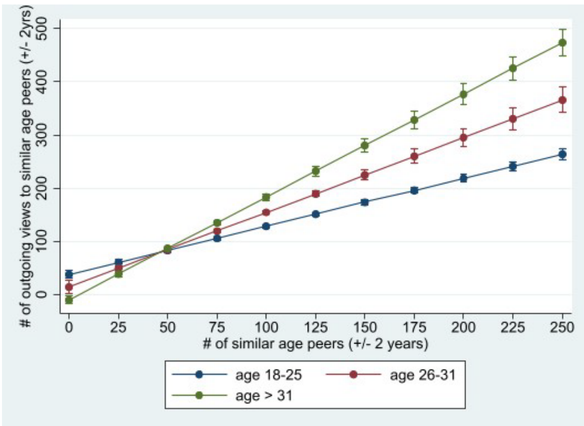


Figure 2: Moderation of similar age peers on viewing behavior by age



a) Total # of outgoing views



b) Outgoing views to similar age peers

Appendix

Table A1. Course Waves with Multiple Cohorts (N = 38 cohorts)

Wave #	Year	Month	Nb. cohorts
1	2017	June	2
2	2017	October	2
3	2017	September	2
4	2018	February	2
5	2018	March	2
6	2018	October	2
7	2018	September	2
8	2019	February	2
9	2019	May	2
10	2020	July	2
11	2020	June	2
12	2020	March	2
13	2020	May	4
14	2020	October	2
15	2020	September	2
16	2021	February	2
17	2021	May	2
18	2022	May	2