Forum size and content contribution per person: a field experiment

Jiye Baek (Hong Kong University of Science and Technology) and Jesse Shore¹ (Boston University)

Abstract

Promoting contribution of content is a key challenge for platforms that support the collective creation or transfer of knowledge. We study the role of forum size (number of people in a forum) on contribution of content per person with a field experiment on a massive open online course (MOOC). We find that larger forums elicit more contribution per person. The number of questions and other help-seeking threads posted per person was unchanged by size, but replies and other more conversational posts increased sharply. Most of the positive effect of size was in a subset of socially responsive subjects. The implication of social responsiveness driving our results is that the unequal distribution of contribution on online platforms is unlikely to be easily changed: if more contributions are elicited from infrequent contributors, the greatest contributors would contribute even more due to there being more to respond to.

Keywords: online engagement, discussion forums, group size, user-generated content, Massive Open Online Courses, free riding, lurkers, knowledge sharing

1. INTRODUCTION

Digital platforms make it easy to bring together large groups of people with shared interests in online discussion communities or forums to share and learn from each other. These online forums are important sites for the creation and transfer of knowledge (Faraj et al., 2016) and can be convened as a stand-alone entity or by an organization as a complement to its official products or services (Jeppesen and Frederiksen, 2006) for users to help each other, innovate (Faraj et al., 2016), or develop a greater engagement with the firm's community or brand (Brodie et al., 2013). However, a common challenge is that only a small minority of users actively contribute to discussion by posting content or asking or answering questions. Promoting greater engagement and active contribution is thus considered a key challenge for online sites or enterprise systems that support the communal creation or transfer of knowledge (Beenen, et al, 2004, Farzan, et al., 2008; Ray, et al., 2014, Hwang et al., 2015).

¹ Correspondence to jccs@bu.edu

This paper studies the connection between those two defining features of online forums – their potential for large size and their low engagement rates – in a randomized field experiment. Size (hereafter “size” refers to the number of people in one online space for interaction) is known to affect engagement, but predictions from prior literature are conflicting, with some studies (Zhang and Zhu, 2011; Huang, et al., 2017) saying size promotes engagement and others (e.g. Chidambaram and Tung 2005; Latane et al. 1979; Mao et al., 2016; Kim, 2013) saying size depresses engagement per person. While existing theory may explain the effects of size on contribution in some settings and under specific interventions, we argue below that it is inconclusive about the effects of size in a broadly-used technology used to support community and the sharing of knowledge: discussion forums.

We conducted a randomized field experiment on forum size, in order to assess the role of size in causing an increase or decrease in forum participation per person. Our setting is the discussion forums on a Massive Open Online Course (MOOC), with participants randomized into forums of size 125, 500, or 2000. We show that larger size causes more participation per person on average, but does not change the proportion of people who post.

The increase in posting in larger cohorts was predominantly in the form of replies -- by a subset of especially socially-responsive people -- to existing threads. To a lesser extent, size also increased the posting of non-help-seeking new threads. In sum, our interpretation is that size gives socially responsive people more to respond to and thus motivates a sharp increase in posting from them. For those seeking information or those without specifically social motivations, the effect of size is much smaller or non-existent.

A novel implication of social responsiveness driving our results is that the inequality of contribution on online platforms is unlikely to be easily changed. Any intervention that increased posts from low-contributors would also result in more posts from socially-responsive high-contributors by giving them more to respond to.

2. THEORETICAL BACKGROUND

2.1 Evidence for a positive effect of size

2.1.1 *Social capital and size*

An explanation that people commonly give for their own forum posting behavior is the presence of established relationships, sometimes called relational social capital (Wasko and Faraj, 2005) or common bonds (Ren et al., 2007). Two major studies of the effects of size on contribution demonstrate the interaction of social capital and size. First, Zhang and Zhu (2011) document the effect of the Chinese government unexpectedly blocking Chinese-language Wikipedia from access by users and editors within mainland China. Immediately after the block, there was a sharp reduction in editing activity by non-blocked editors who had previously interacted heavily with blocked editors.

Zhang and Zhu's results (2011; table 4) show that after controlling for how much a non-blocked editor collaborated with blocked editors prior to the block, the main effect of the block on contribution to Wikipedia was no longer statistically significant at the five percent level: people who had relationships that were severed contributed much less in the short run, while others carried on more or less as normal. Thus, the results of this study remain somewhat ambiguous: are the reductions in contribution an effect of smaller size or simply a natural consequence of severing specific established relationships, where people had concentrated their energies? If the reduction of effort they observed was the consequence of severing specific relationships, then it could be the case that it is not small size *per se* but rather the forcible reduction in size that depresses contribution.

Huang et al. (2017) report a similar result in the opposite direction. People posted more content to a review website after integration with the social networking site Facebook caused their reviews to be visible to their friends. People interact more with their online friends – that is, people with whom they have an established relationship online; adding friends promotes engagement, while severing ties

depresses engagement. The more general question of the effects of size remains unclear from these studies, however.

2.1.2 Content and Resource availability

In online forums, each individual's comments become available as a resource for future users; thus, larger groups and communities have more resources and are more valuable and attractive to new users (Valacich et al. 1992, Butler 2001). For example, individuals will have a greater chance of finding interesting information or the answer to a question in a bigger discussion forum where many participants have already engaged. By this logic, the larger the community, the more new users will be attracted to the resources provided by that community. On Wikipedia, for example, content attracts more editors who create more content, creating a virtuous cycle (Aaltonen and Seiler, 2015, Kane and Ransbotham, 2016).

In other words, online communities display “network effects” or “network externalities.” The theory of network effects says that the value of an interaction technology or platform increases with the number of users on that platform (Katz and Shapiro 1985, 1994).

What is not clear from the “resource availability” or “network effects” perspectives is whether an increased tendency to *join* larger communities translates into an increased per-person tendency to *participate* in them. The presence of more resources (usually in the form of archived existing discussions) means that there is less need to interact: if the answer to my question is already archived, I can just read the existing answer without actively asking again. Indeed, Kane and Ransbotham (2016) show for Wikipedia articles that positive network effects on individual articles attract more readers, but diminish contributions from editors as the content becomes increasingly complete and refined. Similarly, in online education, Beaudoin (2002) reports that people who do not contribute to class discussions feel that there is no need when others have already posted similar points.

2.2 Evidence for a negative effect of size

In larger collaborative groups, contribution per person is lower than in smaller ones (Thomas and Fink, 1963; Bales and Borgatta 1966; Wheelan 2009). Much prior work argues that greater size creates challenges for both online and in-person groups, and that the amount of interaction per person goes down as group size increases because there is more free-riding: obtaining value from a group without contributing to that group (Albanese and van Fleet 1985). Similar to free-riding, prior research has studied “social loafing” (Ingham, et al., 1974; Latane et al., 1979) and the “bystander effect” (Latane and Darley, 1968), showing that a person is less likely to take action or assume responsibility when there are others in a group or otherwise co-present, and that the probability of contribution of effort or helping others is inversely related to the number of people present (Barron and Yechiam, 2002; Bray, et al., 1978; Lowry et al., 2006). Thus, in smaller groups, individuals tend to contribute more time and energy to interact and share information with others because they feel responsible to the group.

It has long been known that, in traditional offline organizations, people in larger groups feel that they matter less, make less of a difference, and that others may not recognize their contributions (Gooding and Wagner, 1985; Kerr 1989; Kerr and Bruun, 1983). Large group size also leads to higher communication and coordination costs (Brooks, 1975; Pendharkar and Rodger 2009) and it can be harder to tell how much and how each individual has contributed (Jones 1984; Kerr and Bruun 1981). Some studies on collaborative groups show a positive effect of size on collective output (e.g. number of posts on a forum), but not contribution per person (Koh et al. 2007; Carillo and Okoli 2011), though in complex tasks, gains from coordination can outstrip losses from individual effort (Mao et al., 2016).

Large group size is less of an obstacle to a high level of individual contribution online than it is offline, because digital collaborators do not suffer from the “production blocking” effect (Diehl and Strobe 1987, Gallupe et al. 1992) in which group members must wait for each other to finish before initiating their own engagement; however, larger groups still have lower participation per person than

smaller groups in digital collaboration. For example, empirical evidence on problem-solving within technology-supported groups shows that size is correlated with a decrease in participation per person, idea quantity per person, decision quality and group cohesiveness because of more free riding (Chidambaram & Tung 2005, Alnuaimi et al. 2010; Valacich et al. 1995, Yap and Bock, 2006).

Even online, increased group size makes interaction between group members more difficult and complex (Riopelle et al., 2003) and leads to moral disengagement (Alnuaimi et al., 2010). Members of larger groups also have more difficulty establishing relationships with others and can be overwhelmed by the high volume of communication, which can lead to reduced contribution or attrition from online groups (Jones, et al., 2004; Wang et al., 2013).

Effort per person in contests and competitions is also reduced when there are more people or teams competing for the same prize (Boudreau et al, 2011). People who are competing against each other are typically not also working collaboratively with each other and thus the reduction in effort cannot be called free riding. Yet, the reason for the reduction is still conceptually related: in both collaboration and competition, the incentive or output is at the collective level but the effort is individual. In competitions, greater size means participants have a lower chance and greater uncertainty about winning a prize; in collaborations, greater size means participants' responsibility for the group's work is diluted.

Results in the context of online education have been equivocal. We are not aware of any research results on the effect of size in MOOCs per se, but several studies present evidence from semester-long online university courses. One field experiment shows that students in smaller online discussion forums read more posts and interact more with other students (Kim 2013). This study employed real-time class discussions, however, raising the possibility that production blocking or other coordination losses may have been responsible for the connection between size and depressed performance. A retrospective study on observational data showed the opposite result: students in larger classes contributed more to class discussions (Qiu et al. 2012), but size and student contribution are

endogenously correlated, and it is unclear whether engaging classes simply attracted more students or whether the presence of more students caused more engagement to occur. The causal arrow, if any, could have pointed in either direction.

2.3 Diverse motivations for and barriers to contribution to online discussion

We turn to a stream of literature that has addressed a related but distinct question: why do people contribute or lurk in online discussions? Although this research is not intended to tell us the effects of size on participation, it can give us important clues if we ask how motivations to contribute would be affected by larger or smaller cohorts of people.

2.3.1 Social responsiveness: contributors are motivated by the presence of other people

Among those who do post online, a common motivation for contribution is a desire to be social, interact with, and help members of a community (Butler, et al, 2002; Budhathoki and Harythornthwaite, 2013; Constant, et al., 1996; Wasko and Faraj, 2000, Wiertz and De Ruyter, 2007, Sun et al., 2014). Feelings of both common identity and common bonds are associated with greater contribution to online discussion (Ren, et al., 2007), particularly in the form of replies to others' posts (Bateman et al, 2011). People who post report having a desire to tell stories and participate in conversations, and be a member of a community (Nonnecke et al., 2006), while lurkers do not. Some contributors are motivated specifically by helping others (Wasko and Faraj, 2000; Wasko and Faraj, 2005; Kankanhalli, et al. 2005).

Many people who contribute to online forums do so because it enhances their reputations (Wasko and Faraj, 2005), identity (Ma and Agarwal, 2007), or social status (Ackerman, 1998; Lampel and Bhalla, 2007). For these people, helping others is a means of presenting oneself as knowledgeable and making oneself more visible online. For example, reputation systems promote more posts overall, but do not increase the number of individuals who contribute posts (Coetzee, et al, 2014).

Status-seeking, community-seeking, conversation-seeking, shared identity with the group, and the desire to help others are different psychological mechanisms, but they are all essentially social motivations to contribute. This is to say that contribution occurs specifically because of the presence of other people or their posts. For this reason, we place all of these motivations for contribution under the broader term “social responsiveness.” Each post by another person on a forum is an opportunity to fulfill socially responsive motivations, whether they are rooted in altruism, or a desire for conversation or status. Therefore, we reason that larger forums have more posts in total to respond to and thus provide more opportunity for socially responsive people to post.

2.3.2 Some contributors are motivated by access to information and knowledge

Some people post in online forums simply because they want the answer to a question, or in perceived exchange for knowledge they received in the past or expect to receive in the future (Constant, et al. 1996; Wasko and Faraj, 2000; Chiu, et al., 2006; Kankanhalli, et al. 2005). The effects of size for information seekers may depend on how individualized knowledge is: if anyone could answer a question, size may not matter past a certain critical mass, but if knowledge very individualized, a larger community means a greater chance of receiving a good answer. Although the effect of size to information seeking is of unclear magnitude, it should not be negative.

2.3.3 Reasons for not contributing

For some people, the presence of other people is a barrier to posting. Some people are simply more self-confident and confident in what they know, and those who are not confident are less likely to contribute (Kankanhalli, et al. 2005; Nonnecke, et al., 2006; Sun, et al., 2014). Lurkers – those who never contribute to an online discussion – report feeling shyness, a lack of confidence, that they have nothing to offer the discussion, and/or that they don’t belong in the community (Preece, et al, 2004; Nonnecke, et al.,

2006; Sun, et al., 2014). Although social comparison can motivate increased participation in some cases (Burtch, et al, 2017), when people's differing abilities are being compared, it is demotivating for low-ability individuals (Ashraf, et al., 2014).

Lurkers also report that their information needs are met without posting, which could be interpreted as free riding, (Preece, et al, 2004; Nonnecke, et al., 2006; Sun, et al., 2014). However, others say that they would post if only other posters had not posted something similar already (Beaudoin, 2002; Nonnecke, et al., 2006), suggesting more of a crowding out effect than a free-riding effect. Some feel that they are contributing to the community by refraining from posting (Preece, et al., 2004). Finally, lurkers are sometimes concerned with privacy and security (Nonnecke and Preece 2001).

The reasons lurkers give for lurking are likely to be exacerbated by size. A lack of confidence or shyness would likely be worse in a larger cohort. If people free ride off of others' ideas or their ideas are crowded out by others' posts, then this would be more pronounced in a larger cohort. And if a desire for privacy prevents posting, this would only be worse in a larger cohort.

2.3.4 Posters are motivated by size; lurkers are demotivated by size (but don't post anyway)

According to our read of prior literature on motivations to contribution, size ought to have different effects on different people. Greater size should provide greater opportunity for social reward, whether they are to help others or build one's own reputation or image. On the other hand, greater size also provides a greater barrier to contribution by those who are shy, private, or feel that they don't belong. In sum, the literature on motivations can be read to support a net increase in posting per person in larger cohorts: size increases posting by those who post, and would decrease posting by those who do not post, if there were anything to decrease.

3. RESEARCH METHODOLOGY

We conducted a field experiment in Boston University's 'Sabermetrics 101'² course offered on edX, which began on July 7th, 2015 and ran for 10 weeks. The edX platform was launched through a partnership between Harvard University and the Massachusetts Institute of Technology in May 2012 and has become one of the largest MOOC providers, along with Coursera and Udacity. Participating institutions have expanded to include 48 core "charter universities" (including Boston University, which provided access to the course we studied), along with many additional partner schools and organizations. It hosts online university-level courses to internet users all around the world and provides an online discussion forum for user interaction.

In many ways, MOOC discussion forums are typical of discussion forums in other settings. The forums are a complement to the main content provided by the MOOC, much as a technology company might provide a platform for a "user community" (Di Gangi et al, 2010). Even if the providers of the main content or product do all they can to create something usable, clear and self-explanatory, there are always unanticipated questions, misunderstandings and novel ideas that users have, and forums provide a space for discussion of these. MOOC forums are also typical of other forums in their large scale and low engagement rate. MOOCs can have over 100,000 students learning from a single professor posting digital content (Breslow et al. 2013, Hood et al. 2015). Typically, however, only 3-5 percent of users contribute posts to MOOC discussion forums (Breslow 2013, Rosé et al. 2014) and more than 90 percent of users eventually drop out of MOOCs before the course end date.

MOOCs also have distinctive features that facilitate experimental study. MOOCs are advertised in advance so that learners can register for the course before a clearly defined launch date. The pool of pre-registered users can then be randomized into comparable subject pools that are exposed to

² "Sabermetrics" denotes the statistical analysis of baseball player and team records; the word derives from the acronym SABR, for the Society for American Baseball Research. According to the course website, the Sabermetrics 101 course primarily covers the "basics of data science and how it applies to the study of baseball" and the "fundamentals of the R and SQL programming languages."

different treatment conditions, such that the causal effects of the treatments can be estimated in a straightforward way.

3.1 Dependent Variables

We measured forum contributions several different ways. We considered all posts made by a user (*# posts*), as well as distinguishing between posts that are new threads (*# threads*) on the forum from comments made on an existing thread (*# replies*). We read the text of each new thread and classified it as help-seeking or not, tabulating the number of such comments (*# help-seeking*). We test the effects of the treatment on the number of posts for each variable, as well as on a dichotomous indicator of whether the variable is greater than zero (*any posts*, *any threads*, *any help-seeking*, *any comments*, and *any help-seeking*).

We measure *grade* as an average grade of all assignments, quizzes and the final exam, including any 0s for any that were not completed. Because grade is highly correlated with the number of assignments completed, we also consider the grade per completed assignment (*grade/assignment*). We also use two dichotomous variables set to different thresholds of significant effort in the course: whether the user passed the course by earning 60% of available points (*passed course*), and whether a user earned greater than 90% of available points (*grade >90%*). Course completion is highly correlated with grade but somewhat distinct from it, and we measure it with the number of assignments – lecture questions, quizzes and final exam – that had a non-zero score (*# assignments*) as well as a binary version, indicating whether a user completed at least one assignment (*any assignments*).

Additionally, although most students are auditors in the sense that they take the course for free and do not receive any official credit, edX provided the option to work toward an edX *ID Verified Certificate* by paying a \$25 fee and achieving a passing grade. For the present course, a user could opt-in to the ID-verified track up through August 13th, 2015, which was approximately halfway through the

course. Whether or not the student passed and thus actually received the certificate was determined at the end of the course. We treat the decision to pay the fee (*paid verification*) as a binary dependent variable that indicates a high intention to complete the course.

3.2 Experimental manipulations

Our experimental design focuses on manipulating the number of students who could interact with each other in the discussion forum portion of the edX platform. In a typical edX course, all students use the same discussion forum to communicate with each other; in our experiment, we replaced the single catch-all forum with multiple forums serving cohorts of different numbers of students. Each student in our study population was assigned uniformly at random to one and only one cohort, and student forum posts were only visible to other members of their own cohort. Students were aware that other cohorts existed, but they could not see or participate in discussions in those cohorts.

Randomization into separate cohorts was performed one week before the course's official start date, just prior to the time that the course website was made accessible to registered students. At this point in time, there were just over 6000 pre-registered students, and we randomized 6000 of these pre-registered students into experimental treatments.

We picked our cohort sizes assuming a high rate of initial attrition. Prior research on MOOCs had shown that 60-70% of initially enrolled students do not end up participating in the course (Clow, 2013). We therefore created three conditions for cohort size of 125 people per cohort (treatment "S"), 500 people per cohort (treatment "M"), and 2000 people per cohort (treatment "L"), with the expectation that the effective sizes would be on the order of 40, 160, and 640 people per cohort after initial attrition, with a smaller fraction participating heavily. Because the overall number of people randomized to each treatment group is the same, there are different numbers of cohorts within each treatment (see Table 1).

We chose exponentially increasing sizes for two reasons. First, we wanted our results to be relevant across the orders of magnitude of sizes found in the literature, ranging from teams in organizations to sizeable online communities. Second, prior literature gives us reason to think that the positive and negative effects of increased size may be sub-linear or logarithmic, rather than linear. In the early work on social loafing, effort appears to decrease with the log of group size (Latane, et al., 1979); on the positive side, matching of rare tastes may also imply logarithmic positive effects of size. If we take the distribution of interests and tastes tend to be highly skewed (creating a heavy-tailed distribution of tastes as in the “long tail”), then exponentially larger population sizes would be necessary to find matches for rarer tastes. Finally, we assumed that individual ability to respond to others is not infinite – even for the most socially-motivated subjects -- and thus there may be an attenuation of the positive effects due to limits on individual time. Thus, we reasoned that an exponentially increasing cohort sizes would increase the chance of detecting differences among larger cohorts.

Within the smaller two treatments, we also ran a two-way treatment in which we informed half of the cohorts that forum participation was required to get full course credit (see Appendix 4 for more details on how the treatments were communicated to users). In traditional learning environments, instructors often grade students on their participation in the discussion to motivate students’ engagement and we adapted this to our study. Additionally, the literature on online engagement emphasizes the importance of norms and we wanted to guard against the possibility of a low-engagement norm emerging. Thus, we implemented the participation requirement by notifying students by email and by posting the requirement on the course webpages (treatment “R”). Students in the non-treated groups received comparable messages that participation in the forums was encouraged (treatment “E”). Subjects received messages informing of their participation requirement (or encouragement) at the end of each assignment (there were 33 assignments in total), to ensure they were aware of the forums and expectations for posting in them.

Other than the forums, all cohorts had access to the same course materials. Many students continued to register for the course after the date at which we assigned students into treatments, but they were not included in the experiment and instead assigned to a “default” cohort, which grew over time as students enrolled. Finally, students remained in a single cohort for the entire duration of the MOOC.

Two teaching assistants acted as primary forum moderators for all cohorts; additionally, the professor and course coordinator stepped in on a handful of occasions. In the experiment design phase, we worked with the teaching team to ensure that all questions were treated equally, in order to maintain the equivalence of the cohorts except for the treatments. Section 4.4 describes the effects of the teaching team in greater detail.

Table 1: Field experiment design

	Treatment S 125-person cohorts	Treatment M 500-person cohorts	Treatment L 2000-person cohort
Treatment E (posting encouraged)	SE 8 cohorts	ME 2 cohorts	LE 1 cohort
Treatment R (posting required)	SR 8 cohorts	MR 2 cohorts	(empty)

3.3 Other independent variables

Users provide self-reported data to edX at the point of sign-up for the platform, which for our experimental subjects was strictly prior to the course beginning. We use self-reported gender, self-reported level of education, self-reported age, number of weeks since signing up for edX, days elapsed since signing up for the course, and whether signing up for edX and the course coincided in time as additional control variables.

Additionally, we use a free text self-reported “goals” for the course as a source for additional controls as follows. Users that included any of the words “career”, “work” or “job” were coded with a 1 on the variable *goals: work*. Users that included the word “baseball” were coded with a 1 on the variable *goals: baseball*. Users that included variants of the word “sabermetrics”, the phrase “data science”, variants of the word “statistics,” variants of the word “analytics” or “SQL” or “R” were coded with a 1 on the variable *goals: statistics*. Additionally, users that left that field blank were coded with a 1 on the variable *no goals given*.

We assessed the effectiveness of the randomization procedure in generating balance in the above covariates across treatments. The difference in means (or for continuous variables, standardized means) was less than 0.06 for all pairwise comparisons between treatments for all variables except age. The difference in standardized means for age between the ME and SR treatments was 0.107 and significant in a t-test. Thus, we control for age in regression models, but not the other exogenous variables. In the tables, below, age is centered and scaled to unit variance. The appendix contains models that include the exogenous covariates omitted in the main paper.

3.4 Endogenous measures of motivation

In addition to the dependent exogenous variables discussed above, we also consider two endogenous variables in certain models that capture different aspects of the relative engagement and motivation that subjects had for the course and forum. Both variables are dichotomous variables, indicating whether the subject engaged in certain ways at the beginning of the course. Neither variable is significantly correlated with the size treatment, but as post-treatment covariates, their inclusion could bias causal estimates of the treatment effects (e.g. Montgomery, et al, 2016). Thus, we include them only in models intended to investigate mechanisms behind the causal effects by illustrating heterogeneous effects of the treatments.

The first endogenous variable we examine is *quick start*, which indicates whether learners watched a video within the first hours after the course began. The first video content was posted at approximately 3pm Eastern Time on June 30th, 2015. A small subset of learners watched that content before midnight, Eastern Time, that same day. Approximately 6.3% of learners in the participation encouraged treatment and 7.1% of the participation required treatment watched a video that day (the difference was not significant: Chi-squared: 1.1421, df=1, $p = 0.285$). The fraction of learners watching a video did not vary significantly by size (participation encouraged: Chi-squared = 0.66521, df = 2, $p = 0.7171$; participation required: Chi-squared = 0.10544, df = 1, $p = 0.7454$).

The second endogenous variable we examine is denoted *introduced self*. On 6/30/2015 (the date that the “pre-module” was released), the teaching assistants (TAs) posted the same message in all cohort forums, introducing themselves and asking the learners in that forum to introduce themselves as well. We read through all posts – including replies and new threads -- made until 7/7/2015 (the date that the first course module was released). Approximately 7.8% of learners in the participation encouraged cohorts introduced themselves and 10.8% of learners in the participation required cohorts introduced themselves (the difference in proportions was significant: Chi-squared = 14.97, df=2, $p = 0.0001$). Holding the participation treatment constant, there were no significant differences in the proportion of learners that introduced themselves due to size (for participation encouraged: Chi-squared = 3.0554, df = 2, $p = 0.217$; for participation required: Chi-squared = 0, df = 1, p -value = 1). We restrict measurement of *introduced self* to the seven days before the course officially began to reduce the amount of endogenous variation in this variable and because no assignments had been posted yet, raising our confidence that any such self-introductory posts were not merely a preamble to a request for help and thus less likely to serve an instrumental purpose.

In analyses below, we use the two variables *quick start* and *introduced self* to assess how the treatments affected people with different levels of motivation and engagement differently. The distinction

between the two variables is important. *Quick start* reflects immediate engagement with the course content and we interpret this variable as an indicator of the subset of subjects who had an overall motivation for the course content but not necessarily motivation for posting. *Introduced self* captures a more social dimension to engagement at the point of the course's beginning. We reason that people who introduced themselves in response to a request to do so were more socially responsive or socially oriented than those who did not.

The two variables are correlated ($r=0.196$) but distinct, and we use the distinction to better understand the mechanism behind the treatment effects. If we find a significant interaction between larger size and *quick start* in a model of some dependent variable of interest, then that is evidence that large size had a greater effect for the subset of students that were especially motivated overall. If, on the other hand, we find a significant interaction between larger size and *introduced self*, then that is evidence that large size had a greater effect for the subset of students that were especially socially responsive (however, because these variables are endogenous, the discussion section (below) considers alternative interpretations of the results). For easiest interpretation, we parameterize models with dummy variables indicating which of four groups a subject is in: those who were positive on neither *quick start* or *introduced self*, those who were positive on *quick start* only, those who were positive on *introduced self* only, and those who were positive on both variables.

3.5 Model choice

The count variables we study show evidence of over-dispersion: the variances of these forum engagement variables are substantially larger than their means. When the data are over-dispersed, the standard errors in Poisson models are biased downward and, therefore, quasipoisson or negative binomial (NB) regression models are preferred (Cameron and Trivedi 2013). Another characteristic of our data is that very few students in our sample posted questions and comments. Of the total sample of 5981 subjects,

only 1070 posted on the forum; most participants have a count of zero for the forum engagement. Thus, we also considered zero inflated negative binomial (ZINB) regressions. ZINB models account for overdispersion and model the presence of excessive zeroes. We fit quasipoisson, NB, and ZINB regression models for count dependent variables and compared them with Vuong’s likelihood ratio test for choice among non-nested models (Vuong, 1989). NB and ZINB were always chosen over quasipoisson models. After AIC-correction for model complexity, NB models were chosen over ZINB models. However, we find that all three models give comparable estimates of effect sizes, and the choice of model does not affect our conclusions. In reports of parameter estimates, we use cluster-robust standard errors to account for within-cohort nesting of observations. For models of the dichotomous variables such as *any posts*, *any assignments* or *grade > 0*, we used logistic regressions with cluster-robust standard errors.

Our main sets of models estimating the causal effects of the treatments on posting (Tables 3 and 4) and performance and attrition (Table 5) contain only the treatment variables plus age. A follow up set of models (Table 6 intended to describe heterogeneous treatment effects also includes *quick start*, *introduced self* and their interactions with each other and the treatment variables. Finally, the appendix contains models built from treatments, exogenous control variables, and their interactions, such that variables were included in the final model only if they improved cross-validation error over simpler models.

4. RESULTS

4.1 Causal effects of size

Larger size causes more forum posts per person on average, but does not increase the number of people posting. The increase in posting in larger cohorts is primarily composed of an increase in replies to

existing threads; posting of new threads per person did not vary significantly by size. Because new thread posting was roughly constant, while replies per person increased, the ratio of replies per thread was higher in larger cohorts. Requiring participation increased all forum posting behaviors.

Raw averages are presented in Table 2, parameter estimates are in Tables 3 and 4, and detailed results follow, below. In general, parameter estimates for medium-sized cohorts were statistically indistinguishable from zero, but consistent with a trend caused by size. Thus we report their effect sizes and confidence intervals in the text, along with the significant estimates for the large cohort. All reported effect sizes are per-person (not per-cohort) effects.

Table 2: Average values of dependent variables by treatment condition

treatment	# posts	# threads	# replies	# help-seeking	# replies: hst	# assignments	grade	verified	passed
MR	2.183	1.075	1.108	0.259	0.173	3.609	0.085	0.073	0.074
SR	1.844	0.891	0.953	0.294	0.160	3.363	0.080	0.057	0.068
LE	0.709	0.248	0.461	0.098	0.100	2.995	0.070	0.061	0.056
ME	0.505	0.199	0.305	0.095	0.047	2.792	0.066	0.046	0.052
SE	0.445	0.220	0.225	0.090	0.027	2.836	0.066	0.053	0.055

Table 3: Treatment effects on forum posting

	<i>Dependent variable:</i>						
	# posts (1)	# threads (2)	# replies (3)	# help-seeking (4)	# other threads (5)	#replies: hst (6)	#replies: other (7)
Large	0.524*** (0.134)	0.239 (0.151)	0.707*** (0.134)	0.049 (0.187)	0.331* (0.153)	1.075*** (0.251)	0.618*** (0.120)
Medium	0.168 (0.122)	0.066 (0.134)	0.245 (0.143)	-0.117 (0.185)	0.164 (0.136)	0.258 (0.262)	0.234 (0.130)
Participation req.	1.529*** (0.122)	1.643*** (0.138)	1.428*** (0.137)	1.196*** (0.193)	1.853*** (0.142)	1.543*** (0.255)	1.397*** (0.124)
Age	0.295*** (0.052)	0.323*** (0.067)	0.274*** (0.046)	0.259* (0.106)	0.359*** (0.066)	0.354*** (0.097)	0.255*** (0.049)
Constant	-0.925*** (0.134)	-1.705*** (0.151)	-1.533*** (0.134)	-2.413*** (0.187)	-2.322*** (0.153)	-3.495*** (0.252)	-1.673*** (0.120)
Observations	5,981	5,981	5,981	5,981	5,981	5,981	5,981
θ	0.082***	0.065***	0.081***	0.049***	0.067***	0.053***	0.086***

Note:

Estimates are from negative binomial regression models. *p<0.05; **p<0.01; ***p<0.001

Table 4: Logistic models of posting v. not posting

	<i>Dependent variable:</i>						
	any posts (1)	any threads (2)	any replies (3)	any help-seeking (4)	any other threads (5)	any replies: hst (6)	any replies: other (7)
Large	0.053 (0.073)	0.188 (0.133)	0.059 (0.077)	0.053 (0.178)	0.217 (0.129)	0.552** (0.210)	0.033 (0.074)
Medium	0.047 (0.078)	0.088 (0.105)	-0.013 (0.098)	-0.011 (0.134)	0.079 (0.088)	0.079 (0.210)	-0.022 (0.096)
Participation req.	0.632*** (0.074)	0.901*** (0.124)	0.636*** (0.093)	0.699*** (0.158)	1.075*** (0.119)	1.108*** (0.203)	0.656*** (0.089)
Age	0.242*** (0.034)	0.297*** (0.043)	0.218*** (0.040)	0.274*** (0.057)	0.326*** (0.045)	0.214** (0.073)	0.228*** (0.042)
Constant	-1.815*** (0.073)	-2.538*** (0.135)	-2.041*** (0.078)	-3.026*** (0.179)	-2.851*** (0.131)	-3.635*** (0.211)	-2.094*** (0.075)
Observations	5,981	5,981	5,981	5,981	5,981	5,981	5,981

Note:

*p<0.05; **p<0.01; ***p<0.001

People in the large cohort posted roughly 169% as often (i.e. 69% more often than) people in the small cohorts overall (95% confidence interval from cluster-robust standard errors (CI): 130%-219%), while people in the medium cohorts posted roughly 118% as often (CI: 93% - 150%). People posted replies to existing threads 203% as often in the large cohort relative to the small cohorts (CI: 156% - 264%), and 127% as often in the medium cohorts (CI: 97% - 169%). Of these replies, replies to help-seeking threads posted by other subjects almost tripled (relative to a very low baseline) in the large cohort to 293% of the rate in small cohorts (CI: 179% - 479%); in the medium cohorts, replies to help-seeking threads were posted at 129% of the rate in small cohorts (CI: 77%- 216%). Replies to other, non-help-seeking threads were posted at 186% of the small cohort rate in the large cohort (CI: 147%-235%) and 126% of the small cohort rate in the medium cohorts (CI: 98% - 163%).

Although posting of all new threads did not vary significantly by size (CI for the large cohort: 94% - 171%), posting of the subset of non-help-seeking threads was significantly different by treatment group: subjects in the large cohort posted at 139% (CI: 103% - 188%) and subjects in the medium cohorts posted at 117% (CI: 90% - 154%) of the rate of subjects in the small cohorts. The evidence is not strong for an overall causal effect of size on posting new threads, but we return to this point below as we investigate heterogeneous treatment effects.

In absolute terms, the greatest effect of size is in increasing replies, specifically replies to non-help-seeking threads. The 1997 Subjects in the large cohort posted 1416 times in total. Holding the effects of age constant, they would have posted 545 times fewer if they would have been divided into 16 small cohorts (rates are calculated by exponentiating the estimated coefficients or sum of coefficients to get the posting rate per person and then multiplying by the number of subjects). These 545 additional posts include approximately 444 more replies to existing threads (321 of these would have been replies to non-help-seeking threads), and only about 98 new threads (of these only 9 would have been new help-seeking threads)³.

Logistic regressions of dependent variables indicating whether subjects posted at least once overall or at least one of any of the sub-types of posts are in Table 4. Overall, size did not change the number of people posting, posting threads or posting replies. In larger cohorts, a greater fraction of subjects posted at least one reply to another subjects' help-seeking thread, but this is merely a change in the types of posts that people made, not a change the number of people posting replies overall.

Larger size does not have an overall causal effect on retention or performance, except that those in larger cohorts were more likely to pay \$25 to enter the id-verified track in anticipation of passing the course. In the large cohort, the odds of pre-paying for the id-verification were 128% times the odds in the small cohorts (CI: 102% - 162%), and in the medium cohort the odds of pre-paying were 109% of odds of pre-paying small cohorts (CI: 84% - 141%). Converting to probabilities, this amounts to a probability of pre-paying of 0.048 in small cohorts, 0.052 in medium cohorts, and 0.06 in the large cohort.

Requiring participation in the discussion forums had a substantial relative effect (if somewhat small absolute effect) on overall retention in the MOOC. In cohorts with required participation, subjects worked on 124% (CI: 112% - 138%) of the assignments that subjects in small cohorts did, the odds of pre-paying for the verified track were 132% (CI: 100% - 174%) of the odds in the small cohorts, and the

³ Estimates for different types of posts are calculated from different models and do not exactly add up to the estimate for the total number of posts.

odds of passing the course (i.e. achieving a grade greater than 60%) were 133% (CI: 113% - 157%) of the odds in small cohorts. In absolute terms, effects from requiring participation were small; subjects in these cohorts only earned 101.6% (CI: 100.7% - 102.4%) of the total course grade earned in participation encouraged cohorts.

As we discuss below, the overall causal effects can be better understood by examining how the treatments affected different people differently. Although there are substantial average effects of size and participation required, the fact is that the majority of enrolled subjects still had little or no activity registered on any dependent variable of interest. In that sense, average effects by themselves are somewhat misleading.

Table 5: Performance and retention

	<i>Dependent variable:</i>						
	# assignments <i>negative binomial</i>	grade <i>OLS</i>	grade/assignment <i>OLS</i>	any assignments <i>logistic</i>	Verified <i>logistic</i>	passed <i>logistic</i>	grade > 90% <i>logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Large	0.089 (0.055)	0.005 (0.004)	0.0001 (0.0002)	0.002 (0.059)	0.252* (0.119)	0.051 (0.080)	-0.118 (0.097)
Medium	0.042 (0.054)	0.003 (0.004)	0.00005 (0.0002)	0.002 (0.060)	0.089 (0.131)	0.036 (0.087)	0.071 (0.116)
Participation req.	0.215*** (0.054)	0.016*** (0.004)	0.001* (0.0002)	0.112 (0.061)	0.279* (0.141)	0.287*** (0.085)	0.174 (0.112)
Age	0.132*** (0.019)	0.011*** (0.002)	0.0003** (0.0001)	0.084*** (0.024)	0.109** (0.040)	0.211*** (0.037)	0.317*** (0.094)
Constant	1.002*** (0.055)	0.065*** (0.004)	0.004*** (0.0002)	-1.091*** (0.059)	-2.997*** (0.118)	-2.900*** (0.080)	-3.674*** (0.104)
Observations	5,981	5,981	5,981	5,981	5,981	5,981	5,981
θ	0.086***						

Note:

*p<0.05; **p<0.01; ***p<0.001

4.2 Descriptive results: motivations to contribute and treatment effects of size

The effects of size on posting behaviors are primarily due to an increase in activity among those who introduced themselves and to a much smaller extent those who watched a video immediately when the course started. The most socially active individuals contribute more per person in larger cohorts. Thus, there is a positive effect of size on forum posting, but it primarily affects the most socially engaged individuals by giving them more opportunity to be social. Size had a much smaller effect posting behavior on subjects who got a *quick start* but did not introduce themselves. We fit models including *quick start*, *introduced self*, and their interactions with treatment variables and each other. Effect sizes for

posting behavior are visualized in Figure 1 and parameter estimates are in Table 6. We omitted the age control variable from this series of models, as it is correlated with the endogenous motivation variables and further complicates the interpretation of the estimates.

After including covariates representing *quick start*, *introduced self*, and their interaction, the main effects of size on the total number of posts and comments are no longer significant at the 5% level. A look at Figure 1 gives a more holistic view: the majority of posts and the majority of the gain in posts due to size were associated with the more motivated subsets of subjects, especially those who introduced themselves. Additionally, although there was no main effect of size on the number of threads, subjects who introduced themselves posted almost twice as many threads in the large cohort as in the small cohorts.

Table 5 shows that evidence for an overall causal effect of size on performance and retention was not strong: only paying for verification increased with size. There were several significant heterogeneous treatment effects of size on performance on retention with a large relative magnitude but small absolute magnitude. The small minority of subjects who were positive on both *quick start* and *introduced self* were much more likely to pass the course in the large cohort than small cohorts (34.6% passed in large, 18.3% passed in small), but in absolute terms, this only amounted to a handful of people. Larger size seemed to change who achieved a grade of over 90% for the course: decreasing the fraction of subjects achieving this among those who were negative on both *quick start* and *introduced self*, but increasing it for subjects who were positive on one or both of those measures. Altogether for the large cohort, parameter estimates from Table 5, model 6 would translate into 15 fewer less-motivated people and 27 more more-motivated people achieving a grade of more than 90%, relative to the rates in the small cohorts.

Figure 1: visualization of treatment effects for different values of *quick start* and *introduced self*. Predicted values are average posts per-person.

None of this rules out the possibility of a true treatment effect on performance; it could be that with so many students not even attempting a single assignment, we just cannot detect the signal as statistically significant due to the number of people who were not engaged with the platform at all (and thus not strongly affected by the treatment). If we were to restrict our analysis to those who have attempted any assignments at all, for instance, then we would find that students in the Large cohort attempted 8% more assignments than those in the small cohorts ($p=.043$). Thus, we can say that larger

cohorts caused modestly more retention in the course for those who engaged with the course, but not more retention overall. We return to this issue below in section 4.4.

Table 6 also shows heterogeneous treatment effects of requiring participation. Required participation has negative interaction with the endogenous motivation parameters, but we interpret those to mean that the people who introduced themselves after being told they were required to participate were less motivated on average than those who introduced themselves without this requirement. Thus, it is not a surprising or particularly meaningful finding. Requiring participation also had heterogeneous effects on performance and retention. Like size, the positive effects seem to be mostly limited to motivated subjects; the main effect of requiring participation on performance and retention is insignificant, but interactions with measures of motivation are positive and significant.

Table 6: Heterogeneous treatment effects

	<i>Dependent variable:</i>						
	# posts <i>negative binomial</i>	# threads <i>negative binomial</i>	# replies <i>negative binomial</i>	# assignments <i>negative binomial</i>	verified <i>logistic</i>	passed <i>logistic</i>	grade >90% <i>logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Large	0.149 (0.192)	-0.088 (0.182)	0.401 (0.240)	0.062 (0.085)	0.266 (0.159)	-0.095 (0.142)	-0.604** (0.186)
Medium	-0.119 (0.175)	-0.113 (0.202)	-0.119 (0.224)	-0.002 (0.085)	-0.112 (0.163)	-0.005 (0.154)	-0.179 (0.233)
Participation req.(PR)	1.623*** (0.178)	1.506*** (0.189)	1.761*** (0.231)	0.061 (0.086)	0.327 (0.170)	0.011 (0.155)	-0.320 (0.257)
Quick start only (QSO)	1.506*** (0.280)	1.654*** (0.296)	1.265** (0.437)	1.291*** (0.281)	1.216** (0.418)	1.367* (0.551)	1.353* (0.678)
Introduced self only (ISO)	2.766*** (0.172)	2.269*** (0.257)	3.183*** (0.179)	1.573*** (0.147)	1.978*** (0.233)	1.909*** (0.303)	1.293** (0.398)
Quick start & intro. self (QS&IS)	3.349*** (0.287)	3.065*** (0.340)	3.655*** (0.310)	1.927*** (0.156)	2.247*** (0.600)	1.795*** (0.437)	1.880*** (0.485)
Large * QSO	0.661* (0.280)	0.444 (0.296)	0.956* (0.437)	0.127 (0.281)	-0.294 (0.418)	0.569 (0.551)	1.167 (0.678)
Medium * QSO	-0.018 (0.341)	-0.059 (0.359)	0.041 (0.445)	-0.051 (0.232)	0.610 (0.439)	-0.044 (0.457)	0.486 (0.585)
Large * ISO	0.547** (0.172)	0.662** (0.257)	0.361* (0.179)	0.080 (0.147)	0.219 (0.233)	0.246 (0.303)	1.073** (0.398)
Medium * ISO	0.442* (0.208)	0.297 (0.275)	0.510* (0.229)	0.073 (0.140)	0.394 (0.260)	-0.025 (0.298)	0.411 (0.396)
Large * QS&IS	0.375 (0.287)	0.191 (0.340)	0.337 (0.310)	0.034 (0.156)	0.196 (0.600)	0.857* (0.437)	0.907 (0.485)
Medium * QS&IS	0.174 (0.295)	0.185 (0.364)	0.138 (0.297)	-0.098 (0.184)	0.203 (0.534)	0.309 (0.399)	0.277 (0.344)
PR * QSO	0.026 (0.318)	0.054 (0.325)	0.047 (0.436)	0.069 (0.231)	-0.674 (0.416)	0.301 (0.473)	0.651 (0.582)
PR * ISO	-0.517** (0.194)	-0.074 (0.259)	-0.846*** (0.211)	0.181 (0.144)	-0.244 (0.267)	0.406 (0.299)	0.951* (0.425)
PR * QS&IS	-0.756* (0.300)	-0.512 (0.350)	-1.014** (0.325)	0.202 (0.187)	0.064 (0.531)	0.982* (0.433)	0.811 (0.421)
Constant	-1.756*** (0.192)	-2.314*** (0.182)	-2.610*** (0.240)	0.644*** (0.085)	-3.439*** (0.159)	-3.289*** (0.142)	-3.918*** (0.186)
Observations	5,981	5,981	5,981	5,981	5,981	5,981	5,981
θ	0.147***	0.107***	0.206***	0.096***			

Note:

*p<0.05; **p<0.01; ***p<0.001

4.3 Level versus type of motivation

We have been arguing that size causes socially responsive individuals to post more. One point of ambiguity in our argument is that introducing oneself is a costlier action than simply pressing play on an online video. Perhaps – a reasonable counter-argument might go – *introduced self* is just an indicator of more motivation, not social responsiveness. The coefficients in Table 6 do show that *introduced self* is associated with higher rate of course completion. By this counter-argument, size merely causes more motivated individuals to post more and there is nothing especially “socially responsive” about the mechanism of contribution increase (setting aside, for the time being the fact that the majority of the increase in posting activity is in the form of replies rather than new posts).

Figure 2: visualization of treatment effects for different values of *quick start* and *introduced self* for subjects who attempted at least one assignment. Predicted values are average posts per-person.

There is something to this critique, but further examination supports the notion that *introduced self* captures not only slightly higher motivation, but indeed a different type of motivation. If we restrict our examination to those individuals who have done at least one assignment, then the difference in number of assignments completed between *quick start* and *introduced self* becomes negligible. Among those who did at least one assignment, when we compare subjects with *quick start* only to subjects with *introduced self* only, there are no significant differences in the proportions of subjects who remained engaged past the first, second, third, fourth, fifth or sixth module. T-tests for difference in means between the number of assignments and the $\log(\text{number of assignments})$ were both insignificant, and quantile regressions for difference in 10th, 50th and 90th percentiles of the distribution of number of assignments completed were all insignificant (N.B. although there were no differences between subjects with only *quick start* and those with only *introduced self*, subjects who were positive on both variables stayed engaged with the course significantly longer and those who were negative on both stayed engaged significantly shorter, than those who were positive on only one of those variables, according to all tests above.). Restricting our analysis to this subset with a more balanced level of motivation to complete the course, the effect of size still works through those who introduced themselves: Figure 2 visualizes effect sizes for the subset of users who have attempted at least one assignment.

4.4 Forum moderators and performance implications

Despite the large effect of size on posting, it is notable that there were not more marked effect of size on performance and retention. Our evidence suggests that the members of the teaching team who were forum moderators made up for the lack of user-generated content posted by the learners taking the MOOC (i.e. our experimental subjects) by responding to more help-seeking threads, suggesting that this equalized performance in the smaller cohorts to a great extent.

Forum moderators treated each question equally regardless of which forum it originated in. If a question already had a suitable answer, there was no need for moderators to add one, but if it had not been answered, moderators would step in to help. This design (rather than, say, allocating a fixed number of moderator answers per person to each cohort) was driven by the teaching team’s commitment to ensuring that learners had an equal chance to master the material, regardless of the experimental condition they were in.

Table 7 gives descriptive statistics for moderator responses to help-seeking threads. Larger cohorts had a greater percentage of help-seeking threads that had replies from other learners but not the moderators (see Table 7: “fraction with replies but no mod reply”). The difference between the LE cohort (with 24.6%) and the SE cohorts (with 12.2%) was statistically significant (Chi-squared = 4.0945, $df = 1, p = 0.04$).

Table 7: Moderator replies to help-seeking threads by treatment

	LE	ME	SE	MR	SR
fraction with no moderator (mod) reply	0.342	0.277	0.278	0.417	0.365
fraction with replies but no mod reply	0.246	0.181	0.122	0.241	0.202
fraction of mod replies within 4 hours	0.426	0.471	0.477	0.457	0.516
mean(log(mod 1st response time))	5.669	5.426	5.437	5.571	5.323
sd(log(mod 1st response time))	2.025	1.996	2.048	2.083	1.930

Among help-seeking threads with a moderator reply, there were no significant differences between moderator first response times by treatment, or between or the fraction of help-seeking-threads that received a moderator reply within four hours. There are slight trends toward smaller cohorts receiving more rapid replies but these are statistically insignificant.

In all, moderators treated questions equally, but this resulted in subjects in larger cohorts having more of their questions answered by other subjects and smaller cohorts having more of their questions answered by moderators in the absence of help from peers.

The different ratio of moderator replies to peer replies in the different treatments raises an alternative interpretation of the findings to be ruled out. Specifically, one might imagine that in the smaller cohorts, given the greater proportion of replies by moderators to requests for help, people may learn over time to diminish their own responsive behavior in expectation that moderators would reply. Thus a norm may develop that exaggerates the estimate of the treatment effect beyond what would occur without moderators.

Many forums do have official moderators who take varying degrees of responsibility for responding to requests for help; for such forums, the behavior of the moderators in our data is a feature that increases external validity, not a bug. For forums without such moderators however, different levels of moderator involvement by forum may introduce some bias in treatment estimates.

We reason that a norm of waiting for the moderators could explain different levels of responsiveness for those replies that occurred prior to a moderator response to a thread; that is, one might expect fewer replies in the (smaller) cohorts, where people were waiting for a moderator reply. However, such a norm of deferral to or crowding out by the moderators would not explain any differences in replies posted on threads to which a moderator has posted the first reply; if the moderator has already replied, people are no longer waiting for a moderator reply. We estimated a model of treatment effect on replies to peers' help seeking threads to which a moderator has provided the first reply, so we could compare the coefficients to the results presented in Table 3.

Estimates for the effects of the treatments on replies to peer's help-seeking threads to which a moderator has already responded are almost identical to the full data. The negative binomial estimate for the effect of large size was 1.004 (compare to 1.075 for the full data), and the estimate for the effect of medium size was 0.241 (compare to 0.258 for the full data). Thus, while the moderators' involvement may have biased our estimates, the overall bias is slight at most.

5. DISCUSSION AND CONCLUSION

5.1 Summary of experimental findings

We find that larger cohorts cause more posting per person but do not cause more people to post. Our strongest specific result was that larger cohort size causes more replies to existing posts made per person. This effect did not apply to all subjects equally: almost all of the effect was on more motivated subjects – particularly those that were socially responsive. Since more motivated people post more, this means that larger size causes the greatest absolute increase in posting among those who post most.

We did not find clear evidence that size had an overall effect on grade or retention, but we did see a larger number of users pay for the ID verified enrollment track, indicating a greater investment in the course among larger cohorts. We also find evidence of some modest heterogeneous treatment effects such that the most motivated people stayed longer in larger cohorts. The limited effects of treatments on performance may have been because the forum moderators for this MOOC were dedicated to giving each student an equal chance at learning the material, and ended up answering more posts in smaller cohorts. In this way, they equalized access to help to a large extent. If subjects had to rely only on user-generated content for help, we posit that there would have been significant overall performance implications of size.

Echoing prior research (e.g. Zalmanson and Oestreicher-Singer, 2015), simply stating that participation was required led to large and statistically significant increases in forum contribution across all variables. It also caused a small but significant increase in average grade and number of assignments completed.

5.2 Discussion

5.2.1 Distinction between social responsiveness and resource availability and network effects

Previous literature has identified theories that are related to social responsiveness but substantially different. Specifically, the theories of network effects (Katz and Shapiro, 1994) and

resource availability (Butler, 2001) predict that a greater number of users on a platform at time 1 would attract more users to the platform at time 2, because a platform with more users provides more resources and indeed value to newcomers. However, neither perspective makes a clear prediction about how much people contribute per person in cross section: these theories predict a virtuous cycle between content and number of people using the platform, but they do not tell us how much people post on a per person basis. In fact, Kane and Ransbotham (2016) provide evidence that such a virtuous cycle would result in less contribution per person, even as the number of users increased, as the content on the platform became more complete and refined. Social responsiveness does not contradict the theories of network effects or resource availability, but rather complement them by predicting the cross sectional, heterogeneous effects of size.

5.2.2 Practical applications

In other settings than online forums, smaller groups have resulted in more participation per person, implying the possibility that more engagement might be possible if people were divided up into more intimate cohort sizes. But in our experiment larger cohorts encourage more posting and reduce attrition from the platform as a whole. For a given population of forum users, the greatest contribution of posts per person would happen if the whole population shared a single forum, rather than being divided up into multiple smaller forums. Future research should investigate how this effect interacts with the diversity of topics discussed in a single forum; it could be that including discussion of unrelated topics in a single forum might increase or suppress engagement, separately from the effects of size.

Our evidence does not support the idea that size changes the number of people who participate versus lurk. If there is an effect, it is too small to be statistically significant in our data. People have diverse reasons for their participation or lack of participation online. We argue that prior literature shows that those who do participate have motivations that are essentially social and thus result in greater participation per person when they are part of a larger cohort. However, for people without these social

motivations, increasing size may act as a further barrier to posting. To increase participation among those without socially-responsive motivations, other incentives, conditions or rules must be provided. For example, 24.3% of subjects posted in our required participation treatment, compared with only 14.7% in the participation encouraged treatment.

However, an implication of social responsiveness driving participation is that inequality of participation rates – the so-called “90-9-1” rule -- may be very difficult to change. Even requiring participation did not reduce inequality among those who did post. Indeed, the Gini coefficient for posts per person in the SR treatment (0.567) was actually higher than the Gini coefficient in the SE treatment (0.495); similarly, the Gini coefficient in MR (0.584) was higher than the Gini coefficient in ME (0.503). Inequality of participation is not just due to the individual attributes of the people attracted to the platform; it is due to a combination of those attributes and the interaction of people on the platform. The heaviest participators are especially socially responsive, so eliciting more contribution from lower-participating individuals provides the socially-responsive high-contributors even more to respond to, maintaining the unequal distribution of posting.

5.3 Limitations and future extensions

With a single MOOC and maximum cohort size equal to 2000 people, we cannot say much the functional form of size’s effects on posting. Over the range that we measure, the total effect appears to be approximately linear or slightly sub-linear, but it could well be that over another order of magnitude of sizes that the effects of size would be attenuated more markedly, as the capacity for any individual to respond to the forum posts of others is not infinite.

A closely related point is that our data prevent us from distinguishing between specific types of social responsiveness. For example, we cannot tell the difference between people helping each another for pure altruistic reasons and helping each other because they want to achieve higher status within the

community. If different psychological mechanisms underlying social responsiveness scaled differently with respect to forum size, a better understanding of these might help predict the functional form of the effect of size on engagement across more orders of magnitude.

In our analyses, we consider heterogeneous treatment effects to make our arguments. In so doing, we use indicators of the level and type of motivation that are measured post-treatment. Our causal results show that larger size causes a set of people who reply to a request to introduce themselves to reply more to others' posts. We cannot say for certain whether the people who introduce themselves in a large cohort are drawn from the same population as the people who introduce themselves in a small cohort.

As such, we cannot distinguish between two competing explanations for our experimental results. Our interpretation is that socially responsive people respond to requests to introduce themselves and also respond to other people's threads. Being in a group with many people gives socially-responsive people more to respond to. Non-socially responsive people remain low-engagement regardless of size. The competing explanation is that different people become socially responsive in different sized cohorts. Some people respond a lot when they are in the presence of many peers but not when they are in the presence of few peers. Other people respond a little when they are in the presence of few peers but not when they are in the presence of many peers.

The key practical difference is that this competing explanation implies that there is a subset of people who would engage more if only there were in a smaller group, which seems quite plausible. This plausible implication, however, comes at the expense of the hard-to-swallow requirement that the more socially responsive people who respond (a lot) in large groups would stop engaging in small groups. We are also not aware of any evidence in prior literature for people who only engage in a certain group size and whose level of engagement is correlated with that group size. Size does not have a statistically significant relationship with the incidence of *quick start* or *introduced self* per cohort, and Chi-squared

tests of covariates associated with posting behavior (age >50, *goals: statistics*, recent registration on edX) show no significant differences between who introduces themselves in different treatments, which supports our interpretation further. Having said all of that, even if the competing explanation were in fact the true mechanism behind the results, one could almost interpret it in the same way as we describe in the paper: people who respond a lot to the presence of many people could almost be called “socially responsive.”

A related and mostly untested implication of the literature is that large size might further reduce the engagement of people without social motivations. We do find evidence that in the large cohort there was a shift in who completed the course to achieve a grade of over 90% toward the most motivated and away from the less-motivated individuals. However, this relative shift is based on a small absolute number of people, and should be replicated before concluding there is a real effect. A different study design – one with greater ability to measure the motivations and attributes of experimental subjects before treatment – would be necessary to resolve the above limitations of the present study. Specifically, future research should measure social responsiveness and other motivations prior to experimental treatment. Survey methods would be helpful, but nonresponse bias is a serious concern in low-engagement settings, so great care (and possibly substantial financial incentives) may be required to obtain a valid sample. While laboratory experiments offer greatest control and opportunity to measure individual subjects’ motivations, a field experiment would be necessary to obtain a population of participants with the real motivations that drive behavior.

We have examined a setting in which motivations to participate and contribute are at the individual level, and thus there is no free-riding or similar dilution of effort due to the juxtaposition of a collective incentive and individual effort (see section 2.2). However, some settings provide a mix of individual and collective incentives or motivations. For example, in open source programming communities, there is collaboration on a single collective product, but individual contributors have

diverse personal incentives, such as wanting to use the software themselves, career signaling, social capital, or just being creative and having fun (Lerner and Tirole, 2002; Lakhani and Wolf, 2005). Future research would be necessary to establish how reductions in effort due to size in collective collaborative tasks would combine with increases in effort due to size's interactions with social motivations.

Presumably, the net effect would be to offset each other to some degree, but it is not clear a priori which effect would dominate, whether there would be substantial implications for inequality of participation, or how situational factors might change outcomes in unexpected ways. Thus we urge caution before directly extrapolating the present results to more collaborative or competitive settings. A further complication is that the presence of monetary incentives can dramatically alter motivations and behavior (Li and Zhang, 2016). Future work should work to shed further light on settings with mixed attributes to better understand the effect of size on contribution with different incentive structures.

5.3.1 Performance and retention

The present manuscript is focused on understanding the effects of size on posting per person to better understand the drivers of contribution of information and knowledge online. Our results do not show any significant effect of size on performance and retention, which is somewhat surprising, given the size of the effects on posting behavior. With the evidence we have, we cannot say for sure why this was the case. Section 4.4 did provide evidence that the moderators may have equalized learning opportunities across sizes, and Appendix 3, below provides suggestive evidence of an effect on retention among those that did engage. However, people who enroll in MOOCs are diverse in their approach to and goals for learning, and it is difficult to analyze outcomes without more granular measures of learner attributes (Littlejohn et al., 2016). Thus, future research, employing better exogenous measures of user motivations and abilities, is necessary to dig deeper into the causal effects of size on retention and knowledge-transfer.

5.4 Conclusion

Online communities and discussion forums are pervasive and important sites for the creation and transfer of knowledge. Size was known to be a major driver of participation across different contexts, but with diverging and unclear predictions for online discussion. Here we show that the effect of size is to increase posting per person on average, but not the fraction of people that post. Thus, the effect is mostly on the most motivated and social people. By giving this subset of individuals more to respond to, they contribute more per person. An implication is that the commonly observed inequality of participation online cannot be easily changed: any intervention that increases participation among low-contributors will elicit a concomitant increase in engagement from the greatest contributors.

REFERENCES

- Aaltonen, Aleksi, and Stephan Seiler. "Cumulative growth in user-generated content production: evidence from Wikipedia." *Management Science* 62, no. 7 (2015): 2054-2069.
- Ackerman, Mark S. "Augmenting organizational memory: a field study of answer garden." *ACM Transactions on Information Systems (TOIS)* 16, no. 3 (1998): 203-224.
- Albanese, R., & Van Fleet, D. D. (1985). Rational behavior in groups: The free-riding tendency. *Academy of Management review*, 10(2), 244-255.
- Alnuaimi, O. A., Robert, L. P., & Maruping, L. M. (2010). Team size, dispersion, and social loafing in technology-supported teams: A perspective on the theory of moral disengagement. *Journal of Management Information Systems*, 27(1), 203-230.
- Ashraf, N., Bandiera, O., & Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior & Organization*, 100, 44-63.
- Bales, R., & Borgatta, E. (1966). Size of group as a factor in the interaction profile. In A. P. Hare, E. Borgatta, & R. Bales (Eds.), *Small groups: Studies in social interaction* (pp. 495-512). New York: Knopf
- Barron, G., & Yechiam, E. (2002). Private e-mail requests and the diffusion of responsibility. *Computers in Human Behavior*, 18(5), 507-520.
- Bateman, P. J., Gray, P. H., & Butler, B. S. (2011). Research note-the impact of community commitment on participation in online communities. *Information Systems Research*, 22(4), 841-854.
- Beaudoin, Michael F. "Learning or lurking?: Tracking the "invisible" online student." *The internet and higher education* 5, no. 2 (2002): 147-155.
- Beenen, Gerard, Kimberly Ling, Xiaoqing Wang, Klarissa Chang, Dan Frankowski, Paul Resnick, and Robert E. Kraut. "Using social psychology to motivate contributions to online communities." *Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW)*, pp. 212-221. ACM, 2004.

- Boudreau, K. J., Lacetera, N., & Lakhani, K. R. (2011). Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science*, 57(5), 843-863.
- Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty, and sex on group performance and member reactions. *Journal of Personality and Social Psychology*, 36(11), 1224.
- 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.
- Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of business research*, 66(1), 105-114.
- Brooks, Frederick Phillips (1975) *The mythical man-month: essays on software engineering* Reading, Massachusetts: Addison Wesley
- Budhathoki, N. & Haythornthwaite, C. (2013). Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist*, 57(5), 548 – 575
- Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2017). Stimulating online reviews by combining financial incentives and social norms. *Management Science*, 64(5), 2065-2082.
- Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, 12(4), 346-362.
- Butler, Brian, Lee Sproull, Sara Kiesler, and Robert Kraut (2002). "Community effort in online groups: Who does the work and why." *Leadership at a distance: Research in technologically supported work*, S. Weisband & L. Atwater (Eds.): 171-194.
- Cameron, A. C., and Trivedi, P. 2013. *Regression Analysis of Count Data* (Vol. 53), Cambridge, UK: Cambridge University Press
- Carillo, K., & Okoli, C. (2011). "Generating quality open content: A functional group perspective based on the time, interaction, and performance theory." *Information & Management*, 48(6), 208-219.
- Chidambaram, L., & Tung, L. L. (2005). "Is out of sight, out of mind? An empirical study of social loafing in technology-supported groups." *Information Systems Research*, 16(2), 149-168.
- Chiu, Chao-Min, Meng-Hsiang Hsu, and Eric TG Wang. "Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories." *Decision support systems* 42, no. 3 (2006): 1872-1888.
- Coetzee, D., Fox, A., Hearst, M. A., & Hartmann, B. (2014). Should your MOOC forum use a reputation system? In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 1176-1187). ACM.
- Constant, David, Lee Sproull, and Sara Kiesler. "The kindness of strangers: The usefulness of electronic weak ties for technical advice." *Organization Science* 7, no. 2 (1996): 119-135.
- Di Gangi, P. M., Wasko, M. M., & Hooker, R. E. (2010). Getting customers' ideas to work for you: learning from dell how to succeed with online user innovation communities. *MIS Quarterly Executive*, 9(4).
- Diehl, M., & Strobe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, 53, 497-509.
- 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(4), 668-684.
- Farzan, Rosta, Joan M. DiMicco, David R. Millen, Casey Dugan, Werner Geyer, and Elizabeth A. Brownholtz. "Results from deploying a participation incentive mechanism within the enterprise." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 563-572. ACM, 2008.
- Gallupe, R. B., Dennis, A. R., Cooper, W. H., Valacich, J. S., Bastianutti, L. M., & Nunamaker, J. F. (1992). Electronic brainstorming and group size. *Academy of Management Journal*, 35(2), 350-369.

- Gooding, R. Z., & Wagner III, J. A. (1985). A meta-analytic review of the relationship between size and performance: The productivity and efficiency of organizations and their subunits. *Administrative Science Quarterly*, 462-481
- Hood, N., Littlejohn, A., & Milligan, C. (2015). Context counts: How learners' contexts influence learning in a MOOC. *Computers & Education*, 91, 83-91.
- Huang, Ni, Yili Hong, and Gordon Burtch. "Social network integration and user content generation: evidence from natural experiments." *MIS Quarterly* 41, no. 4 (2017): 1035-1058.
- Hwang, E. H., Singh, P. V., & Argote, L. (2015). Knowledge sharing in online communities: Learning to cross geographic and hierarchical boundaries. *Organization Science*, 26(6), 1593-1611.
- Ingham, Alan G., George Levinger, James Graves, and Vaughn Peckham. "The Ringelmann effect: Studies of group size and group performance." *Journal of Experimental Social Psychology* 10, no. 4 (1974): 371-384.
- Jeppesen, L. B., & Frederiksen, L. (2006). Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization Science*, 17(1), 45-63.
- Jones, G. R. (1984). Task visibility, free riding, and shirking: Explaining the effect of structure and technology on employee behavior. *Academy of Management Review*, 9(4), 684-695.
- Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research*, 15(2), 194-210.
- Kane, Gerald C., and Sam Ransbotham. "Content as community regulator: The recursive relationship between consumption and contribution in open collaboration communities." *Organization Science* 27, no. 5 (2016): 1258-1274.
- Kankanhalli, A., Tan, B. & Wei, K. (2005). Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly*, 29(1), 113-143.
- Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *American Economic Review*, 75(3), 424-440.
- Katz, M. L., & Shapiro, C. (1994). Systems competition and network effects. *The journal of economic perspectives*, 8(2), 93-115.
- Kerr, N. L. (1989). Illusions of efficacy: The effects of group size on perceived efficacy in social dilemmas. *Journal of experimental social psychology*, 25(4), 287-313.
- Kerr, N. L., & Bruun, S. E. (1981). Ringelmann revisited: Alternative explanations for the social loafing effect. *Personality and social psychology bulletin*, 7(2), 224-231.
- Kerr, N. L., & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and social Psychology*, 44(1), 78.
- Kim, J. (2013). Influence of group size on students' participation in online discussion forums. *Computers & Education*, 62, 123-129.
- Koh, J., Kim, Y. G., Butler, B., & Bock, G. W. (2007). Encouraging participation in virtual communities. *Communications of the ACM*, 50(2), 68-73.
- Lakhani, Karim R., and Robert G. Wolf. (2005). "Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects." *Perspectives on Free and Open Source Software*, Feller, Fitzgerald, Hissam, and Lakhani, Eds., Cambridge: MIT Press
- Lampel, Joseph, and Ajay Bhalla. "The role of status seeking in online communities: Giving the gift of experience." *Journal of computer-mediated communication* 12, no. 2 (2007): 434-455.
- Latane, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of personality and social psychology*, 10(3), 215.
- Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of personality and social psychology*, 37(6), 822.

- Lerner, Josh, and Jean Tirole (2002). "Some simple economics of open source." *The Journal of Industrial Economics* 50, no. 2: 197-234.
- Li, Xitong, and Jiayin Zhang (2016), "The effects of monetary incentives and social comparison on MOOC Participation: a Randomized Field Experiment", Proceedings of the *International Conference on Information Systems*, Dublin, Ireland, Dec, 2016.
- Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education*, 29, 40-48.
- Lowry, P. B., Roberts, T. L., Romano, N. C., Cheney, P. D., & Hightower, R. T. (2006). The Impact of Group Size and Social Presence on Small-Group Communication Does Computer-Mediated Communication Make a Difference?. *Small Group Research*, 37(6), 631-661.
- Ma, M., & Agarwal, R. (2007). Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information systems research*, 18(1), 42-67.
- Mao, A., Mason, W., Suri, S., & Watts, D. J. (2016). An experimental study of team size and performance on a complex task. *PloS one*, 11(4), e0153048.
- Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. "How conditioning on posttreatment variables can ruin your experiment and what to do about it." *American Journal of Political Science* (2016).
- Nonnecke, Blair, Dorine Andrews, and Jenny Preece. "Non-public and public online community participation: Needs, attitudes and behavior." *Electronic Commerce Research* 6, no. 1 (2006): 7-20.
- Nonnecke, B., & Preece, J. (2001). Why lurkers lurk. *Americas Conference on Information Systems* 2001.
- Oestreicher-Singer, Gal, Lior Zalmanson. 2013. Content or community? a digital business strategy for content providers in the social age. *MIS Quarterly* 37 591–616.
- Pendharkar, P. C., & Rodger, J. A. (2009). The relationship between software development team size and software development cost. *Communications of the ACM*, 52(1), 141-144.
- Preece, Jenny, Blair Nonnecke, and Dorine Andrews. "The top five reasons for lurking: improving community experiences for everyone." *Computers in human behavior* 20, no. 2 (2004): 201-223.
- Qiu, M., Hewitt, J., & Brett, C. (2012). Online class size, note reading, note writing and collaborative discourse. *International Journal of Computer-Supported Collaborative Learning*, 7(3), 423-442.
- Ray, Soumya, Sung S. Kim, and James G. Morris. "The central role of engagement in online communities." *Information Systems Research* 25, no. 3 (2014): 528-546.
- Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization studies*, 28(3), 377-408.
- Riopelle, K., Gluesing, J. C., Alcordo, T. C., Baba, M., Britt, D., McKether, W., Monplaisir, L., Ratner, H.H., and Wagner, K. H. (2003). Context, task, and the evolution of technology use in global virtual teams. *Virtual teams that work: Creating conditions for virtual team effectiveness*, 239-264.
- Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., & Sherer, J. (2014, March). Social factors that contribute to attrition in MOOCs. In *Proceedings of the first ACM conference on Learning@ scale conference*(pp. 197-198). ACM.
- Sun, Na, Patrick Pei-Luen Rau, and Liang Ma. "Understanding lurkers in online communities: A literature review." *Computers in Human Behavior* 38 (2014): 110-117.
- Thomas, E. J., & Fink, C. F. (1963). Effects of group size. *Psychological Bulletin*, 60(4), 371-384
- Valacich, J. S., Dennis, A. R., & Nunamaker, J. F. (1992). Group size and anonymity effects on computer-mediated idea generation. *Small Group Research*, 23(1), 49-73.
- Valacich, J.S.; Wheeler, B.C.; Mennecke, B.E.; and Wachter, R. The effects of numerical and logical group size on computer-mediated idea generation. *Organizational Behavior and Human Decision Processes*, 62, 3 (1995), 318–329.

- Vuong, Q.H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57:307-333.
- Wang, X., Butler, B. S., & Ren, Y. (2013). The impact of membership overlap on growth: an ecological competition view of online groups. *Organization Science*, 24(2), 414-431.
- Wasko, M. M., and S. Faraj. "'It is what one does': why people participate and help others in electronic communities of practice." *The journal of strategic information systems* 9, no. 2-3 (2000): 155-173.
- Wasko, M. M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS quarterly*, 35-57.
- Wheelan, S. A. (2009). Group size, group development, and group productivity. *Small Group Research*.
- Wiertz, Caroline, and Ko de Ruyter. "Beyond the call of duty: Why customers contribute to firm-hosted commercial online communities." *Organization studies* 28, no. 3 (2007): 347-376.
- Yap, Lee-Xian, & Gilbert Bock. (2006). Investigation of the Effects of Members' Closeness Size and IT in Virtual Communities: The Social Network Perspective. *AMCIS 2006 Proceedings*, 526.
- Zalmanson, Lior and Oestreicher-Singer, Gal, 'Your Action is Needed': The Effect of Website-Initiated Participation on User Contributions to Content Websites (May 2015). Available at SSRN: <https://ssrn.com/abstract=2606236>
- Zhang, X., & Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *The American Economic Review*, 101(4), 1601-1615.

Appendix: Additional details, results, and alternative model specifications

A1. MODELS WITH EXOGENOUS COVARIATES

Our main series of models in the main paper include only the treatment variables and age. Random assignment resulted in exogenous covariates other than age being balanced across treatments. In this appendix, we report a series of models that include these covariates to check the results for robustness to alternate specifications. To limit over-fitting, we used 5-fold cross-validation and included only those covariates that reduced cross-validation error. Results are in Tables A1 and A2. Candidate covariates for inclusion in the model included main effects and interactions with treatment variables.

Table A1: Forum participation with covariates selected by cross-validation

	<i>Dependent variable:</i>					
	# posts <i>negative binomial</i>	# threads <i>negative binomial</i>	# replies <i>negative binomial</i>	any posts <i>logistic</i>	any threads <i>logistic</i>	any replies <i>logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Large	0.459*** (0.110)	0.151 (0.128)	0.695*** (0.118)	0.059 (0.069)	0.202 (0.132)	0.065 (0.076)
Medium	0.077 (0.125)	-0.043 (0.130)	0.175 (0.153)	0.058 (0.073)	0.109 (0.104)	-0.008 (0.096)
Participation req.	1.674*** (0.158)	1.808*** (0.174)	1.581*** (0.161)	0.661*** (0.069)	0.950*** (0.130)	0.667*** (0.093)
age	0.318*** (0.066)	0.325*** (0.079)	0.406*** (0.058)	0.029 (0.094)	-0.098 (0.120)	0.266*** (0.030)
age ²				0.249*** (0.070)	0.387*** (0.088)	
weeks since EdX sign-up	-0.323*** (0.061)	-0.400*** (0.087)	-0.219 (0.164)	-0.244*** (0.047)	-0.200** (0.065)	-0.274*** (0.045)
(weeks since EdX sign-up) ²			-0.152 (0.161)			
ln(days since course sign-up)				-0.192*** (0.028)	-0.158*** (0.033)	-0.195*** (0.032)
EdX and course sign-ups coincided		-0.306* (0.121)				
goals: statistics	0.292 (0.186)	0.320 (0.239)	0.406*** (0.120)	0.566*** (0.113)	0.321 (0.185)	0.592*** (0.123)
goals: baseball	0.410 (0.217)	0.566* (0.242)			0.232 (0.146)	
no goals given					-0.287** (0.106)	
real name	0.013 (0.099)	0.048 (0.168)	0.018 (0.062)		-0.149 (0.079)	
female						0.254** (0.086)
Participation req X real name	-0.487** (0.177)	-0.501** (0.168)	-0.474** (0.153)			
Participation req X age			-0.310*** (0.088)	-0.149** (0.057)	-0.150* (0.068)	-0.179** (0.065)
Constant	-1.007*** (0.118)	-1.695*** (0.164)	-1.636*** (0.123)	-1.943*** (0.071)	-2.479*** (0.169)	-2.210*** (0.080)
Observations	5,981	5,981	5,981	5,981	5,981	5,981
θ	0.086***	0.069***	0.086***			

Note:

All continuous covariates are centered and scaled to unit variance. *p<0.05; **p<0.01; ***p<0.001

Overall results are generally comparable to the series of models in the main paper, but the effects of size are slightly smaller. Users in the large cohort posted at 158% the rate of those in the small cohorts (compared to 169% estimated without covariates). For replies the estimate was nearly identical: the estimated posting rate in the Large cohort with covariates is 200% of the rate in the small cohorts (compared to 203% without covariates).

The main effect of requiring participation on the number of posts, threads and replies appears higher with covariates. However, there is a countervailing negative interaction effect between requiring participation and *real name* (an indicator variable that is positive when the name used by a subject to register for edX matches the screen name that displays to other users when they make comments on the forum). In other words, requiring participation is not as effective in causing people who use their real names to participate, on average. We interpret this as further evidence in favor of prior literature's (Nonnecke and Preece 2001) claim that concern for privacy can act as a barrier to participation: some of those who created a screen name using personally identifying information were less willing to comply with a stated requirement to participate.

Another consistent pattern of results is that the longer ago people registered for edX, the less likely they were to participate heavily. This could be for many different reasons, for example, long-time users may register for new courses more readily (and without as much motivation to learn about the specific subject), but further research would be necessary to determine the reason. Additionally, people who registered for the course long before it began were also less likely to participate at all, perhaps because they lost interest or forgot about the course by the time it started. Finally, people who signed up for edX and the course within 7 days of each other posted markedly fewer new threads. Prior literature finds that newcomers are hesitant to post while they learn about a community (Preece et al., 2004).

Those with a stated goal related to statistics, Sabermetrics or analytical software (*goals: statistics*) were more likely to participate and participate more. Those with a goal related to baseball posted many more threads. Finally, female participants were more likely to post at least one comment.

Table A2: Performance and retention with covariates selected by cross-validation

	<i>Dependent variable:</i>					
	# assignments	grade	any assignments	verified	passed course	grade > 0.90
	<i>negative binomial</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>logistic</i>	<i>logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Large	0.041 (0.051)	0.006 (0.004)	0.005 (0.058)	0.646*** (0.147)	0.056 (0.079)	-0.118 (0.095)
Medium	-0.001 (0.053)	0.003 (0.004)	0.009 (0.059)	0.319* (0.143)	0.039 (0.086)	0.094 (0.113)
Participation req.	0.195*** (0.053)	0.016*** (0.004)	0.107 (0.060)	0.271 (0.140)	0.283*** (0.084)	0.177 (0.112)
age	-0.220*** (0.051)	-0.017*** (0.004)	-0.783*** (0.176)	-0.161* (0.063)	0.200*** (0.036)	0.607*** (0.079)
age^2	0.359*** (0.045)	0.028*** (0.004)	0.687*** (0.134)			
No age given			-0.768** (0.235)	-0.909*** (0.232)		1.444*** (0.204)
weeks since EdX sign-up	-0.266*** (0.069)	-0.017*** (0.005)	-0.116 (0.125)	0.263 (0.182)	-0.292** (0.113)	
(weeks since EdX sign-up^2)			-0.133 (0.102)	-0.805** (0.258)		-0.450** (0.160)
ln(days since course sign-up)	-0.086* (0.035)	-0.006 (0.003)	-0.156*** (0.030)	-0.136** (0.047)		
EdX and course sign-ups coincided	-0.279** (0.092)	-0.022** (0.009)	-0.127 (0.113)		-0.437** (0.148)	-0.513 (0.281)
goals: statistics	0.388*** (0.083)	0.035*** (0.009)	0.405*** (0.073)	0.563*** (0.156)	0.402** (0.132)	0.728*** (0.191)
goals: baseball				0.804*** (0.210)		
no goals given				-0.074 (0.202)		0.231 (0.148)
Large X goals: baseball				-0.775*** (0.194)		
Medium X goals: baseball				-0.946*** (0.213)		
Large X no goals given				-0.663*** (0.183)		
Medium X no goals given				-0.185 (0.230)		
Constant	1.062*** (0.057)	0.069*** (0.005)	-1.009*** (0.074)	-3.189*** (0.162)	-2.807*** (0.092)	-4.012*** (0.167)
Observations	5,981	5,981	5,981	5,981	5,981	5,981
θ	0.088***					

Note:

All continuous covariates are centered and scaled to unit variance. *p<0.05; **p<0.01; ***p<0.001

Size did not have a substantial effect on the majority of performance and retention variables, and controlling for covariates does not change this. However, size did affect the number of people choosing to purchase an ID-verified certificate; controlling for covariates paints a more complex picture. To begin with, the main effects of size are larger with controls. Converting parameter estimates to

probabilities, the model without covariates estimates that 4.8% of users in small cohorts and 6% of users in the large cohort purchased certificates. Adding controls, we estimate that 4% of users in small cohorts and 7.3% of users in the large cohort purchased certificates. However, we also find countervailing negative effects. Users with no goals given (about 50% of all users) purchased the certificate at about the same rate, regardless of size. Users with a goal related to baseball (only about 7% of all users) were more likely to purchase the certificate in the small cohorts than the medium or large cohorts. Taken together with results from the main paper, the picture is one of heterogeneous treatment effects. Large size leads to more engagement and greater value of the course for people who are engaged, especially so for those with socially-responsive motivations. Those with different motivations (learning about baseball, but not statistics) may even have gotten less value out of larger cohorts where there was more discussion.

A2. ALTERNATIVE MODEL SPECIFICATIONS

Tables A3, A4 and A5 report models of treatment effects on the main forum participation variables estimated with different functional forms. Vuong's test indicates that negative binomial regressions fit these variables best, but to demonstrate that our conclusions are robust to alternate specifications, we include them here.

There are slight differences in parameter estimates, but all models show large and significant increase in total posts, and replies, with a smaller, mostly insignificant increase in new threads. Parameters estimated by OLS imply a larger effect of size than we see in our preferred negative binomial (NB) models. NB estimates of the effect of large size relative to small size are 169%, 127% and 203% for posts, threads, and replies, respectively; with OLS, the estimates are 190%, 149% and 225% for posts, threads, and replies, respectively. OLS estimates are expected to be biased upwards in the presence of overdispersion.

Quasipoisson (QP) models estimate slightly more posting in the smaller cohorts and have generally smaller standard errors, leading to an impression of greater precision than NB models. Estimate magnitudes for medium and large cohorts vary relative to the NB models, however. Estimates are smaller for the large cohort for total posts and replies, but larger for the number of new threads, even giving a significant estimate for the effect of large size on posting of new threads. For medium-sized cohorts, the pattern is different: estimates are larger for total posts and new threads, but smaller for replies.

Zero-inflated NB (ZINB) models are harder to interpret, as there is a parameterized zero-inflation component to the model in addition to the count model. For total posts, the ZINB model gives a larger probability of a subject in the large cohort having zero posts as part of the zero-inflation part of the model, but also a larger number of posts in the count part of the model. For new threads, that pattern is reversed (lower estimate for the zero-inflation component and lower estimate in the count model). The zero component of the ZINB model of replies shows evidence of misspecification, with extremely large estimates and standard errors, so we will not interpret that model.

In sum, different model specifications differ slightly from the best-fitting NB model, but are consistent with respect to the overall magnitude of effects. The only change in interpretation that might occur if we used another type of model is related to new threads. The quasipoisson model indicates that the effect of large size on the number of new threads is distinguishable from zero, while the NB model does not. In any case, our analyses in section 4.3 in the main paper suggests that the answer depends strongly on the motivations of the subject in question, so the statistical significance of the main effect is of secondary importance.

Table A3: Comparison of model types: total posts

	<i>negative binomial</i>	<i>ZINB:</i>		<i>quasipoisson</i>	<i>OLS</i>
		<i>count</i>	<i>zero</i>		
	(1)	(2)	(3)	(4)	(5)
Large	0.524*** (0.134)	0.687*** (0.145)	0.446 (0.340)	0.487*** (0.103)	0.340*** (0.098)
Medium	0.168 (0.122)	0.176 (0.119)	0.022 (0.170)	0.173 (0.102)	0.214 (0.157)
Participation req.	1.529*** (0.122)	1.410*** (0.122)	-0.258 (0.396)	1.430*** (0.097)	1.522*** (0.156)
age	0.295*** (0.052)	0.197*** (0.053)	-0.315 (0.199)	0.277*** (0.047)	0.287*** (0.039)
Constant	-0.925*** (0.134)	-0.459 (0.561)	-0.598 (1.525)	-0.866*** (0.106)	0.374*** (0.098)
Observations	5,981	5,981	5,981	5,981	5,981
θ	0.082***				

Note:

*p<0.05; **p<0.01; ***p<0.001

Table A4: Comparison of model types: # threads

	<i>negative binomial</i>	<i>ZINB:</i>		<i>quasipoisson</i>	<i>OLS</i>
		<i>count</i>	<i>zero</i>		
	(1)	(2)	(3)	(4)	(5)
Large	0.239 (0.151)	0.090 (0.111)	-0.243 (0.168)	0.243* (0.106)	0.082 (0.044)
Medium	0.066 (0.134)	0.030 (0.114)	-0.118 (0.176)	0.151 (0.078)	0.089 (0.064)
Participation req.	1.643*** (0.138)	1.232*** (0.112)	-0.639 (0.507)	1.529*** (0.100)	0.765*** (0.064)
Age	0.323*** (0.067)	0.158* (0.074)	-0.336* (0.164)	0.298*** (0.052)	0.138*** (0.022)
Constant	-1.705*** (0.151)	-0.624 (1.114)	0.646 (1.785)	-1.677*** (0.109)	0.168*** (0.044)
Observations	5,981	5,981	5,981	5,981	5,981
θ	0.065***				

Note:

*p<0.05; **p<0.01; ***p<0.001

Table A5: Comparison of model types: # replies

	<i>negative binomial</i>	<i>ZINB:</i>		<i>quasipoisson</i>	<i>OLS</i>
		<i>count</i>	<i>zero</i>		
	(1)	(2)	(3)	(4)	(5)
Large	0.707*** (0.134)	1.179*** (0.143)	8.552 (13.321)	0.650*** (0.116)	0.257*** (0.066)
Medium	0.245 (0.143)	0.389** (0.138)	5.796 (12.022)	0.193 (0.147)	0.125 (0.117)
Participation req.	1.428*** (0.137)	1.526*** (0.157)	2.108 (1.873)	1.344*** (0.116)	0.757*** (0.115)
Age	0.274*** (0.046)	0.254*** (0.047)	-0.200 (0.124)	0.260*** (0.045)	0.149*** (0.025)
Constant	-1.533*** (0.134)	-1.585*** (0.145)	-9.210 (13.349)	-1.456*** (0.118)	0.206** (0.065)
Observations	5,981	5,981	5,981	5,981	5,981
θ	0.081***				

Note:

*p<0.05; **p<0.01; ***p<0.001

Table A6 reports parameter estimates analogous to those in Table 3 in the main paper, but estimated only on the subset of cohorts in the treatment encouraged condition. Results are very close to the full data, with the exception of the estimates for treatment effects on the number of non-help-seeking threads. For this latter dependent variable, there does not appear to be a significant effect, but as we noted in the main text, the evidence is not strong for an effect of size on new-thread posting. For the number of replies (the overall number and the number of replies to non-help-seeking threads), that is more central to the theory of social responsiveness, medium cohorts are found to have significantly more forum posting than the small cohorts.

Table A7 reports estimates from models that combine small and medium cohorts together for the purposes of estimation. These models are a response to the unbalanced design of the experiment (see Table 1). The LE treatment had 2000 subjects assigned to it, which is twice the number that all of the other treatment groups. Conceivably, positive parameter estimates for the large cohort could be the result of sampling a skewed distribution more heavily and ending up with better representation of the tail of the distribution. By combining the small and medium cohorts together, the LE cohort would be compared against comparable numbers of subjects in the small/medium participation encouraged and small/medium participation required groups. Estimates for the large size are as expected in these models: slightly smaller than the main series of models (because now the reference category includes heavier-posting medium cohort subjects), but still in the same direction and statistically significant.

Table A6: Treatment effects on forum posting, participation encouraged only

	<i>Dependent variable:</i>						
	# posts (1)	# threads (2)	# replies (3)	# help-seeking (4)	# other threads (5)	#replies: hst (6)	#replies: other (7)
Large	0.546** (0.171)	0.188 (0.182)	0.789*** (0.172)	0.162 (0.251)	0.186 (0.168)	1.235** (0.379)	0.692*** (0.150)
Medium	0.178 (0.181)	-0.084 (0.216)	0.375* (0.174)	0.036 (0.324)	-0.163 (0.172)	0.553 (0.423)	0.342* (0.149)
Age	0.384*** (0.076)	0.437*** (0.090)	0.359*** (0.068)	0.424** (0.147)	0.468*** (0.053)	0.469*** (0.070)	0.338*** (0.087)
Constant	-0.952*** (0.171)	-1.667*** (0.182)	-1.622*** (0.171)	-2.547*** (0.251)	-2.202*** (0.168)	-3.693*** (0.383)	-1.752*** (0.148)
Observations	3,985	3,985	3,985	3,985	3,985	3,985	3,985
θ	0.076*** (0.004)	0.058*** (0.005)	0.075*** (0.005)	0.048*** (0.006)	0.061*** (0.006)	0.042*** (0.006)	0.080*** (0.006)

Note:

*p<0.05; **p<0.01; ***p<0.001

Table A7: Treatment effects on forum posting, small and medium cohorts combined

	<i>Dependent variable:</i>						
	# posts (1)	# threads (2)	# replies (3)	# help-seeking (4)	# other threads (5)	#replies: hst (6)	#replies: other (7)
Large	0.436*** (0.092)	0.208 (0.110)	0.573*** (0.103)	0.108 (0.158)	0.255** (0.098)	0.930*** (0.208)	0.492*** (0.092)
Participation Req.	1.529*** (0.125)	1.647*** (0.136)	1.418*** (0.151)	1.199*** (0.197)	1.870*** (0.139)	1.527*** (0.267)	1.390*** (0.137)
Age	0.293*** (0.051)	0.323*** (0.066)	0.271*** (0.044)	0.254* (0.107)	0.355*** (0.067)	0.352*** (0.097)	0.252*** (0.047)
Constant	-0.837*** (0.091)	-1.674*** (0.110)	-1.398*** (0.102)	-2.472*** (0.157)	-2.245*** (0.098)	-3.350*** (0.209)	-1.546*** (0.091)
Observations	5,981	5,981	5,981	5,981	5,981	5,981	5,981
θ	0.082*** (0.003)	0.065*** (0.003)	0.081*** (0.004)	0.049*** (0.004)	0.067*** (0.004)	0.052*** (0.005)	0.085*** (0.004)

Note:

*p<0.05; **p<0.01; ***p<0.001

A3. SURVIVAL ANALYSIS OF USERS WHO WATCHED AT LEAST ONE VIDEO

Figure A1 shows a Kaplan-Meier survival plot, such that a subject is considered to have “died” on the day that they watch their last video lecture, unless that day was after the date that the final module was released, in which case they are considered to have “survived” the whole course. The analysis is limited to the subset of subjects who watched at least one video, and thus describes the effect of the treatment on duration of engagement for those who engaged. Median engagement increases with size and the participation requirement; the differences are suggestive, but not statistically significant. Like the other post-treatment subset analyses, the results should not be extrapolated back to an overall causal effect.

Figure A1: Kaplan-Meier survival plot of course video watching for subjects who interacted with at least one video. Median survival times are indicated with the dotted lines and increase in SE, ME, LE, SR and MR order.

A4. DETAILS OF TREATMENT

Learners received messages by email at the launch of each module and on the Courseware pages (at the end of the assignments portion of each of 6 modules) informing and remind them of their discussion forum group. The text below is what a learner in the SE (small, treatment encouraged) condition received:

This year, we highly encourage all participants to interact with their fellow learners on the forum. Throughout the Courseware, you will see places where we ask you to post to the forum to interact with your group. You are currently in a group of 125 other learners, and your posts will only be seen by your group members (and the staff of course!). Please use the discussion forum link below to post something about this track. You may start a new thread or respond to a fellow learners thread; the only recommendation is that you post something about the course using the link below.

For comparison, the text below is what a learner in the MR condition received:

This year, we will also be requiring students to post to the Discussion Forum as part of their course grade. Throughout the Courseware, you will see places where we ask you to post to the forum to interact with your group. You are currently in a group of 500 other learners, and your posts will only be seen by your group members (and the staff of course!). Please use the discussion forum link below to post something about this track. You may start a new thread or respond to a fellow learners thread; the only recommendation is that you post something about the course using the link below.

After completing your post in the discussion forum using the link below, please check the following box to receive credit for your post.

For students in the participation required conditions, the above language was then followed by a check box as depicted in Figure A2, below

Figure A2: check box for learners in the participation required treatments.

A5. COVARIATE BALANCE

Subjects were placed into cohorts uniformly at random (covariates were not available at the time of randomization so blocking was impossible). To assess possible covariate imbalance, we compared standardized means (for continuous covariates) and raw means (for dichotomous covariates) between each treatment pair (e.g. the proportion of people reporting a male gender in LE was compared to the proportion in each of ME, SE, MR, and SR) and the largest between-treatment difference was recorded.

We then compared these differences in the real data to 1000 permutations of the data (effectively re-randomizations) to assess how severe the imbalance was, as realized in the data. As noted in the main text, the age variable had the largest imbalance, and thus we control for it in parametric models. Figure

A3 plots the observed imbalance against the distribution of random imbalances. Other than age, all other variables are within expected values.

Figure A3: observed imbalance (solid line) plotted against imbalance from permuted data. The 90th percentile of the distribution derived from permuted data is shown with a dashed line.

With 6000 observations, ten covariates and 5 treatment groups, some level of imbalance is expected and normal. To illustrate, we randomly permuted (i.e. re-randomized) the treatment groups of our participants 1000 times and calculated the maximum between-treatment difference in standardized means (for the continuous variables) or raw means (for the dichotomous variables). We then calculated the quantiles of these maximum between-cohort differences to establish a distribution of expected imbalance. For reference, the out-of-balance “age” covariate in the real data is greater than 907 of 1000 re-randomizations – it is a 90.7th percentile imbalance, and all other covariates were less than or equal to the 75th percentile imbalance from our permuted data.

Of the 1000 permuted runs only 394 runs had no covariate imbalance greater than the 90th percentile of expected imbalance and 271 runs had two or more covariates greater than the 90th percentile

of expected imbalance. Thus our real data, with one covariate exceeding the 90th percentile of expected imbalance, was typical of possible randomization outcomes.