

Forum size and content contribution per person: a field experiment

Jiye Baek, Hong Kong University of Science and Technology
Jesse Shore, Boston University

1. SUMMARY

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. Prior literature on size is inconclusive, providing evidence for both positive [Zhang and Zhu 2011; Huang et al. 2017] and negative [Latané et al. 1979; Mao et al. 2016] effects of size, with the specifics of prior study settings introducing further possible confounds. We conducted a randomized field experiment on a massive open online course (MOOC), and 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.

2. METHODS

We conducted a field experiment in Boston University's 'Sabermetrics 101' course offered on edX, which began on July 7th, 2015 and ran for 10 weeks. 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.

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. 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. 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.

To investigate mechanisms behind the causal effects in the experiment, we examine two endogenous variables that help illustrate 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 second endogenous variable we examine is denoted *introduced self*, which indicates whether course participants responded to a forum moderator request to introduce themselves in the week preceding the official course launch. Neither *quick start* nor *introduced self* is significantly correlated with size, according to Chi-squared tests.

3. RESULTS

3.1 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.

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%). Parameter estimates from negative binomial models of posting behaviors with cluster-robust standard errors are in Table 1.

Table I. : 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 with cluster robust standard errors. *p<0.05; **p<0.01; ***p<0.001

3.2 Heterogeneous effects of treatments

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.

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.

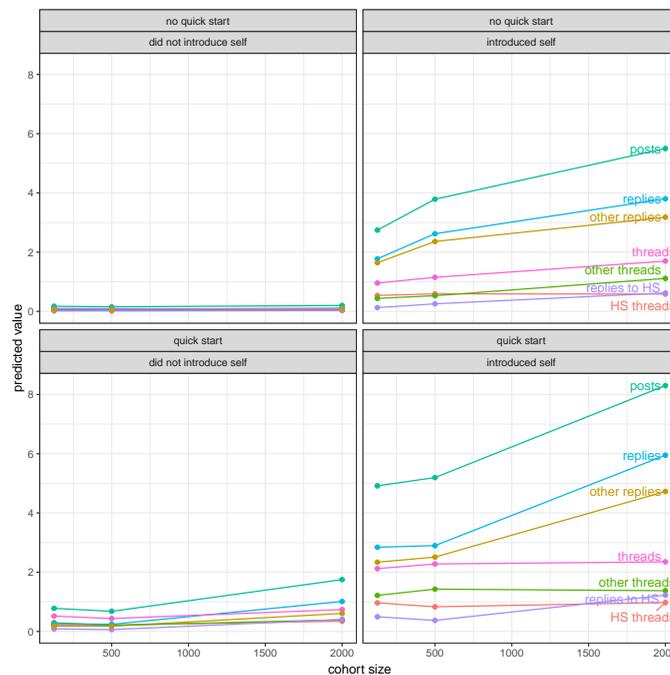


Fig. 1: visualization of treatment effects for different values of *quick start* and *introduced self*

4. IMPLICATIONS FOR INEQUALITY OF PARTICIPATION

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, in the small treatment, the Gini coefficient for posts per person in the participation required treatment (0.567) was actually higher than the Gini coefficient in the participation encouraged treatment (0.495); similarly, in the medium treatment, the Gini coefficient in the participation required treatment (0.584) was higher than that in the participation encouraged treatment (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.

REFERENCES

- Doug Clow. 2013. MOOCs and the funnel of participation. In *Proceedings of the third international conference on learning analytics and knowledge*. ACM, 185–189.
- Ni Huang, Yili Hong, and Gordon Burtch. 2017. Social network integration and user content generation: Evidence from natural experiments. *MIS Quarterly: Management Information Systems* 41, 4 (2017), 1035–1058.
- Bibb Latané, Kipling Williams, and Stephen Harkins. 1979. Many hands make light the work: The causes and consequences of social loafing. *Journal of personality and social psychology* 37, 6 (1979), 822.
- Andrew Mao, Winter Mason, Siddharth Suri, and Duncan J Watts. 2016. An experimental study of team size and performance on a complex task. *PloS one* 11, 4 (2016), e0153048.
- Xiaoquan Michael Zhang and Feng Zhu. 2011. Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review* 101, 4 (2011), 1601–15.