The social dimension of participation and completion in MOOCs

This is the author's manuscript

Original Citation:
The social dimension of participation and completion in MOOCs / Goglio V., Parigi P.. - (2018), pp. 85-89. ((Intervento presentato al convegno Learning with MOOCs tenutosi a Madrid (ES) nel 26-28 settembre 2018.

Availability:
This version is available http://hdl.handle.net/2318/1677700 since 2018-10-05T17:07:49Z

Publisher:
IEEE

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
The social dimension of participation and completion in MOOCs

Valentina Goglio  
Department of Cultures, Politics and Society  
University of Turin  
Turin, Italy  
Institute for Research in the Social Sciences  
Stanford University  
Stanford (CA), USA  
valentina.goglio@unito.it

Paolo Parigi  
Lead Trust Scientist  
Airbnb  
San Francisco (CA), USA  
Institute for Research in the Social Sciences  
Stanford University  
Stanford (CA), USA  
pparigi@stanford.edu

Abstract— The rapid and impressive development of MOOCs in less than half a decade has brought about contrasting arguments about the social dimension of MOOCs. This paper investigates how the socio-economic background of learners affects their own experience in the on-line course and their chances of course completion. We test whether learners from low SES have lower chances of completing the on-line course and whether participation in online discussion forums mediates the role of SES. Analyzing data from two MOOCs provided by Stanford University, we find that in both cases the hypothesis of a negative association between low SES and course completion is supported. Moreover, we find that forum participation further reinforces the advantage of well-educated learners enrolled in one case, while does not have any significant effect for the other course. We conclude with some design implications for creators of on-line courses.

Keywords— MOOCs, SES, completion, participation, online forum

I. INTRODUCTION

Between 2008 and 2012, the outburst of Massive Open Online Courses (MOOCs) in the USA created great excitement about their potential, placin MOOCs among the key challenges for the future of HE. Now, the media hype that characterized the earlier period seems to be gone (own elaboration based on Factiva). Notwithstanding, the number of registered users to the main MOOCs platforms is impressively high and keeps growing [1]. MOOCs growing enrollment has brought about contrasting arguments. Enthusiastic views argue that these resources have opened up a series of opportunities for reducing social inequalities and promoting growth and employment. On the other side, skeptical views have questioned the actual empowering and equalizing effect of MOOCs. In particular, scholars point to a gap in literature about the social dimension of MOOCs learners, as very little is known about MOOCs’s access and returns to education. There is still limited empirical research on dropouts, and overall there are not enough empirical data to reach a definitive conclusion [2]–[4].

The objective of this paper is to investigate the social dimension of MOOCs, namely how the socio-economic background of learners affects their own experience in the on-line course and their chances of course completion. Given the high numbers of enrollments, the composition of MOOCs’ learners’ body is heterogeneous under several aspects. Different motivations, socio-economic backgrounds and learning patterns drive the learning process and produce different outcomes. The empirical research we present here investigates only the role of socio-economic background on the participation to learning activities and on the completion of the MOOC. Other factors may be relevant and will require further investigation.

II. THEORETICAL BACKGROUND

Previous research has highlighted that MOOCs are characterized by massive attrition rates [5] and that MOOCs participants tend to be well-educated individuals living in developed countries, mainly English-speaking [3], [6], [7]. Yet some authors maintain that the motivations of each learner can be very different and with such high numbers of enrollment, even a small percentage of completion may result in a large number of successful completers [8]. While on one side there is a growing body of literature on the pedagogical aspects of online learning, the social and equitable dimension of MOOCs is far less investigated. The way in which socio-economic background affects the participation to course activities and on the completion of the on-line course remains quite under-investigated.

We hypothesize that a low socio-economic background has a direct negative association with completion. Further, we hypothesize that participation in the course activity has an indirect relationship to course completion. The theoretical model we propose is represented in fig. 1. As it happens for the more standard in-presence education, the low socio-economic status of the learner negatively effects the completion of the particular type of online courses [9]–[11]. This relationship is further mediated by the level of participation of the learner. Indeed, research on in-presence education showed that participation to class activities is both negatively associated to socio-economic background (students from lower SES interact less than peers from high SES) and positively associated to successful completion of the course (students who participate more have higher chances of completing successful than their peers). We assume that this relationship holds true in the digital environment of MOOCs as well [12], and hypothesize that the relationship between low socio-economic background and completion of MOOCs is also mediated by participation in forum discussions. We expect to find that learners from low socio-economic background have lower chances of completing the course attaining a certificate; moreover, we expect to find that learners from low SES also show lower chances of participation to online forum discussion and this further reinforces their disadvantage on course completion.
III. DATA AND METHOD

The data used in this paper come from two different MOOCs provided by Stanford University through Lagunita, a Stanford instance of the OpenEdX platform. We selected two courses: ‘Statistical Learning’ and ‘America’s Poverty and Inequality Course’ based on their different field of study and on the relatively high number of forum posts available. We use demographic and performance data of learners registered, as well as survey data for those who completed the pre-course survey. The data come from the Winter 2016 edition of Statistical Learning and Fall 2016 edition of Poverty and Inequality. The empirical strategy uses quantitative methods of analysis. We investigate the direct and indirect relationship between socio-economic background and MOOCs completion using logistic regression analysis.

We use the certificate of completion obtained at the end of the online course in order to build our dependent variable ‘completion’. This may not be the perfect variable for completion, as not all learners may be interested in downloading the certificate at the end of the course [13], [14]. An advantage of using the completion certificate is that such information is available for both courses included in our analysis; the certificate is provided free, downloadable from the course website, although the rules change slightly: learners must have earned at least 50% of graded assignments for Statistical Learning, while at least 75% for the Poverty and Inequality course. This difference can lead to a more restricted sample of completers for the second course considered. The key independent variable, socio-economic background, refers to the economic and cultural resources available to the learner determined by both his family of origin and his current situation. Typically, literature on social stratification refers to level of education and occupational category of parents and of the individual, as well as family and personal income as proxies for SES. However, information about parents’ education is only available for the course on Poverty and Inequality. Thus, we opted for using the following proxies for the socio-economic status of the learner that were available for both on-line courses: self-declared employment status and level of education of the learner. For the first course, Statistical Learning, we also add information on ethnicity. The relative role of ethnicity can change, given that learners can take the course from whatever country in the world. We combined the information on ethnicity with the place of birth and operationalized this variable equal to 1 if the individual belongs to the prevalent ethnicity in the country of birth (reliable information on the country of residence is not available). Regarding the educational level of learners, the distribution of educational titles is skewed toward high qualifications as typical of MOOCs [6], and thus we set the reference category to associate degrees or lower. Regarding employment, we used involuntary exclusion from the labor market (unemployment) as indicator of a disadvantaged condition. The proxy for ‘participation’ in the course—the mediating factor in our model—is defined here as participation to the online discussion forum of the course. It is a dummy variable equal to 1 if the learner wrote at least one forum post. Finally, we include controls for sex and age. We run separate models for the two courses, due to the different sample size, definition of the dependent variable ‘completion’ and of the other independent variables coming from survey data.

IV. FINDINGS

Table 1 provides a summary of the characteristics of learners in the two courses. The sample size of both demographic and survey data varies widely between the two courses, with Statistical Learning attracting a much higher number of learners than Poverty and Inequality (25 times more learners). On the other hand, Poverty and Inequality course shows better engagement proxies with a higher proportion of: i) respondents to the pre-course survey, ii) certificates attained (despite the more restrictive rule for the Statement of Accomplishment) and iii) written online forum posts. As far as the other variables are concerned, women are underrepresented in the Statistical Learning course, while the distribution of educational qualifications is stable. Employed and unemployed people are equally represented in the two courses, while Statistical Learning has a higher proportion of students and Poverty and Inequality a higher proportion of retired people. Overall, the estimates from the logistic regressions show a robust and significant direct effect of socio-economic background on completion, but a less straightforward indirect relationship mediated by forum participation. First, we analyze the findings for the Statistical Learning course (Fig. 2). The condition of unemployment, contrary to what hypothesized, shows a positive association with completion. This may suggest that unemployed learners may be more motivated in gaining and updating skills and may view or use the certificate of completion as a signaling tool for increasing their employment chances. Their potentially disadvantaged condition seems to works as an incentive for an instrumental use of the course. The results for the educational attainment support the hypothesis of a relative disadvantage of learners with low educational qualifications: indeed, learners with graduate education (master or PhD) have higher chances of completing the course compared to their peers with no college degree. Finally, we find that in some particular contexts belonging to the prevalent ethnicity is negatively associated to the completion of the course. Namely, results for being Black in Africa suggest that this population is indeed disadvantaged even in the MOOCs context.

We then test whether participation to the online forum has a mediating role on this relationship between SES and completion (Fig. 3). The interaction models show that participation to online discussions, by writing at least one forum post, does not have any mediating role for unemployed people. This indicates that participating in the forum activity does not make any difference on the chances...
of completing the course for unemployed people; the direct
effect of their employment situation on the chances of
completing the course is stronger than the mediated
effect. On the other hand, participation to forum discuss-
ion seems to be beneficial for learners with already higher level of
education (higher than Associate). This positive mediating
role of participation to forum adds up to their original
relative advantage observed in the direct effect, further
reinforcing the advantage of well-educated learners, instead
of representing a tool for the engagement of learners from
low SES. Finally, participation seems to even further
worsening the disadvantage of middle-eastern learners,
whose participation to forum is negatively associated to
completion.

With respect to the second course analyzed, Poverty and
Inequality, the estimates support our hypothesis with some
interesting peculiar features. Indeed, learners from better off
socio-economic background (as guessed by their
occupational status) have a relative advantage in terms of
completion (Fig. 4). People involuntarily excluded from the
labor market (unemployed) do not have different chances of
completing the course, compared to their employed peers.
Yet, this holds true for a residual category, which we can
reasonably consider representing purely inactive people (not
seeking job, not in education). In this case the instrumen-
tal value of the course seems to fail in favor of a more value-
oriented approach or recreational function of the course
(also supported by the fact that when excluding the control
for age, retired learners are positively associated to course
completion). With respect to the mediating role (Fig. 5),
forum participation does seem to further reinforce the
relative advantage of inactive people. Summing up,
evidence suggests that learners who can afford not to be
in the labor market, have a relative advantage in completing
the course compared to their less advantage peers. Moreover,
participation to online forum discussion further enhances
their chances of completion.

As a final step, we run robustness checks in order to detect
possible bias due to collinearity or specification error, and a
goodness of fit test. All the tests support and confirm the
results of the models presented in Figures 2 to 5.

<table>
<thead>
<tr>
<th>TABLE I. SAMPLE CHARACTERISTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>total N</td>
</tr>
<tr>
<td>took certificate of completion</td>
</tr>
<tr>
<td>wrote at least one forum post</td>
</tr>
<tr>
<td>wrote post &amp; took certificate</td>
</tr>
<tr>
<td>no. posts written</td>
</tr>
<tr>
<td>demographic data</td>
</tr>
<tr>
<td>age</td>
</tr>
<tr>
<td>female</td>
</tr>
<tr>
<td>educational attainment</td>
</tr>
<tr>
<td>less than BA</td>
</tr>
<tr>
<td>Bachelor</td>
</tr>
<tr>
<td>Doctorate</td>
</tr>
<tr>
<td>Masters/prof. degree</td>
</tr>
<tr>
<td>Withheld</td>
</tr>
<tr>
<td>survey data</td>
</tr>
<tr>
<td>total N</td>
</tr>
<tr>
<td>employment status</td>
</tr>
<tr>
<td>FT employed</td>
</tr>
<tr>
<td>FT employed</td>
</tr>
<tr>
<td>student</td>
</tr>
</tbody>
</table>
In this paper we analyzed the relationship between SES and MOOC completion in order to test whether some of the forms of social inequalities observed in the domain of traditional in-person education transfer to the digital environment of MOOCs. We also tested whether this disadvantage could be mediated (positively or negatively) by the participation to discussions on the online forum of the course, with the aim of detecting whether participation through these tools may leads to different outcomes compared to participation in in-person education.

Overall, findings from our models show that in the two courses analyzed, learners with higher SES have more chances of completing the course reaching a certificate of completion, thus supporting the hypothesis of a negative (direct) association between low SES and course completion. Moreover, participation to online forum discussions has an ambiguous mediating role. Indeed, the participation to online forum further reinforces the relative advantage of already highly educated learners (for the Statistical Learning course), and of inactive learners (in the case of Poverty and Inequality course). Finally, it is worth noting that the course on Statistical Learning seems to play an instrumental value for learners. Indeed, we can argue that the higher chances of completion for people excluded from the labor market suggests that learners use this course for their own re-skilling and as a signaling tool for potential employers. On the contrary, it can be argued that the course on Poverty and Inequality may play more a recreational than instrumental function, although with the data at hand we are not able to define any further the function assigned by learners to the course.

An implication for course design that can be drawn from these results regards online forum participations. Indeed, if participation seems to further reinforce the advantage of well-educated people or of those who can afford not being in the labour market, a simple intervention in the course design aimed at incentivizing participation from all learners may contribute to reshape the trend. Previous research shed light on some particular types of learners who tend to monopolize the online forum discussion, marginalizing other learners and making it difficult for them to participate [15]. Yet, participation to forum can be incentivized by making it a substantial part of the total grading; by making it more attractive and user-friendly; by providing regular message alerts; by regulating the maximum length or number of posts per learner (in order to avoid ‘super-posters’) or even by making it mandatory for attaining the certificate.

This paper provides a first exploration of the social dimension of MOOCs, which is often overlooked in empirical research. It also presents some limitations that will need to be addressed in future research: results are limited to two courses although in different fields of study; more comprehensive results can be reached by including a higher number of courses and disciplines. Moreover, the topic of participation to online forum can be further investigated by including information on the actual content of the posts written by learners.

ACKNOWLEDGMENT

The authors thank CAROL (Center for Advanced Research through Online Learning) for the provision of data. We thank Mitchell Stevens for his support and suggestions.

REFERENCES


This project has received funding from the H2020-MSCA-IF-2016 programme under grant agreement No 750242.


