IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**The social dimension of participation and completion in MOOCs**

**This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available http://hdl.handle.net/2318/1677700 since 2020-03-19T11:22:24Z

*Publisher:*

IEEE

*Published version:*

DOI:10.1109/LWMOOCS.2018.8534620

*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)

12 October 2024

**This is a preliminary version. To quote refer to:**
**Goglio V., Parigi P. 2018 "The social dimension of participation and completion in MOOCs"**
**LWMOOCS 2018 Conference Proceedings, IEEE (ISBN 978-1-5386-6533-6)**

# *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 brought about contrasting arguments and scholars acknowledge that there is a gap in literature about the social dimension of MOOCs. Very little is known about the dimension of institutional consequences of MOOCs as well as social implications in terms of access and returns to education. The objective of this paper is to investigate the social dimension of MOOCs, by focusing on how the socio-economic background of learners affects their own experience in the course and their chances of course completion. We test whether learners from low SES have lower chances of completing the course and whether participation to online discussion forums may mediate such a negative association, or rather reinforce their original disadvantage. Analyzing data from two MOOCs provided by Stanford University, we find that in both cases the hypothesis of a negative association between low SES background 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.**

***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, putting them among the key challenges for the future of HE. The hype that characterized the earlier period seems to be gone, MOOCs do not longer enjoy the same attention on media that they had in 2012-2013 (own elaboration based on Factiva) however, the number of registered users to the main MOOCs platforms is impressively high and keeps growing [1]. This phenomenon brought about contrasting arguments. Enthusiastic views welcomed these resources opening up a series of opportunities for reducing social inequalities and promoting growth and employment. On the other side, skeptical views 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 the dimension of institutional consequences of MOOCs as well as social implications in terms of 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 course and their chances of course completion.

With such high numbers of enrollments, the composition of MOOCs' learners' body is heterogeneous under several aspects. Different motivations, socio-economic backgrounds and learning patterns drive them through the learning process and lead them to different outcomes. The empirical research we present here investigates the role of socio-economic background on the participation to learning activities and on the completion of the MOOC.

## II. Theoretical Background

Previous research has highlighted that MOOCs are characterized by massive attrition rates [5] and MOOCs tend to fail the target population they originally aimed to: participants tend to be polarized into 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 course remains quite under-investigated.

The main research question of the paper addresses the direct and indirect effect of the socio-economic background on the completion of the MOOC course. We hypothesize that a low socio-economic background has a direct negative association with completion, which is mediated by the level of participation in the course activity (indirect relationship). The theoretical model we propose here is represented in fig. 1. The main idea is that, as it happens in in-presence education, the completion of the particular type of online courses as MOOCs is negatively affected by the low socio-economic status of the learner [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 to 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.

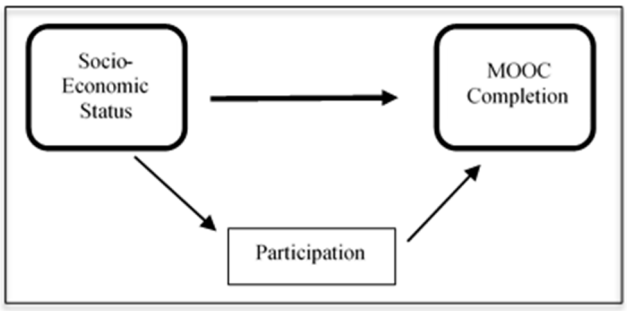


Fig. 1. Theoretical background

## 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.

In this work the key concepts considered are operationalized on the basis of previous research but also on data availability, due to the constrains posed by data. For the dependent variable 'completion' we rely on the certificate of completion obtained at the end of the course. 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, but it still can be considered as a good proxy[13], [14]. Moreover, this type of information is available for both courses; 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 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. However, this information is only available for the course on Poverty and Inequality, and thus we decided to flank it with other proxies for the socio-economic status of the learner. For both courses, we considered: self-declared employment status and level of education of the individual. 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 thus 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 discussion 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 instrumental 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 being 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 I. SAMPLE CHARACTERISTICS

| | **Statistical Learning** | | **Poverty and Inequality** | |
|---|---|---|---|---|
| | ***N*** | ***%*** | ***N*** | ***%*** |
| total N | 72,854 | | 2,908 | |
| took certificate of completion | 4491 | 6.16 | 358 | 12.31 |
| wrote at least one forum post | 990 | 1.36 | 397 | 13.65 |
| wrote post & took certificate | 499 | 0.68 | 209 | 7.18 |
| | Mean | StdDev | Mean | Std Dev |
| no. posts written | 3.48 | 11.39 | 6.21 | 13.23 |
| *demographic data* | Mean | Std Dev | Mean | Std Dev |
| age | 31.41 | 9.53 | 40.65 | 14.67 |
| | N | % | N | % |
| female | 13,937 | 19.13 | 1,609 | 55.33 |
| educational attainment | | | | |
| less than BA | 6,035 | 8.28 | 298 | 10.25 |
| Bachelors | 23,481 | 32.23 | 763 | 26.24 |
| Doctorate | 7,155 | 9.82 | 373 | 12.83 |
| Masters/prof. degree | 25,781 | 35.39 | 1,173 | 40.34 |
| Withheld | 9,180 | 12.60 | 292 | 10.04 |
| *survey data* | | | | |
| total N | 10,787 | | 982 | |
| employment status | | | | |
| FT employed | 6,619 | 64.52 | 633 | 69.48 |
| PT employed | 662 | 6.45 | | |
| student | 2,106 | 20.53 | 85 | 9.33 |
| unemployed | 766 | 7.47 | 50 | 5.49 |
| retired | 106 | 1.03 | 88 | 9.66 |
| other | n.a. | n.a. | 55 | 6.04 |
| belongs to prevalent ethnicity | 8,182 | 81.11 | n.a. | n.a. |
| parental education | | | | |
| High school or less | n.a. | n.a. | 270 | 29.54 |
| Associate/Some College | n.a. | n.a. | 273 | 29.87 |
| Bachelor | n.a. | n.a. | 54 | 5.91 |
| Masters/PhD | n.a. | n.a. | 317 | 34.68 |

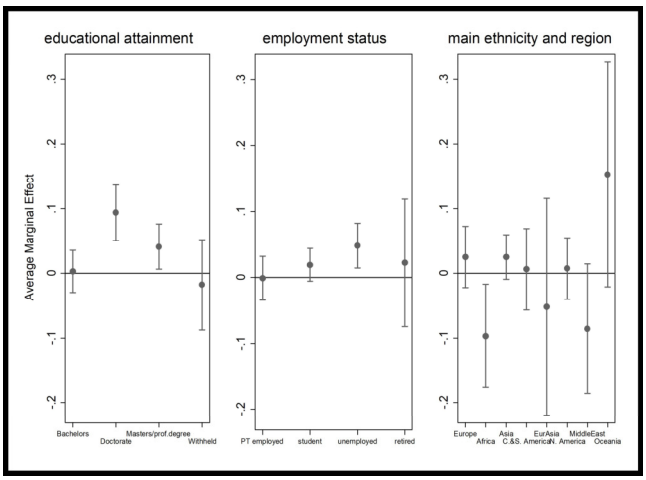


Fig. 2 Statistical Learning
Average Marginal Effects for the full model (95% confidence interval)

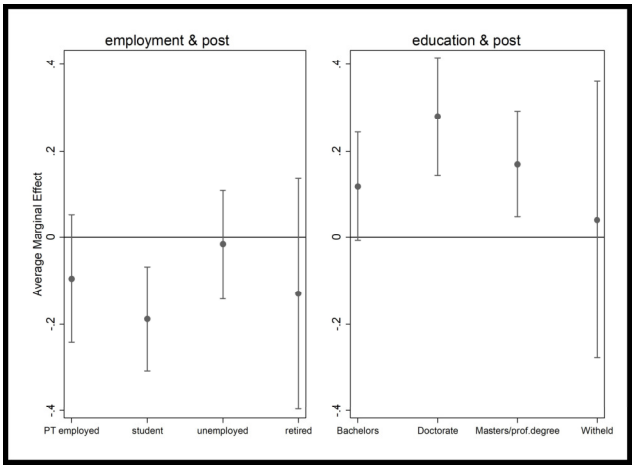


Fig. 3 Statistical Learning
Average Marginal Effects for interactions (95% confidence interval)
Note: interaction between main ethnicity & region & post omitted due to small N and collinearity

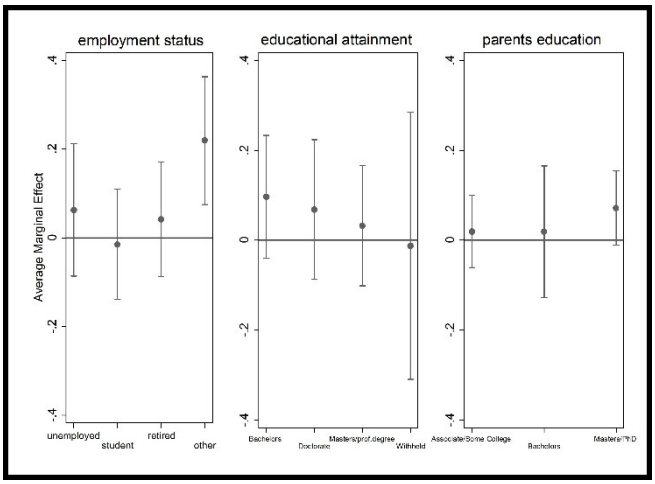


Fig. 4 Poverty and Inequality
Average Marginal Effects for the full model (95% confidence interval)

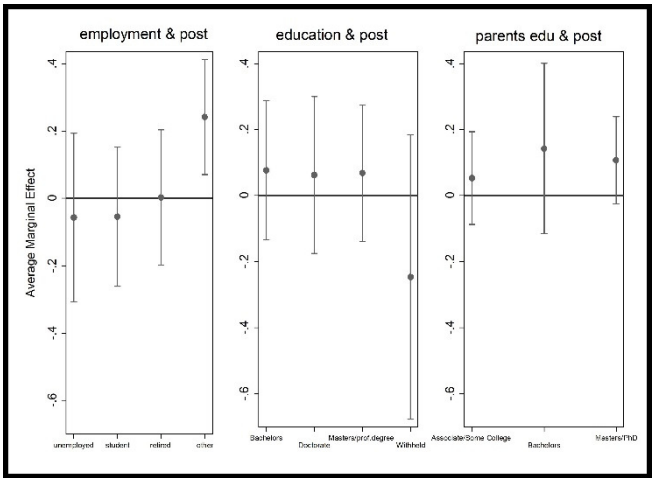


Fig. 5 Statistical Learning
Average Marginal Effects for interactions (95% confidence interval)

## V. Conclusions

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). does not. 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.

Finally, an interesting 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.

Finally, these analyses provide a first interesting exploration of the social dimension of MOOCs, which is often overlooked in empirical research. However, it also presents some limitations that can 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 collaboration and valuable comments.

## REFERENCES

[1] ClassCentral, "A Product at Every Price: A Review of MOOC Stats and Trends in 2017," 2018. [Online]. Available: https://www.class-central.com/report/moocs-stats-and-trends-2017/. [Accessed: 07-May-2018].

[2] I. Literat, "Implications of massive open online courses for higher education: mitigating or reifying educational inequities?," Higher Education Research & Development, vol. 34, no. 6, pp. 1164–1177, Nov. 2015.

[3] V. Rolfe, "A Systematic Review of the Socio-Ethical Aspects of Massive Online Open Courses," European Journal of Open, Distance and E-Learning, vol. 18, no. 1, pp. 53–72, 2015.

[4] Kalz, M., Kreijns, K., Walhout, J., Castaño-Munoz, J., Espasa, A., and Tovar, E., "Establishing a European cross-provider data collection about open online courses.," The International Review of Research in Open and Distributed Learning (IRRODL), vol. 16, no. 6, pp. 62–77, 2015.

[5] A. Margaryan, M. Bianco, and A. Littlejohn, "Instructional quality of Massive Open Online Courses (MOOCs)," Computers & Education, vol. 80, pp. 77–83, Jan. 2015.

[6] G. Christensen, A. Steinmetz, B. Alcorn, A. Bennett, D. Woods, and E. Emanuel, "The MOOC Phenomenon: Who Takes Massive Open Online Courses and Why?," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2350964, Nov. 2013.

[7] A. Ho et al., "HarvardX and MITx: Two Years of Open Online Courses Fall 2012-Summer 2014," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2586847, Mar. 2015.

[8] G. Sharrock, "Making sense of the MOOCs debate," Journal of Higher Education Policy and Management, vol. 37, no. 5, pp. 597–609, Sep. 2015.

[9] Y. Shavit and H.-P. Blossfeld, Persistent inequality: changing educational attainment in thirteen countries. Boulder, Colo.: Westview Press, 1993.

[10] Y. Shavit, R. Arum, A. Gamoran, and G. Menachem, Stratification in Higher Education: A Comparative Study. Stanford University Press, 2007.

[11] M. Triventi, "Stratification in Higher Education and Its Relationship with Social Inequality: A Comparative Study of 11 European Countries," European Sociological Review, vol. 29, no. 3, pp. 489–502, Jun. 2013.

[12] D. J. Zhang, G. Allon, and J. A. Van Mieghem, "Does Social Interaction Improve Learning Outcomes? Evidence from Field Experiments on Massive Open Online Courses," Manufacturing & Service Operations Management, vol. 19, no. 3, pp. 347–367, Jul. 2017.

[13] B. J. Evans and R. B. Baker, "MOOCs and Persistence: Definitions and Predictors," in MOOCs and Higher Education: Implications for Institutional Research, S. J. Blackmon and C. H. Major, Eds. San Francisco: Jossey-Bass, 2016, pp. 69–86.

[14] B. J. Evans, R. B. Baker, and T. S. Dee, "Persistence Patterns in Massive Open Online Courses (MOOCs)," The Journal of Higher Education, vol. 87, no. 2, pp. 206–242, Mar. 2016.

[15] J. Huang, A. Dasgupta, A. Ghosh, J. Manning, and M. Sanders, "Superposter behavior in MOOC forums," in Proceedings of the First ACM Conference on Learning at Scale, 2014, pp. 117–126.

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