Peer assessment in the first French MOOC : Analyzing assessors' behavior

Matthieu Cisel
ENS Cachan
61 av. du Pdt Wilson
94230 Cachan
+33 1 43 37 08 52
mcisel@ens-cachan.fr

Rémi Bachelet
Ecole Centrale de Lille
Cité Scientifique,
59651 Villeneuve d'Ascq
+33 3 20 33 53 53
remi.bachelet@ec-lille.fr

Eric Bruillard
ENS Cachan
61 av. du Pdt Wilson
94230 Cachan
+33 1 47 40 24 57
bruillard@ens-cachan.fr

ABSTRACT
Massive Open Online Courses (MOOCs) have spread incredibly fast over the past two years, triggering a worldwide debate over the impact of online learning on the democratization of higher education. Given the increasing number of students registering to MOOCs for free, course instructors who want to go beyond automated evaluation have no choice but to use peer assessment. In response to recurrent criticism over the unreliability of peer evaluation, techniques such as calibrated peer review have been developed, but still very little is known about the factors that influence assessors' behavior. Based on two editions of *Introduction to Project Management*, the first French xMOOC, we explored the impact of learners' background on their engagement in peer assessment and on the grades they gave. We observed that registrants that took part in peer evaluation differed significantly from other participants in regards to time constraints and demographic variables such as geographical origin. As far as grades were concerned, students tended to give higher grades than other registrants, especially when they could get credentials by completing the course.

Keywords
Peer assessment, MOOC, engagement, bias

1. INTRODUCTION
The impact of Massive Open Online Courses (MOOCs) has considerably deepened since the foundation of edX and Coursera in 2012 [3], fostering hope over the possibility to open up high quality education. Nevertheless, initial enthusiasm has been tempered by recurrent criticism over different aspects of MOOCs such as their low completion rates [1] or the unreliability of the grading process. Many courses rely on peer assessment [9] to evaluate at no cost large amounts of assignments. This grading process is easily scalable, but has faced high level of skepticism given the fact that MOOC participants are not trained examiners.

In response to those critics, calibrated peer review and algorithms were developed in order to either decrease graders' bias *a priori* or correct for it *a posteriori* [8]. Nevertheless, a deeper understanding of the factors influencing the peer grading process is needed in order to increase the efficiency of those mitigation strategies. Indeed, assessing peers is not a common practice in educational systems; some participants may refuse to take part in the process because they do not feel legitimate enough to grade assignments, or merely because they do not want to spend extra time on the course. Even among those who take part in peer assessment, grades assessors give may be influenced by their situation. For instance, one may think that people in management positions are more severe than students because they fell less uncomfortable with giving bad grades. *Introduction to Project Management* is the first French xMOOC; it relies extensively on peer assessment and therefore represents an interesting case study in regards to those issues. How does participants' background influence their engagement in the peer assessment on one hand and the grades they give on the other hand?

2. MATERIAL AND METHODS

2.1 Course description
ABC de la Gestion de Projet (*Introduction to Project Management*) is a MOOC organized by Centrale Lille, it was run twice in 2013, on Canvas.net during spring and on a Canvas LMS-based website in the fall. The first edition took place from March 18th to April 21st; it was four weeks long, 3495 registered and 1332 completed the course. The second one took place from September 16th to October 20th. It was 5 weeks long, 10847 registered and 3301 completed the course. In the second edition of the course exclusively, 579 students from Centrale Lille and several other French institutions of higher education registered. Those students could get credentials by completing the MOOC. Thereafter, we will make the distinction between students who could get credentials by completing the course and other students who registered independently.

Two certificates corresponding to two different workloads were offered - a basic one and an advanced one. According to the professor in charge of the MOOC, completion of the basic certificate and the advanced certificate required respectively around two hours and eight hours per week. In the first edition and the second edition of the course, respectively 894 and 2492 obtained the basic certificate and 438 and 809 obtained the advanced certificate.

The course provided videos, quizzes, weekly assignments and a final examination. To obtain the basic certificate, it was required to complete the quizzes and to pass the final exam. In order to obtain the advanced certificate, participants were required to pass the basic certificate and submit weekly assignments that were peer evaluated based on a precise marking scheme. Peer assessment was not anonymous. Only one assignment could be skipped for those who aimed at getting the certificate. In the second edition of the MOOC, course instructors threatened to lower the grades of the participants who had taken part in the peer assessment process; this threat was never applied though. Assignments were evaluated four times each in the first edition, and five times each in the second. Consequently, over the duration of the MOOC registrants could assess up to 16 and 25 assignments in the first and the second edition, respectively.

2.2 Available data and methods

Participant activity reports, gradebooks and surveys used in this study were downloaded from Canvas Learning Management System. In both editions, participants were asked to fill in a survey at the beginning of the course. It was responded to by 69% and 54% of the registrants, and by 99% and 93% of the 1332 and 3301 of those who completed the MOOC, in the first edition and the second one, respectively.

The following demographic comparisons are only valid under the assumption that responding to the survey is independent from learner's background. IP addresses were not collected; therefore all available data on geographical origin was extracted from surveys. We focused on some demographic variables of interest, and reported the results for the participants who filled in the survey only (Table 1). Some variables like "highest diploma" had significant effects but were partly redundant with variables like *employment status*. They were therefore removed from the models. Among variables without any statistically significant impact, we only reported gender, which stands as a control.

Countries were classified into three categories based on their human development index (HDI), this data was provided by the UN [10]. Countries with a low HDI and those with a very high HDI will thereafter be referred to as Least Developed Countries and More Developed Countries, respectively. For statistical reasons, medium and high HDI were grouped and will be referred to as Intermediate HDI. In addition to common demographic variables, participants were asked how many hours they intended to spend on the course.

For each grade G_{ij} received by the artifact j and given by the assessor i, a grade score GS_{ij} was computed. It is the difference between the grade G_{ij} and the mean grade of the artifact, Mj, as follows $GSij = Gij - Mj$. This score is positive if the grade lies above the mean, and negative otherwise. Each assessor i received a grader score GrS_i, which is the mean of his grade scores. GrS_i= mean (GS_{ij}). This score is positive if the assessor gives in average higher grades than other graders, and negative otherwise. The maximum grade for an assignment was 100. A grader score of 10 meant that the assessor, in average, would give grades 10 points higher than other assessor. We could not get the timestamps associated with each assessment. Consequently, no time series analysis was carried out. Anonymised data was analyzed with the open source statistical software R 2.12.

In order to obtain odd-ratios, we computed logistic regressions (glm procedure, family="binomial"). In the case of Table 1, The higher is the Odd-ratio, the higher is the proportion of registrants who submitted at least an assignment for the given category, compared with the reference (Ref). For instance, participants from More Developed Countries were 1.98 more likely to submit an assignment than participants from Least Developed Countries, which represent the reference. Regarding Table 2, the lower is the odd-ratio, the lower is the proportion of participants who skipped peer assessment. For grader score, we used a linear model; a higher estimate means a higher grader score compared to the reference (Ref).

Table 1. Composition of the MOOC, based on the initial survey, for the spring edition and the fall edition. Reg: number of registrants. Sub: Number of registrants who submitted at least one assignment. O.R.: Odd-ratios. Ref: reference. Higher OR means more participants who submitted at least an assignment, compared to the reference

Variable	Spring edition			Fall edition		
Gender	Reg	Sub	O.R.	Reg	Sub	O.R.
Male	1564	408	*Ref*	3603	738	*Ref*
Female	744	182	0.97	1877	328	0.88
Employment status						
Higher management positions	1156	314	1.27	2322	429	**1.46** *
Lower management positions	296	61	0.79	837	117	0.99
Unemployed	303	98	1.23	585	179	1.06
Students	316	52	0.73	716	101	1.01
Students (in curriculum)	0	0		579	232	**6.04** ***
Others	238	64	*Ref*	442	65	*Ref*
HDI						
Low	551	122	*Ref*	1090	174	*Ref*
Intermediate	208	54	**1.43** *	546	98	1.29
Very High	1521	410	**1.98** ***	3807	786	**1.50** ***
Hours per week						
below 2 h	579	42	**0.31** ***	1783	161	**0.43** ***
between 2 to 4 h	908	182	*Ref*	1794	280	Ref
between 4 to 6 h	529	225	**3.22** ***	982	301	**2.77** ***
above 6 h	304	142	***4.39*** ***	895	315	**3.86** ***
*p-value <0.05, ** p-value <0.01, ***p-value <0.001						

3. RESULTS

3.1 A selective process

Only a fraction of registrants submitted assignments and were therefore allowed to take part in the peer assessment process (Table 1). Dropout rate was relatively low within the advanced

certificate. Among participants who had submitted at least an assignment, 71% and 68% obtained it, in the first and the second edition of the course, respectively.

Given the potential influence of assessors' background on their behavior, we first aimed at determining whether participants who submitted assignments differed significantly from other registrants' as far as demographic variables and time constraints were concerned. Through a logistic regression, we compared these two types of participants (Table 1). Interaction terms were not significant, and were therefore removed from the analysis.

Table 2. Understanding engagement in the peer assessment process and factors affecting graders' bias**.** Sk. OR : Odd-ratios for participants who skipped peer assessment at least once. E. Gr S: Estimates of the linear model for grader score. Ref : Reference

Variable	**Spring**		**Fall**	
Gender	Sk. O.R.	E. GrS	Sk. O.R.	E. GrS
Male	*Ref*	*Ref*	*Ref*	*Ref*
Female	0.74	0.59	0.90	0.37
Employment status				
Higher management positions	0.81	0.88	1.11	0.31
Lower management positions	0.96	0.86	1.95	-1.1
Unemployed	1.30	3.77	1.26	-1.04
Students	1.36	-0.90	1.15	4.62*
Students (in curriculum)			**0.51****	**5.35****
Others	*Ref*	*Ref*	*Ref*	*Ref*
HDI				
Low	*Ref*	*Ref*	*Ref*	*Ref*
Intermediate	1.01	3.12	1.31	0.12
Very High	**0.30****	2.81	**0.48****	0.86
Hours per week				
below 2 h	1.44	-3.83	1.62	0.06
between 2 to 4 h	Ref	*Ref*	*Ref*	*Ref*
between 4 to 6 h	1.03	-0.49	0.91	0.70
above 6 h	1.17	-1.49	1.21	0.44
*p-value <0.05, ** p-value <0.01, ***p-value <0.001				

Geographical origin and time constraints were the main drivers of selection. As shown in Table 1, registrants from More Developed Countries were more likely to submit assignments than those from Least Developed Countries. Time constraints were also a very important driver of selection. Participants who were not able to spend more than two hours per week on the MOOC were unlikely to submit an assignment, and consequently to take part in the peer assessment process.

3.2 Skipping peer assessment

Given that taking part in peer assessment was encouraged but not compulsory to get the certificate, a significant proportion of the participants skipped it for some of the assignments they had submitted. This proportion was higher for the spring edition (32.7%), than for the fall edition (8.3%). Engagement in the peer assessment was related to the participation in the course. In the case of the first edition of the course, the more assignments participants submitted, the less likely they were to skip peer assessment (Figure 1). It ranged from 57% for those who ha submitted only one assignment to only 25% for those who had submitted the all four assignments. This trend was less obvious for the second edition of the MOOC; even among participants who had submitted only one assignment, a very small proportion actually skipped peer assessment.

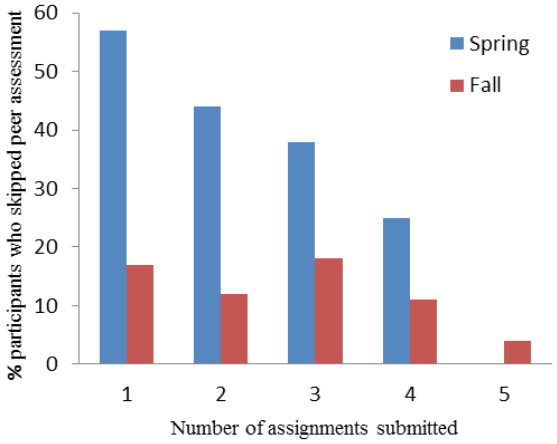


Figure 1. Number of assignements submitted by participants versus engagement in peer assessment for the spring and the fall edition of the MOOC

Based on logistic regression, we aimed at identifying the background of participants who skipped peer assessment, registrants who had not submitted at least an assignment were excluded from the model (Table 2, Sk. OR). Time constraints had no longer any statistically significant impact. Only geographical origin had an impact in the spring edition. Participants from More Developed Countries were 70% less likely to skip peer assessment than participants from Least Developed Countries. This trend was also observed in the fall edition, during which we noticed that students who would get credentials for the course were less likely than others to skip peer assessment.

3.3 Students tend to give higher grades

We have focused so far on the identification of the factors that influenced engagement in the peer assessment process. The first round of selection is the submission of the assignment, since it was required to be allowed to evaluate peers. Afterwards, a fraction of the participants that had passed this first round of selection skipped some peer assessment; it therefore represents a second round of selection. The last step is the grading process itself. We classified participants according to the grades they gave. Based on Student t tests with a 0.05 confidence, we found that for 17% of the 615 assessors of the first edition and for 35% of the 1170 assessors of the second one, distribution of grade scores was statistically different from 0, meaning that in average

they were giving either higher or lower grades than other assessors.

Through a linear regression applied to the grader score metric, we aimed at identifying factors associated with graders' bias. It appeared that employment status was the most influencing factor, but only in the fall edition (Table 2, E. GrS); we could not detect any other factor with a statistically significant impact. More specifically, we noticed that students tended to give higher marks than others in average, especially when they could get credentials by completing the course. For this category of students, the estimate of the grader score was 5.35. It means that in average, compared to other assessors, students gave grades more than 5 points higher.

CONCLUSION

The increasing use of peer assessment in MOOCs is probably one of the main paradigm shifts of the current evolutions of online learning. Not only does it allow course instructors to scale up the evaluation process, but it also enhances the course's pedagogical value. Improving its relevance requires identifying factors that influence engagement in the assessment process and those that impact graders' bias [8]. We observed differences in engagement patterns within the peer assessment process; some participants did all the required peer evaluations whereas others disengaged. Similar patterns were observed at a larger scale for various MOOCs regarding completion rates [6]. Given the amount of time required by the assignments, it is not surprising that participants who started submitting assignments displayed a higher level of engagement than average registrants. Therefore, selection occurred mostly before peer assessment itself.

Employment status and geographical origin were the factors that influenced engagement the most, a trend that had been detected in previous studies [2]. Further investigations are needed to understand why participants from Least Developed Countries show lower levels of engagement than those from More Developed Countries, regarding both submission of assignments and participation in peer assessment. Technical issues have been pointed as a possible explanation, but other reasons may be at stake. Time constraints were also one of the main drivers of selection, which is not surprising given that most registrants follow MOOCs during their free time. Identification of the factors associated with engagement in online programs is not a new issue [5], but the open nature of MOOCs has made this question even more critical.

Regardless of participants' background, traditional incentives such as bonuses and maluses seem to be a simple method to drive participants to take part in the peer evaluation process. Indeed, the threat to get a lower grade for not taking part in the peer evaluation process is the most likely explanation for the decrease of participants who skipped peer assessment in the second edition of the MOOC. Finally, we noticed that grading behavior could depend upon assessors' background; students tended to give higher grades than other assessors, and might therefore be less reliable as far as peer assessment is concerned. This might be due to the fact that peer assessment was not anonymous.

In order to enhance MOOCs' completion rates, algorithms are being developed in order to flag participants about to drop out [4], and course instructors are testing ways to influence learners' behavior [7]. Similar approaches could be followed in the near future to detect disengaging or unreliable assessors in order to influence their behavior. Further investigations on graders' behavior could be helpful for course instructors who aim at designing efficient strategies to increase the accuracy and the relevance of peer assessment.

4. ACKNOWLEDGMENTS

The authors wish to address their grateful thoughts to Rémi Bachelet, the professor of the MOOC, to Unow and Instructure, who hosted the MOOC, and to the team of students and volunteers who took part in the design of the course. I wish to thank Muriel Epstein and Mehdi Khaneboubi, assistant professors in educational studies, for their helpful remarks and feedbacks. This research was reported to the *Commission Nationale Informatique et Liberté* (CNIL), the French institution in charge of online privacy (ticket 1691291).

5. REFERENCES

[1] 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 and Practice in Assessment*, *8*, 13–25.

[2] Cisel, M. (2014). Analyzing completion rates in the First French xMOOC. *Proceedings of the European MOOC Stakeholder Summit 2014*

[3] Daniel, J. (2012). Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility. *Journal of Interactive Media in Education*, *3*(0). Retrieved from http://www-jime.open.ac.uk/jime/article/view/2012-18 (Accessed 18-Sep-2013).

[4] Halawa, S., Greene, D. & Mitchell, J. (2014) Dropout Prediction in MOOCs using Learner Activity Features. *Proceedings of the European MOOC Stakeholder Summit 2014*

[5] Hart, C. (2012). Factors associated with student persistence in an online program of study: A review of the literature. *Journal of Interactive Online Learning*, *11*, 19-42.

[6] Kizilcec, R. F., Piech C., Schneider E. (2013) Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses, LAK'13 *Proceedings of the Third International Conference on Learning Analytics and Knowledge*.

[7] Kizilcec, R., Schneider, E. ,Cohen, G.L & MacFarland, D. A. (2014) Encouraging Forum Participation in Online Courses with Collectivist, Individualist and Neutral Motivational Framings. *Proceedings of the European MOOC Stakeholder Summit 2014*

[8] Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuned Models of Peer Assessment in MOOCs. Retrieved from http://www.stanford.edu/~cpiech/bio/papers/tuningPeerGrading.pdf (Accessed 18-Sep-2013).

[9] Sadler, P. Good, E. (2006). The Impact of Self-and Peer-Grading on Student Learning. *Educational Assessment*, *11*, 1-31.

[10] United Nations Development Programme. (2011). Human development report. Retrieved from

http://hdr.undp.org/en/media/HDR_2011
Statistical_Tables.xls (Accessed 18-Sep-2013)