What affects permanence in a MOOC about Chemistry?

Gabriela Trindade Perry (Corresponding author)

Graduate Program of Informatics in Education, Universidade Federal do Rio Grande do Sul (UFRGS),

Porto Alegre, Brazil.

Henrique Padovani Queiroz Machado

Department of Physics Engineering, Universidade Federal do Rio Grande do Sul (UFRGS),

Porto Alegre, Brazil

Napoliana Souza

Graduate Program of Informatics in Education, Universidade Federal do Rio Grande do Sul (UFRGS),

Porto Alegre, Brazil.

Paola Rosatto

Graduate Program of Applied Mathmatics, Universidade Federal do Rio Grande do Sul (UFRGS),

Porto Alegre, Brazil.

Abstract

In this paper we analyze what influences the permanence of 1606 students in a MOOC on General Chemistry, using navigation records (log files). Permanence – quantified as the amount of course items viewed and tests completed – was compared regarding the following parameters: (1) showing the correct answers after completing evaluative questionnaires; (2) offering certificates of completion; (3) allowing non-linear navigation (free browsing). Results from the Mann Whitney tests revealed that offering a certificate and showing the correct answer to test questions influence the permanence. However, when considering a smaller cut of students - those who completed at least 30% of the activities - none of the parameters influenced permanence. From these results, it can be argued that these two configuration parameters are relevant in relation to permanence, since they are an incentive for students who perform fewer activities, and therefore are those who are at greater risk of evading.

Keywords: MOOC; permanence; dropout

1. Introduction

The interest and dissemination of MOOCs (Massive Open Online Courses), has grown in recent years, whether due to the innovation aspect reinforced by the media (Pappano, 2012) or by the emergence of platforms developed with famous universities worldwide. In 2008, Siemens and Downes made the first

edition of the Connectivism course, which became known as the first MOOC, and was attended by approximately 2,300 students (Downes, 2012). This was the "first generation of MOOCs", also known as "Connectivist" (cMOOC). For Downes (2014), traditional courses are designed as books; you have to follow the chapters by the end of the book, while the "connectivist" courses look more like magazines or newspapers. Siemens (2013) states that connectivist MOOCs are non-linear, and emphasize the importance of learner autonomy.

In contrast to cMOOCs are xMOOCs (the "x" suggests the "extension" of a university course), a model that has grown and attracted crowds to online course platforms. Rhoads et al. (2015), for example, points out that xMOOCs offered potential economic benefits in the form of a more automated and digitized version of university subjects. According to them, xMOOCs tend to be attractive to entrepreneurs, investors, policy makers and administrators seeking solutions to problems of access to higher education. In a matter of a few years the xMOOCs came to dominate the global scenario of this modality.

Anyway, MOOCs are articulated within this contemporary moment of enormous social, demographic, economic and technological changes in which universities reform and move. The next landmark in the recent history of MOOCs was in 2011, when three engineering professors from Stanford University offered computer science courses in the form of MOOCs. The Introduction to Artificial Intelligence course, offered for free to students from around the world, attracted 160,000 students from more than 190 different countries (Stacey, 2013).

From this moment on we see a turnaround in the development of online courses, with the entrance of new actors that form consortiums in robust platforms, offering a new model of MOOC which is pedagogically and philosophically distant from the connectivist ideals of the pioneers of MOOCs. The first letter O of the acronym MOOCs that carried the OPEN principle has undergone significant changes, keeping only the sense of being open to any participant, but no longer sustains itself while free access to the content..

Hollands and Tirthalli (2014) state there is no doubt that the entry of MOOCs into the educational landscape has led many higher education institutions to revisit their distance education strategies. Initially, the overwhelming majority of the MOOC courses came from American universities and was delivered through platforms like Coursera, EdX, Udacity, Udemy, but this picture has changed. Currently, there are platforms from many other countries such as FutureLearn (UK), MiríadaX (Spain), XuetangX (China), MexicoX (Mexico), EduOpen (Italy), ThaiMooc (Thailand). It is possible to see a strong tendency of universities around the world to offer MOOCs and attract the glances for themselves. Boal and Stallivieri (2015) argue that it is possible to say that MOOCs are influencing the process of internationalization of higher education institutions.

Brazil, despite its continental dimension and a network of higher education institutions, has a small production of MOOCs. Brazil does not have any high-impact MOOC platform, despite the potential of a country with continental dimensions and challenging educational needs. Brazilian platforms are Veduca, Unesp Aberta, Open Courses UEMA, and more recently, Lúmina.

On the other hand, the country has amount of students enrolled in courses in international platforms. Allione and Stein (2016) point out that the participants of the course researched by the authors represented 147 countries with the five largest being the United States, India, China, Brazil and Spain. In the University of

London report (2014), Brazil appears as the 5th country in number of participants in the Coursera platform, with 6% of this total.

2. Dropout and Permanence

At the end of any course, whether in person or at a distance, participants expect to receive a diploma, certificate or attestation of completion. Therefore, completing a course is traditionally a relevant moment in the study experience, and this is part of the expectations of all those involved in the process: institutions, teachers and students. Therefore, a relevant data in the evaluation of courses in general is the ration between how many students have entered and how many have concluded, because if the students begin the course and do not conclude, it is a sign that something is not going well. The Brazilian Minstry of Education points out, for example, that in 2014 there was a 49% droput from graduation courses. Regarding MOOCs, dropout and non-completeness of is one of its main characteristics and, consequently, one of its greatest criticisms. It is well known that MOOCs attract thousands of students, but only between 5 and 12% complete the course and earn a certificate of completion (Koller et al 2013, Clow 2013, Perna et al., 2014). Jordan (2015), in an extensive analysis of 221 MOOCs, presents completion rates between 0.7% and 52.1%, with the average being 12.6%, that is, a dropout   higher than 80%. Comparatively, according to the 2016 ABED (Brazilian Association of Distance Education) census, dropout in formal ED courses offered by higher education institutions in Brazil ranged from 11% to 25%.

2.1 What influences dropout?

Onah et al. (2014) present some reasons for high dropout rates in MOOCs: no intention to complete the course, lack of time, difficulty level of the course, lack of support, lack of familiarity with technology, lack of study habits, negative experiences, wrong expectations about the course, late start and peer evaluation. Gütl et al. (2014) categorize dropouts in two groups: "healthy", which concerns a group of participants who are selecting course content and "unhealthy", with students who initially wanted to complete the course but fail for a number of reasons. Gomez-Zermeno and Aleman (2016) identified that the main characteristics of the participants who completed the course are: to have postgraduate degrees, previous online educational experience, commitment to the course, economic stability, proficiency in the use of information technology, advanced proficiency in digital resource creation, intermediate English skills, advanced proficiency in using techniques and methods to organize knowledge and active participation in research networks. On the other hand, those who decided to leave the course indicated problems with the structure and orientation in the course, limitations in the use of information technology or in English, besides the limited availability of time for family or work reasons. Eriksson et al. (2016) say time is the major limiting factor of course completion.

One of the conclusions of Ho's (2015) research is that courses whose participants paid for a certificate had a completion percentage of 59%, while those that did not require payment had completion rates of 5% - it suggests financial investment may be a factor influencing evasion in MOOCs. Ebben and Murphy (2014) argue that the fact that MOOCs are free and do not impose access difficulties often leads participants to start a course without having incentive to continue.

Kizilcec and Halawa (2015) conducted a systematic review about dropout, based on self-assessments and behavioral data from more than 100,000 participants in 21 courses. The research tested hypotheses on the role of psychological factors in self-attributed success by participants. The findings were: the existence of persistence and performance gaps among students of different genres and locations; the relationship between students who identify themselves as successful in online learning and higher levels of focus on goals, social belonging, and personal growth goals; the challenges faced by online learners, the most important of which was the lack of sufficient time, which often appeared related to low levels of willpower. Adamapoulos (2013) says the teacher has an important positive effect on the likelihood of a student successfully completing a course, as well as students' feelings about course tasks and materials. Forums have a partial positive effect on the probability of completion of a course. The difficulty level, the workload and the duration of a course have a negative effect. On the other hand, for the most difficult courses, having an open schedule and greater workload have a positive effect on the likelihood of completion. The authors also concluded that if a certificate is awarded upon successful completion of a course, this factor also positively affects retention of participants. The university's reputation also positively affects the likelihood of completion. The content is also pointed out as important when evaluating the possible reasons for the avoidance of courses. Hohe and El Said (2016) focused their research predominantly on participants' perceptions of MOOC resources rather than individual student characteristics. The findings of this investigation identified that the content of the MOOC has a significant effect on the retention of participants, based on the utility perceived by the participants. In addition, interaction with the teacher / tutor also has a significant effect on the participants' permanence.

3. Methodology

This research is a case study of a MOOC on General Chemistry, available on the Lúmina platform. The MOOCs available in Lúmina maintain log records (or log files) - according to Azevedo et al. (2013) this is a usual practice. In its user agreement contract , Lúmina warns about navigation data collection.

We compared three parameters focusing on permanence, which is conceptualized as the number of activities carried out in the course. It was decided to focus on permanence instead of dropout, because at first all students who do not complete 100% of the activities could be considered as evaded, while permanence refers to the degree of interaction with the course - and can be measured, for example, as time logged in and/or quantity of activities (as in this research).

  The parameters in analysis were:
1. Blocking versus allowing non-linear navigation (free browsing). In the case of linear (restricted) navigation, the student should complete activity A1 before having access to the immediately following activity A2.
2. Showing versus not showing correct answers after submitting answers to evaluation questionnaires.
3. Offering versus not offering completion certificates.

3.1 The MOOC on General Chemistry
The object of study was a MOOC called "General Chemistry" (which had two editions). Both editions had

the same content - they only changed in relation to the configuration parameters analyzed in this research. Both editions were composed of 21 videos (one presentation video and twenty with video lessons), 17 pages with link lists and 17 evaluation questionnaires (between 2 and 7 questions). They also had a video and an introductory text, a student profile survey, a forum for questions and suggestions and a course evaluation survey. Table 1 shows the summary of these courses.

Table 1. Summary of General Chemistry courses

	First edition	*Second edition*
Release date	August 2016	June 2018
Closing date	June 2018	-
Enrollment	490	1116
Percentage of students with 80% to 100% of completed activities	1,24%	31%
Average time to complete between 80% and 100% of activities	78 days	24,5 days
Average scores of completed questionnaires	8,7	8,7

The course versions do not define the grouping of variables, but the dates on which they occurred.

3.2 Students' profile
It was not possible to obtain the profile of the 1606 students, because, due to a configuration error of profile survey, the questions were not marked as mandatory. Therefore, the quantities refer only to the individuals who answered the questions (summarized in Table 2).

Table 2. Profile of the respondent sample

	First edition	*Second edition*
Gender	37 (M) 19 (F)	406 (M) 674 (F)
Schooling	13 (UND) 35 (COL) 9 (POS)	275 (UND) 630 (COL) 140 (POS)
Do you plan to complete the course?	5 (N) 52 (Y)	61 (N) 991 (Y)

M= Male | F = Female
UND = Undergraduate school | COL = College | POS = Post-graduation
N = No | Y = Yes

3.3 Data wrangling routine
Lúmina is based on Moodle, which offers the following reports: activity completion, scores, navigation log and survey data. To obtain the data used in this research, we used only the activity completion report, which has 1 student per line. This log contained the following columns:

- Name and email.
- Date of completion for each activity.

The first step of the data wrangling routine was excluding students who had not performed activities (those

who only registered or just answered the profile search). Users with admin or teacher role were also removed. Finally, duplicate students (for example, students who unsubscribed and then re-subscribed) were also excluded. After this filtering, there were 1606 unique students. Next, the following columns were created:

- Number of completed activities (integer).
- Percentage of activities completed (float).
- Show response (boolean).
- Offer certificate (boolean).
- Free browsing (boolean).

The values of the last three columns – which represent the parameters under analysis in this research – were filled according to the date each parameter was set up.

The data wrangling and statistical tests (Mann-Whitney) were done using the R language. The Mann-Whitney test does not compare the mean, median, or any other distribution parameter, using the mean-rank concept: values are ordered incrementally, and the highest values occupy the highest ranks (when there are equal values, an average value is assigned for the ranks). The mean rank represents the sum of the ranks of each group - in the case of this survey there were three variables, with two groups each. According to Daniel (1990), the Mann-Whitney test has the following assumptions: (1) the data come from random observations; (2) the two samples are independent; (3) the observed variable is continuous (4) the distribution of populations differs only in relation to position - if they differ.

It should be noted that the effects of the configurations are not discriminated by the Mann-Whitney test. This is due to the date each parameter was set up in Lúmina. So, for example, every student who (could) earn a certification also could browse freely.

3.4 Research hypothesis

Before stating the hypothesis, it should be noted that the parameters under analysis were changed only once, so their effects were not separated. The dates of modification were:

- Free browsing was allowed after May 2017.
- Correct answers as feedback in evaluative activities were shown until June 2018 (after this date, no feedback was given).
- Certificates of completion begun in July 2017.

The hypotheses of this research refer to the difference between each of these three parameters, regarding permanence.

Hypothesis 1 - Navigation mode: There were no statistically significant (95% level) differences among the medians of the number of activities carried out by the students enrolled in the course version that did not allow free navigation versus the enrolled ones in the course that allowed free navigation.

Hypothesis 2 – Completion certificate: There were no statistically significant differences (at a 95% level) among the medians of the number of activities carried out by students enrolled in the edition of the

course that did not offer a certificate (and allowed free browsing) versus those enrolled in the one offered (and allowed free browsing).

Hypothesis 3 – Correct answers: There were no statistically significant differences (95% level) among the medians of the number of activities carried out by the students enrolled in the course edition that allowed the visualization of the correct answers after sending the evaluation questionnaires (and allowed free browsing and offered a certificate) versus the enrolled questionnaires in the one that did not allow (and allowed free browsing and offered a certificate).

4. Results

4.1 Free navigation (free browsing)

The difference between the amount of activities completed was compared. May 2017 defines the dates of enrollment for each group – restricted navigation (n = 336), and free navigation (n = 1270). Figure 1 shows the density plot of these two groups.
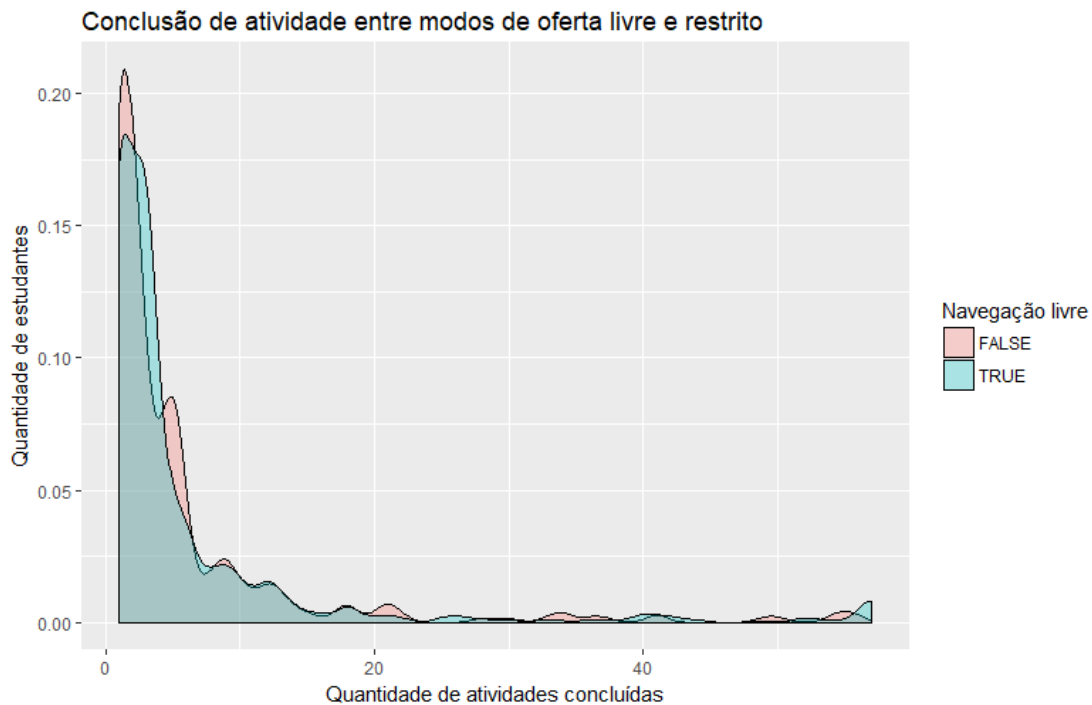

Figure 1 Density plot for the free and restricted browsing groups.

The plot in Figure 1 does not show difference in permanence - measured as the amount of completed activities. The Mann-Whitney (two-tailed) test confirms this observation (p = 0.405), so that the null hypothesis cannot be rejected (there is no difference between the samples). The average rank of the student sample that used linear navigation was 785, while the ranks of non-linear navigation students were 808.

4.2 Offering certificates

The difference between the amount of activities carried out by the students who registered before and after July 2017 (this date defines the change in the configuration of the certification offering), was compared.

Before this date, no certificate was offered (n = 437), and after it was offered (n = 1169). Figure 2 shows the density plot of these two groups.
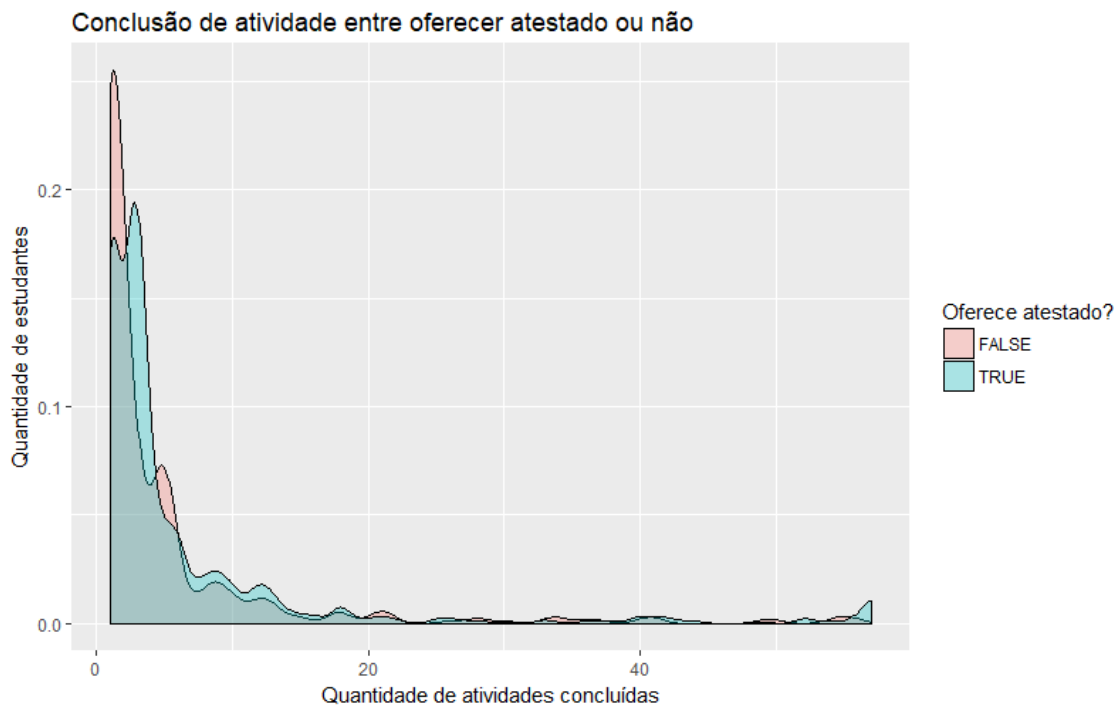

Figure 2 Density plot between providing and not providing certificate.

The plot shown in Figure 2 shows some difference in permanence - measured as the amount of completed activities – especially in the beginning of the course. The Mann-Whitney (two-tailed) test confirms this observation (p = 0.00), so that the null hypothesis is rejected (there is a difference between the samples). The average range of the sample of students who did not receive a certificate was 670, while the rank of the students that received was 843.

The same test was performed, however, considering a smaller cut, containing only the students who performed between 30% and 100% of the activities, that is, between 18 and 56 activities. This cut was arbitrary, and was not based in the literature (no similar study was found). The density plot was used as the criterion for defining these values. The new sample contained 28 students in the group that did not receive certificates and 92 in the group that received a certificate. The average range of the sample of students who did not receive a certificate was 51, while the rank of the students they received was 63. These values are considerably smaller than in the previous test because of the number of observations (sample size). As p = 0,113, the null hypothesis cannot be rejected for this group, which means: to offer or not a certificate does not influence permanence of students who completed at least 30% of the course.

4.3 Showing the correct answer as feedback in evaluative activities

The difference between the number of activities performed by the students who enrolled before and after June 2018 was compared (the date that defines the change in the configuration of how to show the correct answers after submitting evaluative activities). Before this date, the correct answers were shown as

feedback (n = 103), and after they were not shown as feedback (n = 1503). Figure 3 shows the density plot of these two groups.
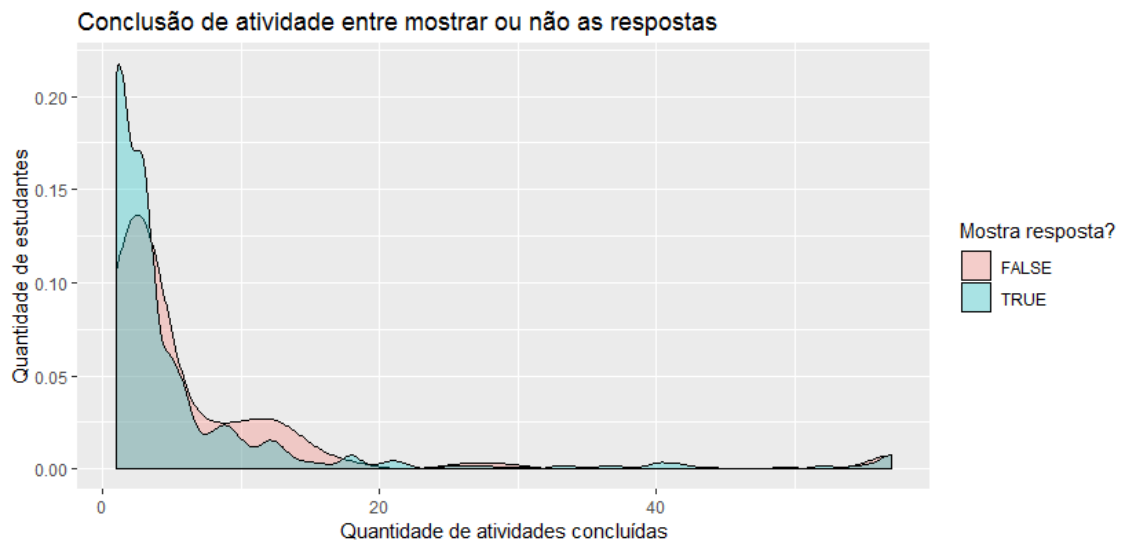

Figure 3 Density plot for to show or not correct answers as feedback

The plot in Figure 3 shows some difference in amount of completed activities in the beginning of the course. The Mann-Whitney (two-tailed) test confirms this observation (p = 0.035), so that the null hypothesis is rejected (there is a difference between the samples). The average rank of the student sample that viewed the answers was 895, while the rank of the students who did not visualize the answers was 797.

The same test was performed, however, considering a smaller cut, containing only the students who performed between 30% and 100% of the activities, that is, between 18 and 56 activities. This cut was arbitrary, and was not based on figures found in the literature (no similar study was found). The density plot was used as the criterion for defining these values. The new sample contained 6 students in the group that did not received feedback with the correct answers and 114 in the group that did. The average rank of the former group was 60, while the rank of the latter group was 72. These values are considerably smaller than in the previous test because they are a function of the number of observations (sample size). As p = 0.415, the null hypothesis cannot be rejected for this group, which means: to show or not the answer does not influence permanence in students who completed at least 30% of the course.

5. Conclusion

The results about the certificate and correct answers parameters were consistent with what was expected - after all, it is reasonable that whenever access is facilitated, permanence will increase. In the case of the navigation style, it was a surprise it did not make a difference.

Regarding showing the correct answers as feedback, it has some pedagogical implications, since it can be argued that when the correct answer is not shown, the student might not feel encouraged to continue. In this scenario, one can imagine that the feeling of failure would lead the student to give up - which seems

to have been the case when considering the total sample (i.e., who performed at least 1 activity). However, as the platform's administrators consider it is necessary to pay attention in class to be approved in the course – Lúmina is not showing this feedback anymore.

In relation to the certificates, it was expected that it would increase permanence. However, when considering the students who performed at least 30% of the activities, the results change, and none of the parameters had any effect. From these results, it can be argued that these configuration parameters are relevant factors in relation to permanence, since they are an incentive for students who perform less than 30% of the course activities, usually those that are at greater risk of dropping out.

6. Acknowledgement

7. References

[1] ADAMOPOULOS, P. What Makes a Great MOOC? *An Interdisciplinary Analysis of Student Retention in Online Courses*. Thirty Fourth International Conference on Information Systems, 2013, 1–21.

[2] ALLIONE, Gloria; STEIN, Rebecca M. Mass attrition: An analysis of drop out from principles of microeconomics MOOC. *The Journal of Economic Education*, Philadelphia, 2017, v. 47, n. 2, p.174-186.

[3] AZEVEDO, R.; HARLEY, J.; TREVORS, G.; FEYZI-BEHNAGH, R.; DUFFY, M.; BOUCHET, F.; LANDIS, R.S. Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In: R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies*, vol. 26, p. 427-449, 2013.

[4] BOAL, H., M., C.; STALLIVIERI, L. 2015. The influence of Massive Online Open Courses (MOOCs) on the internationalization process of higher education. *Digital Universities: Best Practices*, Roma, v. 2-3, n. 2015, p.21-36.

[5] CLOW, D. *MOOCs and the funnel of participation*. In: Third Conference on Learning Analytics and Knowledge (LAK 2013), 8-12 April 2013, Leuven, Belgium, pp. 185–189.

[6] DANIEL, W., W. 1990. *Applied nonparametric statistics*. Boston: PWS-Kent. 2ª edição.

[7] DOWNES, S. Connectivism and connective knowledge: essays on meaning and learning networks. National Research Council Canada, 2012.

[8] DOWNES, S. Like reading a newspaper. 2014. Half an Hour.

[9] EBBEN, M.; MURPHY, J. 2014. Unpacking MOOC scholarly discourse: A review of nascent MOOC scholarship. *Learning Media and Technology* 39(3).

[10] ERIKSSON, T; ADAWI, T; STÖHR, C. 2016. Time is the bottleneck: a qualitative study exploring why learners drop out of MOOCs. *Journal Of Computing In Higher Education*, [s.l.], v. 29, n. 1,

p.133-146,.

[11] GOMEZ-ZERMENO, M., G.; LAGARZA, L., A.. 2016. Research analysis on MOOC course dropout and retention rates. *Turkish Online Journal of Distance Education*, [s.l.], p.03-14, 1 abr. 2016.

[12] GÜTL, C.; CHANG, V.; MORALES, M. 2014. *Attrition in MOOC: lessons learned from drop-out students*. In: Learning technology for education in cloud. MOOC and Big data, 2014, pp. 37e48. Springer International Publishing.

[13] HO, A. D. et al. 2015. HarvardX and MITx : Two Years of Open Online Courses Fall 2012-Summer 2014. *SSRN Electronic Journal*, 2015. n. 10, p. 1–37

[14] HOLLANDS, F., M.; TIRTHALI, D. 2014. *Resource requirements and costs of developing and delivering MOOCs*. The International Review of Research in open and Distributed Learning, v. 15, n. 5.

[15] HONE, K. S.; SAID, G. R. El. Exploring the factors affecting MOOC retention: A survey study. *Computers & Education*, 2016. v. 98, p. 157–168.

[16] KIZILCEC, R; HALAWA, S. 2015. *Attrition and Achievement Gaps in Online Learning. Proceedings Of The Second*. ACM Conference On Learning @ Scale - L@s '15, [s.l.], p.57-66, 2015. ACM Press.

[17] KOLLER, D. et al. 2013. Retention and Intention in Massive Open Online Courses (New Horizons). *Educause Review*. p. 62–63.

[18] ONAH, D. F. O., SINCLAIR, J.; BOYATT, R. 2014.*Dropout Rates Of Massive Open Online Courses: Behavioural Patterns*. EDULEARN14 Proceedings, 5825-5834.

[19] PAPPANO, L. The Year of the MOOC. 2012. *The New York Times*, 1–7.

[20] PERNA, L. W. et al. 2014. Moving Through MOOCs: Understanding the Progression of Users in Massive Open Online Courses. *Educational Researcher*, [s.l.], v. 43, n. 9, p.421-432,

[21] RHOADS, R A. et al. 2015. The Massive Open Online Course Movement, xMOOCs, and Faculty Labor. *The Review of Higher Education*, [s.l.], v. 38, n. 3, p.397-424.

[22] STACEY, P. The Pedagogy of MOOCs.