Attrition patterns amongst participant groups in Massive Open Online Courses

David G. Glance, P. Hugh R. Barrett
The University of Western Australia

In their current instantiation, Massive Open Online Courses (MOOCs) represent a form of educational delivery with little barrier to entry. In part due to the ease of enrolling, low completion rates have been a notable characteristic of MOOCs with as few as 5 - 10% of enrolled students gaining a certificate of completion. Beyond these observations, little is known about the patterns of attrition, nor how they may vary amongst different groups of MOOC participants.

Activity data was analysed from log records collected from 42 MOOCs run by Stanford University between 2011 and 2013 on Coursera. A novel analysis was applied to the data which showed that attrition varies little between those that audited the course and those that actively participated in assessments.

The results indicate that attrition is not influenced by levels of participation in a MOOC and attrition is further not greatly influenced by attributes of the course.

Keywords: MOOCs, student attrition

Introduction

In their current instantiation, Massive Open Online Courses (MOOCs) represent a form of educational delivery with a very low barrier to entry. Students with access to a computer and the Internet can register for a class with the click of a button. Students can decide at what level to engage with the video content, readings, interactive exercises, and assessments and ultimately determine what they get out of the entire experience.

In part due to the ease of enrolling, low completion rates have been a notable characteristic of MOOCs with as few as 5 - 10% of students who enrolled gaining a certificate of completion (Gallagher & Garret, 2013). In examining the issue of disengagement from MOOCs, it has been pointed out that the simple binary categorisation of students taking MOOCs into completers and non-completers may be misleading (Kizilcec et al., 2013). They analyzed 3 MOOCs run by Stanford University on the MOOC website Coursera (Coursera, 2014) using a cluster analysis approach and this revealed several distinct student groups based on their level of engagement with the course. They found that they could split completers and non-completers into four distinct groups. Completers were divided into “Completing” and “Auditing”, with the Auditing group watching videos but not attempting the quizzes or assignments. Non-completers could be divided into “Disengaging” and “Sampling”. Disengagers started the course watching videos and attempting exercises but dropped out, typically in the first third of the course. Samplers typically watched one or two videos before dropping out. In addition to these groups are the “Non-starters”, those students who enrolled but never interact with the course in any way. On Coursera, approximately 55% of students who enrolled, were classified as non-starters.

In another method of grouping MOOC participants, active students were categorised on a spectrum (Seaton et al., 2010); from browsers (auditing) to certificate earners (completing). The basic characteristic that distinguished MOOC participants was whether they only watched videos, or they attempted at least one or more of the activities set as homework over the 14 week period of the course. The researchers found that participants who attempted over 5% of the homework on the course represented only 25% of the overall participants, but that they accounted for 92% of the total amount of time everyone spent on the course. Another interesting aspect of the completers who earned certificates from the course was that 25% of them watched less than 20% of the video content. The overall median time spent by all participants on the MOOC was one hour.

The studies quoted so far refer to MOOCs that are taught in a conventional way with lectures, quizzes and assignments. These MOOCs are often referred to as xMOOCs (Siemens, 2012). The original MOOC term was used to refer to a different format of MOOC called the cMOOCs that emphasised connectivist learning (Siemens, 2005). In an analysis of different formats of engagement of one cMOOC entitled Change11, Milligan et al (2013) identified the same split of completers into what they termed ‘Lurkers’ and ‘Active participants’. The Lurkers equated to the Auditors discussed earlier with Active participants paralleling the Completer group. Because of the nature of the study, they did not quantify disengagers in any way but did discuss a third group called ‘Passive participants’ who persisted with the course passively but were doing so because they were unhappy about the connectivist approach.

In examining further the characteristics of participants and MOOC attrition, Kizilcec et al (2013) found evidence for differences in both the demographics of course participants and student disengagement according to the level of the course. They employed a course difficulty classification that rated the three courses at high school, undergraduate and graduate levels, equating to introductory, intermediate and advanced. The introductory course attracted more female students than the others with larger numbers of students completing.

In a discussion on student engagement in higher education Carini et al (2006) point out the possibly self-evident premise that the “more students study or practice a subject, the more they tend to learn about it”. Although possibly a proxy for learning outcomes, the examination of engagement on its own may not tell us particularly how well students learned as a result or whether their learning was superficial or deep. For this, we would need to address this question specifically within the context of a theoretical model such as Bigg’s 3P model of learning (Biggs, 1989). The issue with engagement and MOOCs however is possibly more relevant than in a face-to-face higher education setting because of the truism that unless the students complete a significant portion of the course, the issue of how much they learned may well be moot. There is a secondary question that should also be tackled which is to examine the difference in learning outcomes between students who audit a course and view only the videos and those that engage fully, completing quizzes, assignments and exams. The question is not really about whether there is a difference in outcomes but rather what learning, if any, occurs for people simply watching videos? This is an important consideration for educational designers as the question of how much effort should be expended to accommodate auditors will depend on whether they are really learning anything or simply being “entertained”. This may be considered a worthwhile goal of MOOCs, especially if they original motivation was to enhance engagement of the public at large with the institution in a general way.

The final question to be tackled when looking at the subject of engagement in MOOCs is what activities are used to measure it. MOOC platforms can yield a wealth of data, allowing for quite sophisticated measures of time spent viewing videos, completing quizzes and assessments and contributing to forums. Of these, the most basic is watching videos and is the activity that all students groups have in common. Thus the simple measure of the number of unique views of videos each week is an effective proxy measure of the number of students participating in the course, either as auditors or as fully-engaged completers (Belanger & Thornton, 2013, Breslow et al, 2013).

Finally, in the discussion of attrition in MOOCs, it is worth considering it in the context of the substantial body of research on attrition of learners in more established online and distance learning environments. In a review of research of early attrition among first time eLearners, Tyler-Smith (2006) highlighted that attrition rates have been quoted as being as high as 70 – 80%. Tyler-Smith has proposed that early attrition from these courses, especially for first-time eLearners may be due to a cognitive overload as students struggle to cope with technology and isolation, on top of new content.

Strategies for engaging online students and reducing attrition rates in traditional online courses have included increasing the level of engagement between academic staff and students (Angelino, et al, 2007). This is not possible to scale to the numbers involved in MOOCs which is why Georgia Tech’s online Master of Science programme, which utilizes MOOCs as the basis of its curriculum, limited enrolled student numbers to just 375 (Georgia Tech, 2014).

The focus of this study was to examine the patterns of attrition in a series of MOOCs run by Stanford University on the Coursera website. This examination proceeded from the initial observation that the decay of activity in MOOCs appeared to follow a classic two-phase decay (Motulsky, 2004) that is seen in a wide range of social, physical and biological systems. The first question was whether there were any differences between courses in these attrition curves or their constants.

In the remainder of this paper, we describe the methods of normalization for the MOOC activity data and their subsequent analysis as two-phase exponential decay curves. Finally, we discuss the significance of the findings of this analysis in the context of the behavior of students taking these courses and what consequences this might have for interpreting the levels of enagement in MOOCs and what MOOC developers might be able to do to engage their students more comprehensively.

Methods and results

Activity data was analysed from log records collected from 42 MOOCs run by Stanford University between 2011 and 2013 on Coursera. There were 17 unique courses, some of which were repeated. The majority of these courses were in Computer Science (32) with the remaining courses being in Mathematics (5), Science (3), Politics (1) and Health (1).

In surveys of the students taking 36 of these courses, of the 120,861 respondents, approximately 77% were male and 23% female. 34% of the survey respondents reported having a college degree (Bachelors, Masters or PhD).

In analysing the activity for each course, we normalised all activity by the number of students who watched the first video. This was 53.97% (+/- 14.92%) of the total number of people who had "enrolled" in the course. Activity was then calculated by counting the number of individuals that had watched a particular video or attempted a quiz. We also normalised the sequence of videos and quizzes in the course. This was necessary because each course had a different number of videos and quizzes. The normalisation process involved making the total number of videos and quizzes 100 and taking the sequence number of the video or quiz as a percentage of that total.

When video and quiz activity is plotted against the normalised sequence of the video and quiz, a characteristic activity profile results that shows a two-phase decay (Motulsky, 2004). Curves of this sort illustrate that there are two different mechanisms at work in the attrition of students from MOOCs. The curve shows an initial fast decay, followed by a slower decay with an overlap between the two. Each phase of the decay has a half-life value that represents the point at which 50% of the initial population has ceased activity. This is illustrated in Figure 1 for the Automata MOOC.

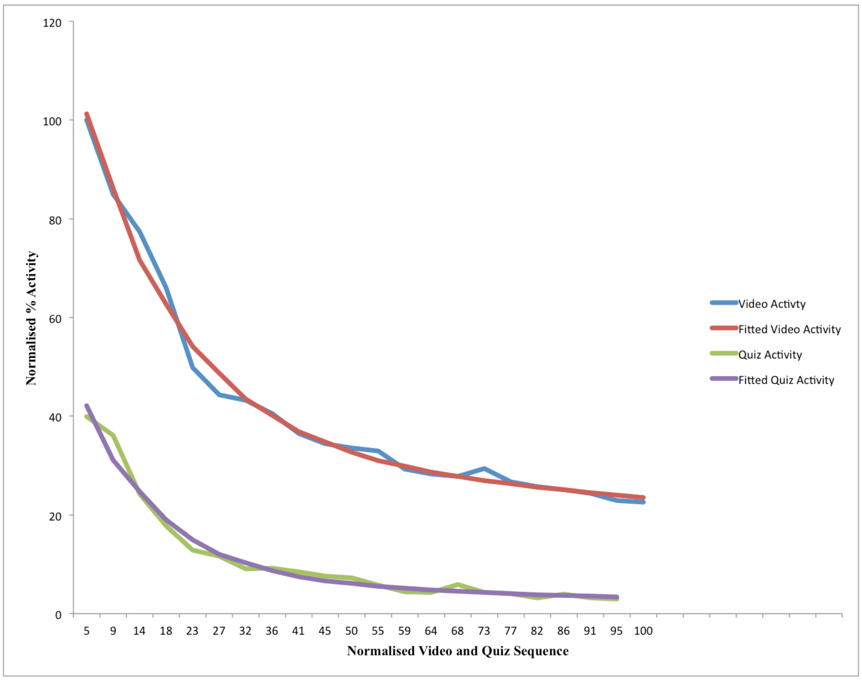

Figure 1: Video and Quiz activity decay curves for the Automata MOOC

Exponential curves were fitted using the Solver function in Microsoft Excel (Laverman, 2013) and the half-life of each phase calculated. A mean half-life from the two component half-lives was also calculated.

Table 1 presents a summary of the courses and the number of people who have subscribed to each course, watched the first and last videos and taken the first and last quizzes. On average, 46% of students enrolled in these MOOCs were non-starters, not engaging in any activity on the MOOC after they had enrolled. A smaller number of students attempted the first quiz and this difference between students who just watched videos, the auditors, and those who were actively taking some or all of the quizzes, the engagers, persisted until the end of the course. On average, 8% of the students who enrolled were engagers.

Table 1: Summary of subscribers and video and quiz activity

Subscribers	% 1st Video	% 1st Quiz	% Last Video	% Last Quiz
51,285 ± 31,197	54 ± 15	44 ± 13	17 ± 6	8 ± 7

The first thing to note in Table 2 is that there is normally a dual phase exponential decay in activity for both video watching and quiz taking. The normal process of decay for video activity was to have a fast component with an average half-life of 7.68 (± 5.6) and a slow component with a half-life of 93.71 (± 31.30). The half-life of the quiz activity was 8.07 (± 7.78) for the fast component and 106.47 (± 161.66) for the slow component.

Using a Pearson's product-moment correlation analysis in SPSS, the data showed that there was no correlation between the half-life of the fast and slow components of either the videos or the quizzes. There was a significant correlation between the mean half-life of video activity and the mean half-life of quiz activity ($r = 0.383$, $n = 42$, $p = 0.14$).

Table 2: Summary of exponential analysis of MOOC video and quiz activity

Video t1/2 Fast	Video t1/2 Slow	Video t1/2 Mean	Quiz t1/2 Fast	Quiz t1/2 Slow	Quiz t1/2 Mean
7.68 ± 5.6	93.71 ± 31.30	48.94 ± 20.86	8.07 ± 7.78	106.47 ± 161.66	34.86 ± 37. 98

Figure 2 shows a summary of all of the normalised fitted activity curves for students watching videos in the MOOCs. The graph illustrates the fact that all courses showed a very similar pattern of attrition with the least variability being displayed in the initial rapid drop-out from MOOCs. This phenomenon was despite the differences in the courses; the individuals giving the courses, their subject matter, their duration in weeks, the number of activities and the level of difficulty of the courses. This similarity of the attrition curve is therefore indicative of common factors being the root cause of students dropping out from the MOOCs.

Figure 3 shows a summary of all of the normalised fitted activity curves for students taking quizzes in the MOOCs. As for the video activity, the attrition curve for quiz activity shows a very common pattern for all MOOCs. The other thing to note is that there were fewer students taking quizzes than watching videos and that the shapes of the curves for both video and quiz attrition are similar.

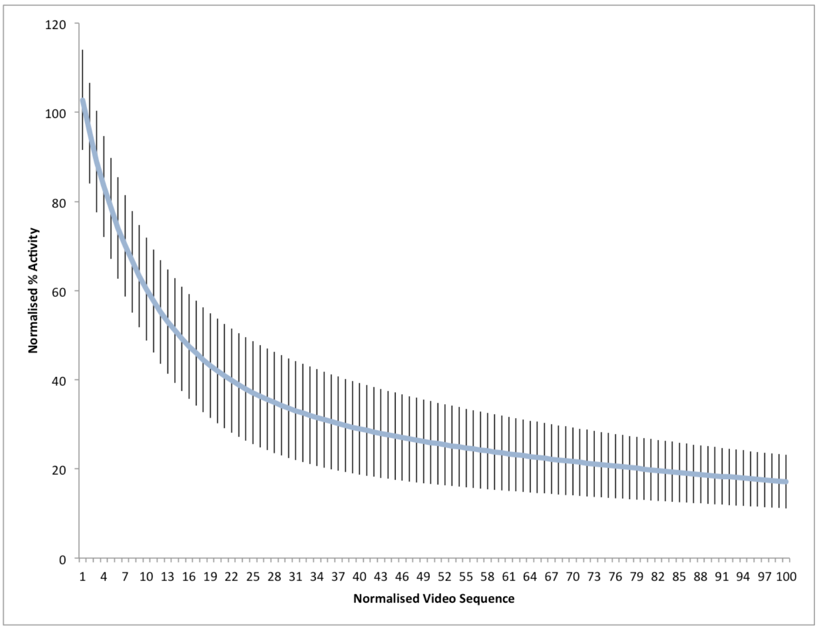

Figure 2: Mean normalised fitted video decay curve for all of the study MOOCs. Bars show ± standard deviation.

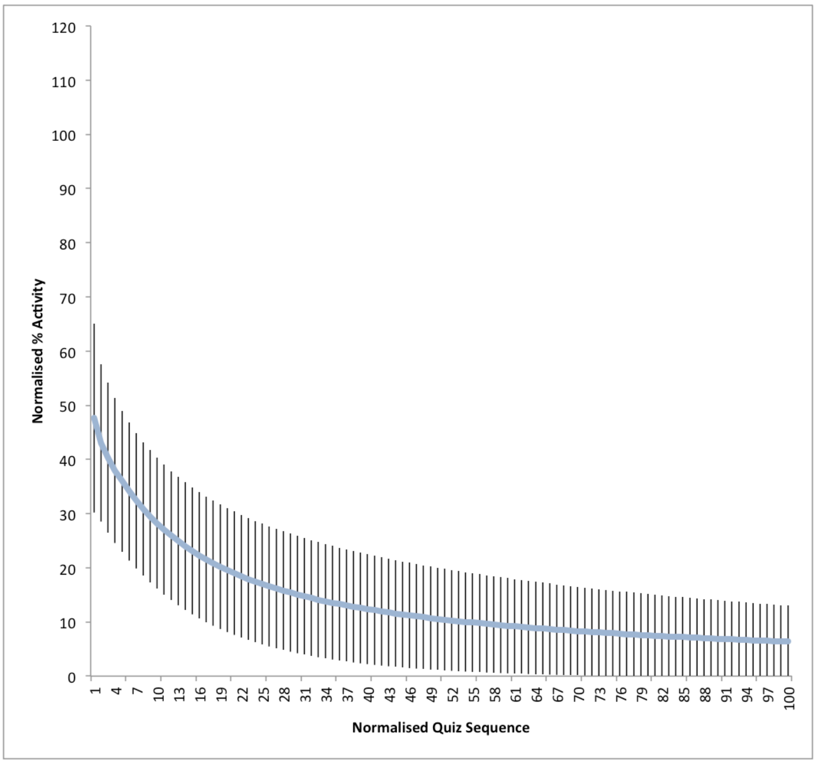

Figure 3: Mean normalised fitted quiz decay curve for all of the study MOOCs. Bars show ± standard deviation.

As a highlight of the deterministic characteristic of MOOC attrition, and illustrated in Figure 4, the number of students who watched the first video in a MOOC was highly correlated to the initial number of enrolled students ($r = 0.809$, $n = 42$, $p < 0.01$). The number of students who attempted the first quiz was also highly correlated to the initial number of enrolled students ($r = 0.92$, $n = 42$, $p < 0.01$).

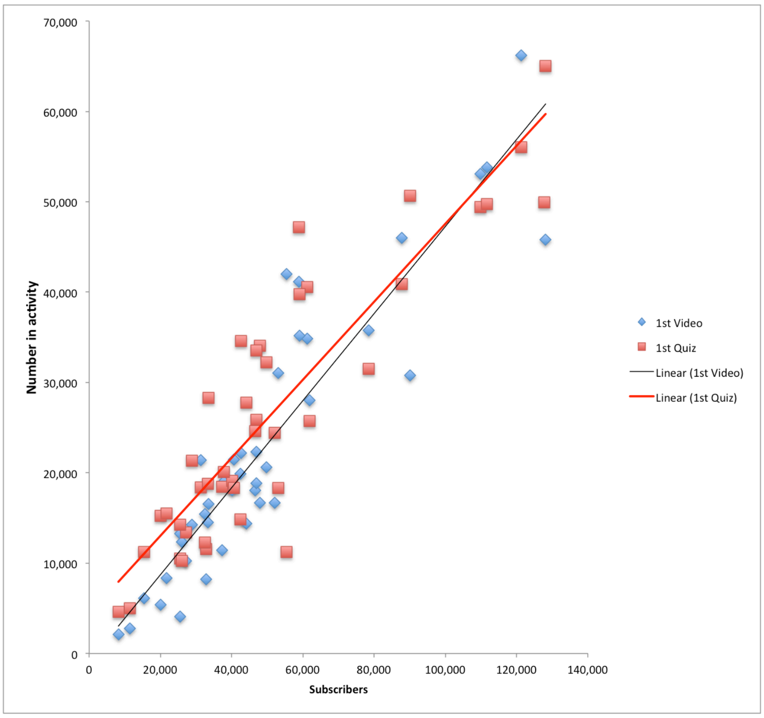

Figure 4: Relationship between number of subscribers and participants who watched the 1st video and attempted the 1st quiz

Figure 5 shows that the relationship between initial numbers of enrolled students and auditors and engagers persisted to the end of the course in the MOOCs in this study. There was significant correlation between the initial number of students enrolled in a MOOC and the number of students who watched the final video ($r = 0.858$, $n = 42$, $p < 0.01$). The correlation between the initial number of students enrolled in a MOOC and the number of students who took the final quiz was ($r = 0.605$, $n = 42$, $p < 0.01$).

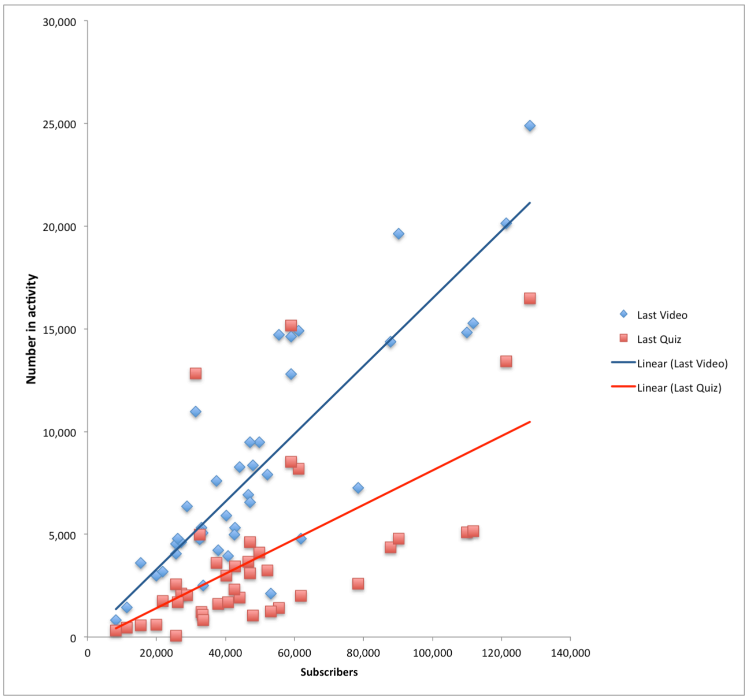

Figure 5: Relationship between number of subscribers and participants who watched the last video and attempted the last quiz

Discussion

The analysis of student activity in terms of the main two components, watching videos and taking quizzes, shows that there are at least two groups of people taking MOOCs. Auditors are a group of students who only watch videos and engagers are those who take quizzes and also watch videos. This categorisation of MOOC students is simpler than that outlined by Kizilcec et al. (2013). Unlike their analysis, this study does not look at the question of whether there is movement between these groups or whether there are people who only take quizzes. However, the overall similarity between the decay curves of the auditors and engagers and especially the measure of mean attrition half-life of these two groups, demonstrates that attrition is not related to level of engagement and is a deterministic process.

The second somewhat surprising finding is that the attrition of students from these courses was largely predictable despite the variation in subject, course duration, course difficulty, number of activities and any other difference between these courses. We have observed that of the students that initially subscribe to a course, approximately 55% will watch the first video and 45% will take the first quiz. Activity will decay rapidly initially and then more slowly to result in a final audience of 17% watching the last video and 8% taking the final quiz. This is keeping with data reported elsewhere (See Gallagher & Garret, 2013).

The normalisation process used to compare courses illustrates the fact that attrition is dependent on the number of activities in a course, represented in this study as watching videos or taking quizzes. This was independent of the duration of time because these activities varied in number and also by the time period in which they were offered. This is likely to be because students principally drop out because they report not having enough time to continue with a MOOC (Khalil & Ebner, 2014). This reason for attrition would certainly fit with the observation of attrition after each activity offered in the course. MOOCs are normally structured to take place over a given time frame and so it is easy to fall behind and give up when it proves difficult or impossible to catch up. It is worth noting however that at least two of the courses examined in this study were self-paced and so the pressure of deadlines would not have been an issue for those courses.

Other factors may additionally be at play, including the fact that students are not at an educational level capable of comprehending or engaging with the MOOCs. However, approximately 34% of the students taking the MOOCs had a prior university qualification including a large number of postgraduates and so this may not have been a factor in their case.

The fact that attrition in MOOCs follows a consistent path highlights that the shorter the duration of the course, and the fewer activities that students have to take, the greater the number of students who will complete the overall course. Although it would be possible to implement strategies to increase engagement and try and slow down the drop-out rate, this study shows that the factors for students not continuing are more likely to be because of external factors such as lack of time rather than individual levels of interest or difficulty with the material. As the courses cost the students nothing to enroll in, there is little incentive to continue with a course once any external pressures become a factor and so it would appear that little can be done to prevent students dropping out.

MOOC attrition curves appear to be similar in their characteristics to those found in other contexts that require effort and time on the part of the consumer. In particular, gym membership retention rate shows an exponential curve of a similar sort, albeit over a longer time period than for a MOOC course (The Retention People, 2014).

The difficulty with the focus on the massive aspect of MOOCs is that it has drawn attention to their inherent and dominant property of high attrition. In fact, as this study has shown, for many of the MOOCs today, the numbers of students actually participating in any way in the MOOC at the midpoint of the course is approximately 25% of the enrolled students. Of course, this can still represent a cohort of several thousand, but in total, the numbers of students taking MOOCs is a tiny fraction of the 200 million people enrolled in tertiary education worldwide (UNESCO, 2012).

Any strategies that are implemented to keep MOOC student attrition down, are likely to come at a cost of a diminution of the actual educational experience and outcomes of the course since the only effective strategy would be to simply have less content. An alternative strategy would be to simply accept that large numbers of students will inevitably drop out and to concentrate on the 17% who will get to the end, either auditing the course or completing it through engagement at all levels.

The methodology for analysis of MOOC attrition presented in this paper presents a standardized means of comparison of attrition across MOOCs, normalizing for the number of activities and the duration of the course. Although this study has highlighted the behavior of MOOC students as a whole, it has not detailed the subtleties of MOOC engagement at the level of the individual. We do not know for example if a student who simply audited a MOOC actually was able to retain any knowledge or gain any understanding of the content.

Although the courses on this MOOC were of mixed type, they were predominantly concerned with computer science. It is still possible that non-computer science MOOCs will show different patterns of attrition, although our work would suggest that if time is the main determinant of MOOC attrition, then this would not be the case.

Future work will help answer these questions and also focus on quantifying the factors that cause students to drop out of a MOOC, confirming perhaps that lack of time is indeed the main determinant of MOOC attrition.

Acknowledgements

The authors would like to acknowledge support of the The Office of the Vice Provost for Online Learning at Stanford University for their assistance in this study and for providing access to the data of their MOOCs.

References

Angelino, L. M., Williams, F. K., & Natvig, D. (2007). Strategies to Engage Online Students and Reduce Attrition Rates. *Journal of Educators Online*, *4*(2), n2.

Belanger, Y., & Thornton, J. (2013). *Bioelectricity: A Quantitative Approach Duke University's First MOOC*. Retrieved from http://hdl.handle.net/10161/6216

Breslow, L. B., 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 & Practice in Assessment* 8. 13-25.

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education* 47, no. 1, 1-32.

Coursera, (2014). Coursera website. Retrieved from http://www.coursera.org

Gallagher, G. (2013). *Disruptive Education: Technology-Enabled Universities*
Retrieved from
http://ussc.edu.au/ussc/assets/media/docs/publications/130801_DisruptiveEducation_GallagherGarrett.pdf

Georgia Tech Online Masters in Computer Science, (2014). Retrieved from http://www.omscs.gatech.edu/uncategorized/oms-cs-update-january-15-2014/

Hanan, K. & Ebner, M. (2014). MOOCs Completion Rates and Possible Methods to Improve Retention-A Literature Review. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, vol. 2014, no. 1, 1305-1313.

Kizilcec, R. F., Piech, C. & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, (pp. 170-179). ACM.

Laverman, L. E. (2013). *Using Solver in Excel for non-linear curve fitting.* Retrieved from http://web.chem.ucsb.edu/~laverman/Chem116/PDF116CL/Solver.pdf

Milligan, C., Littlejohn, A. & Margaryan, A. (2013). Patterns of engagement in connectivist MOOCs. *MERLOT Journal of Online Learning and Teaching*, *9*(2).

Motulsky, H., Christopolous, A. (2004). *Fitting Models to Biological Data Using Linear and Nonlinear Regression. A practical guide to curve fitting.* Oxford University Press.

Seaton, D. T., Bergner, T., Chuang, I., Mitros, P. & Pritchard, D. E. (2010). Who Does What in a Massive Open Online Course? *International. J. Hum.-Comput. Stud* 68, 223-241.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2(1), 3-10. Retrieved from http://www.itdl.org/Journal/Jan_05/article01.htm

Siemens, G. (2012). *MOOCs are really a platform* [Web log post]. Retrieved from http://www.elearnspace.org/blog/2012/07/25/moocs-are-really-a-platform

The Retention People. (2014). *Measuring Retention*. Retrieved from http://www.theretentionpeople.com/insight/research/measuring-retention/

Tyler-Smit, K. (2006). Early Attrition among First Time eLearners: A Review of Factors that Contribute to Drop-out, Withdrawal and Non-completion Rates of Adult Learners undertaking eLearning Programmes. *Journal of Online Learning and Teaching*. *2*(2), 73-85.

UNESCO, (2012). Enrolment statistics. *UNESCO Institute for Statistics*. Retrieved from http://data.uis.unesco.org/Index.aspx?queryid=128

Contact author: David Glance, david.glance@uwa.edu.au

Please cite as: Glance, D. G., & Barrett, P. H. R. (2014). Attrition patterns amongst participant groups in Massive Open Online Courses. In B. Hegarty, J. McDonald, & S.-K. Loke (Eds.), *Rhetoric and Reality: Critical perspectives on educational technology. Proceedings ascilite Dunedin 2014* (pp. 12-20).

Note: All published papers are refereed, having undergone a double-blind peer-review process.