

Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
<http://economics.sas.upenn.edu/pier>

PIER Working Paper 14-031

“Mass attrition: An analysis of drop out from a Principles of Microeconomics MOOC”

by

Rebecca M Stein and Gloria Allione

<http://ssrn.com/abstract=2505028>

Mass attrition: An analysis of drop out from a Principles of Microeconomics¹ MOOC.

**Rebecca M Stein and Gloria Allione
University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.**

Abstract

Though Mass Open Online Courses are very different from each other in their structure, content or audience, they are all characterized by low completion rates. In this paper, we use the Cox proportional hazard model to analyze student retention in the MOOC Principle of Microeconomics. Using two different measures of retention, video watching and quiz submission, we show that students' commitment to the course can be strongly predicted by their participation in the first week's activities. Data collected through a voluntary opt-in survey allow us to study retention in relation to demographics, for a subset of enrollees. We find a higher dropout rate for college students or younger participants. Female attrition is larger when measured by video watching but not when measured by quiz completion. Self-ascribed motivation for taking the course is not a predictor of completion. We conclude that raw completion rates cannot be the criterion to judge the success of MOOCs, as it is in case of traditional courses. The results are consistent with the existing literature which separates MOOCs students into two different groups: Committed Learners and Browsers.

Introduction

Mass open online courses (MOOCs) have been around for only a short time but have already caused a stir in higher education. A Massive Open Online Course (or MOOC) is a course given over the web for free, to anyone who is interested in signing up and has the basic technical capabilities to do so. The movement started in 2008 and took off in 2012, dubbed by the New York Times as 'The Year of the MOOC'. Some large platforms include Coursera, Udacity and edX. The potential of reaching a wide audience, both geographically and in terms of interests and demographics, is exciting. The chance of offering a lower-cost method of delivering education is transformative. MOOCs have already reached millions of students, introducing them to topics as diverse as modern poetry and introduction to artificial intelligence.

It has been clear from the start that though MOOCs enroll tens of thousands of students each, only a very small fraction of them, between 5 and 12 percent, finishes the course and earns a certificate of completion (Koller et al 2013). Low retention rates have been a

¹ We would like to thank Ritika Khandeparkar for excellent research assistance in data retrieval and participants in the American Economic Association Second Annual National Conference on Teaching Economics for helpful comments.

characteristic of MOOCs since the beginning and there has been some interest in both documenting and understanding the phenomenon. This paper uses individual level time-stamped data from a nine-week Coursera course, Principles of Microeconomics, sponsored by the University of Pennsylvania, to analyze patterns of student attrition. For a subset of the students we use demographic information, collected via an opt-in survey, to enhance our understanding of these patterns.

Two questions motivate us. First, whether the patterns of engagement are set early, so that within a short period we can distinguish statistically between two groups of students: Browsers and Committed Learners. The second question is what demographic characteristics are associated with course retention and completion. To answer these questions, we apply a Cox proportional hazard model to the dropout rate from the course as measured by attrition from video watching and quiz taking.

We find that students who engaged in the course during the first week, by either completing the first quiz or by submitting a peer assessment exercise, were less likely to drop out of the course. Among students who took the first quiz, those who also completed the peer assignment were less likely to drop out of completing future quizzes. This suggests that the attrition pattern is not uniform among all enrollees, but rather there are distinct sub-groups of participants who reveal their type early on.

In the subsample of enrollees for whom we have demographic data, we find that those younger than 25 and college students are more likely to abandon the course. Female enrollees are also more likely to drop out as measured by video viewing. On the other hand, those who are over 35 are the most likely to stay in the course. Interestingly, among students who took the first quiz, neither female nor younger students, are more likely to drop out from future quiz submissions than others. This suggests that submitting the first quiz is a strong signal of commitment, since the differences between categories found in video retention disappear when looking at quiz retention, at least for those who took the survey.

As this is, to our knowledge, one of the first papers to attempt a sophisticated statistical analysis of retention in MOOCs, we take care to explain some technical issues in using the Cox proportional hazard model. The model assumes that covariates affect drop out in a proportional manner that is time invariant. We test the proportionality assumption by regressing the scaled Schoenfeld residuals on functions of time. In the cases where the zero-slope hypothesis is rejected, the proportionality assumption does not hold and we adjust our model by including the interaction term between the covariate and time. Though many of the details have been relegated to footnotes, we hope the brief summary included here will be helpful to this developing literature. Full specifications of all regressions are available from the authors.

The paper starts with a brief literature review of attrition from MOOCs. We then describe the Microeconomics Principles Coursera course and the data used. We follow with the statistical analysis of retention patterns and end with conclusions.

Literature Survey

High dropout rates have been a characteristic of MOOCs since the beginning and there has been some interest in both documenting and understanding the phenomenon. Koller and Ng (2013) state that the typical Coursera online course has between 40 and 60K students enrolled, out of whom only 15 to 20% submit an assignment. Of those, about 45% finish the course successfully by earning a statement of accomplishment. These figures are hard to compare with the traditional course structure where the overwhelming majority of students enrolled are expected to complete the course. Koller and Ng rightly explain that the variety of intents of online course takers is very different from the traditional model. In the online framework, some students do indeed plan on completing the course and can be referred to as “Committed Learners.” However, others may just want to understand what the topic is about or find out more about a particular professor or even about MOOCs as a whole. Those are the so-called “Browsers.” Looking at one specific course, Modern Poetry, taught by Al Filreis from the University of Pennsylvania, they observe that the decline in lecture watching is not matched by the exponential distribution, as they would expect if attrition from the course was only due to random life related events that get in the way of completion. They then look at a set of students taking 40 Coursera courses and model them as made up of two distinct groups: active and passive users. Under this assumption, the exponential function is a good predictor of retention in each group. They find a retention rate of 90 to 95 percent per hour of video watching or about 40 to 50 percent per month of lectures.

Katy Jordan (2014) explores enrollment and completion rates of mass online courses over the 18 months of their existence. She finds that though MOOCs are now quite as large as they once were, the completion rates, in most cases defined by earning a certificate of accomplishment, have somewhat increased over this period (though the majority of them still have completion rates below 10% with a median of 6.5%). She observes that across all course sizes, about half of all enrollees becomes active students. Completion rates are negatively correlated with course length and are lower when the course requires relatively more assignments, especially those involving programming or peer assessment exercises.

How to model dropout or its inverse, retention, is discussed in a number of papers. Clow (2013) suggests the idea of Funnel of Participation, where there is a significant drop-off at any stage, from enrollment to course completion. These steep drop-offs are seen as a barrier to the incorporation of MOOCs into use as an alternative for standard academic settings. One of the benefits of an institution, states Clow, is that there are people and resources to predict failure and skilled learning professionals to intervene preemptively. Perna et al (2014) explore further the drop-off points for a set of Coursera courses run by the University of Pennsylvania. They define ‘loss factor’ as the ratio of students who accessed the lecture in the first course module to those who accessed a lecture in the final module. They find a loss factor of 3 to 9 with a median of 6 across 16 Coursera courses. In none of these courses did more than 12% of registrants receive a final grade of 80% or higher.

There is an extensive literature on retention in credit academic online courses (reviewed in Hart 2012). Some of the factors mentioned in this literature as associated with persistence are more easily applicable to the MOOC environment. These include general satisfaction with online learning, motivation, time management skills and family support. Others may be more challenging to create in a MOOC, including peer support, a sense of belonging to a learning community and increased communication with instructor. Our study does not test these factors directly. However, students were asked in the survey about their initial motivation for taking this course and from our results it appears not to be important in explaining retention.

Data

The University of Pennsylvania joined as a partner with Coursera in 2012 and our nine-week course, Principles of Microeconomics, which ran during the spring of 2013, was one of the first courses to run as part of this partnership. Each week included 4-15 lecture video segments and a ten question multiple-choice quiz. There were three peer assessment exercises and three fill-in-the-blank short-answer assignments during the course, plus a multiple-choice final exam at the end. Students with a grade of 65% earned a certificate of completion in the course and those above 85% earned it with distinction.

Table 1 summarizes some basic course statistics that point to the stark difference between enrollment in a MOOC and a traditional course. In terms of size, there were almost 36 *thousand* students enrolled. This size, typical of a MOOC, is equivalent of enrolling a whole state university in one course concurrently. However, enrollment numbers are misleading. Of the thirty-six thousand enrollees, only 18 thousand logged in even once and only fifteen thousand watched, at a minimum, part of one lecture. Initial enrollment, in other words, is more akin to browsing a university brochure than registering for a live semester in campus. By the second week only 6,486 students remained and out of these 2,744 stayed to the end of the course and watched the last video of the last lecture in week nine. The final exam was attempted by 866 students. Only 739 earned a certificate of accomplishment, which is the recognition that a student successfully mastered the material covered in the course.

Table 1: Lecture Activity in Principles of Microeconomics Coursera Course
Spring 2013

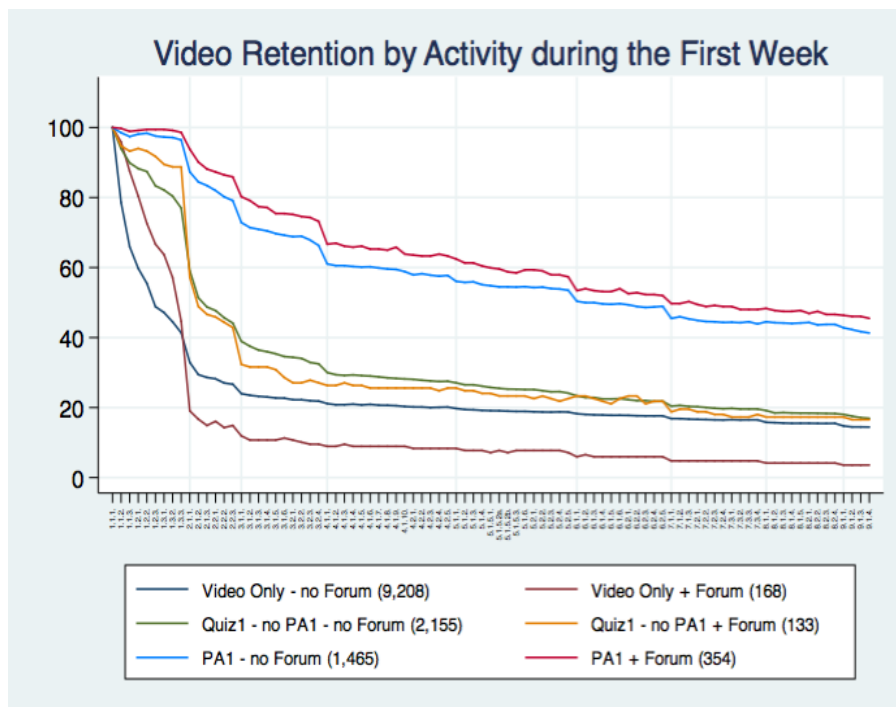
Enrolled in the course	35,819
Logged into the course at least once	18,009
Watched at least part of one video	15,427
Watched first video of the first week	13,483
Watched first video of the second week	6,486
Watched last video of the last week	2,744
Took the Final Exam	886
Earned the Statement of Accomplishment	739

In addition to watching videos, students engaged in the course in a variety of ways. This included participation in online forums (posting questions, responding to other posts, voting on posts to move them up or down on a priority list) as well as solving the weekly quizzes and peer-assessed assignments, assessing other students' assignments and taking the final exam. Table 2 summarizes course engagement data during the first week of the course.

Table 2: First Week's Activity

Watched the first video in the first week	13,483
Took First Quiz	4,571
Submitted First Peer Assessment	1,860
Graded First Peer Assessment (during 2 nd week)	1,457
Participated to Forum Activity	684

Only about a third of the students who watched at least one video took the first quiz and only 18% submitted the first peer assessment. As quizzes and assessment were a major component of the grade in the course, failing to take these suggests that most of those enrolled in this MOOC behaved as passive learners or auditors rather than the fully engaged students we expect in a live, credit bearing course.



The graph above plots the percentage of students who watched each video, out of the initial set who watched at least part of the first video, according to their level of activity during the first week of the course. We can see from this graph both the general downward trend of video watching and the differences across subgroups. In each population there is a decline in ‘attendance’ over the course, but the decline is the slowest for students who submitted the first open-ended peer-assessed assignment and the fastest for those who only watched videos. This graph suggests that activity in the very first week already reflects the students’ level of engagement, an issue that we will explore further in the results section.

Demographic data are not collected routinely on the Coursera platform. We included a voluntary opt-in survey that was available for completion throughout the course. As summarized in Table 3, out of the 3,567 who submitted the survey, 64% percent were male and the predominant age group was 22-35 years old. Over half had taken a Coursera course before. Almost half of the participants were working full time and the vast majority were participating in the course in addition to other big time commitments such as studies, part time work etc. Students represented 147 countries with the largest five being United States, India, China, Brazil, Spain.

In summary, our student-level dataset includes information on video watching, other measures of course participation and, for a subset of enrollees, some basic demographic information. As stated earlier, the question that motivated us was whether engagement patterns are set early so that we can distinguish statistically between “Committed Learners” and “Browsers”. Moreover, using the demographic data, we wanted to know whether there are statistical differences across demographic groups in terms of engagement and commitment to the course.

Table 3: Demographic Data Collected from Voluntary Survey (n=3,567)

Gender	
Male	64.40%
Female	35.60%
Coursera Experience	
First time Coursera User	48.13%
Experienced Coursera User	51.87%
Occupational Status	
Student (pre-college)	5.18%
Student (undergraduate)	17.27%
Student (post-graduate)	10.16%
Employed full-time	47.06%
Employed part-time	3.33%
Self-employed	7.31%

Unemployed	6.10%
Other	3.58%
Age	
Under 18	3.78%
18 - 21	12.81%
22 - 35	57.48%
36 - 50	17.88%
51 - 70	7.46%
71 – 90+	0.59%
Self-ascribed Motivation	
Learn the principles	23.18%
Expand my knowledge	37.62%
Preparation for future course	11.24%
Help with my job	11.27%
Help with my life skills	13.23%
Interested in a course with Penn and /or Rebecca Stein	5.45%

Model and results

We model the probability of dropping the course by means of a Cox proportional hazard model. The hazard rate is the probability of dropping out at time t conditional on surviving until that point. The proportionality assumes that there is an underlying baseline probability $\lambda_0(t)$ of dropping the course (due, for example, to random life events that interrupt the student's ability to continue) which is not directly affected by the independent variables. Rather, the impact of a unit increase in the independent variable is multiplicative with respect to the hazard or dropout rate. The Cox proportional hazard model has the form:

$$\lambda(t | X) = \lambda_0(t) \exp(\beta_1 X_1 + \dots + \beta_n X_n) = \lambda_0(t) \exp(\beta^*)$$

for the hazard at time t and an individual covariate vector X .

The independent variables in vector X , were taken to be either measures of course connectivity (as measured by the student's activity during the first week of class) or demographic data.

As the name suggests, the Cox proportional hazard model assumes that the covariates affect dropout in a proportional manner that is time invariant. The proportionality assumption can be tested by regressing the scaled Schoenfeld residuals on functions of time. We can think of Schoenfeld residuals as the observed minus the expected values of the covariates at each failure time. If the proportionality assumption holds, the slope of the Schoenfeld residuals with respect to time should not be significantly different from zero. In cases where the zero-slope hypothesis is rejected, the proportionality assumption does not hold so we adjust our model by including the interaction term between the covariate and time.

We measure dropping out of the course in two ways: no longer watching videos and no longer submitting quizzes. Obviously, the set of students in the latter specification is smaller, as it is conditional on taking the first quiz.

In all the tables below we provide the hazard ratio instead of the coefficient. For example, a hazard ratio of 1.25 on the dummy variable "female" would suggest that females are 25% *more* likely to drop out in any given period compared to males; a hazard ratio of .7 would suggest that they are 30% *less* likely to drop out.

Using demographic data is only possible for those who took the survey. The graph below suggests that survey participation is not random: students taking the survey are less likely to drop out.

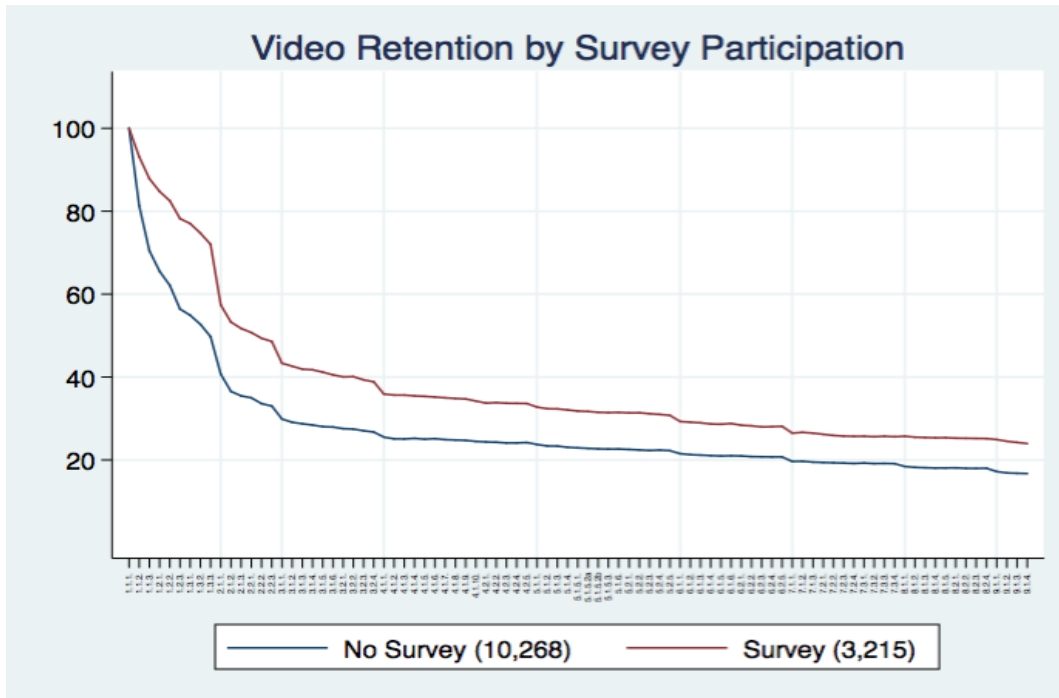


Table 4 confirms the previous statement. Using the Cox proportional hazard model (columns 1 and 3), we see that students taking the survey have a lower hazard of dropping out of the course. In terms of video watching, those taking the survey are 27% less likely to drop the course in each period, while in terms of quiz completion they are 35% less likely to drop the course. The same qualitative results still hold when we adjust the model to include the time interaction term (columns 2 and 4)². However, we find that the gap between those taking the survey and those not taking the survey evolves differently over time between the two measures of course retention. In terms of video retention the difference shrinks by .6% per period while for quiz retention it increases by 7% per period.

There could be two possible reasons underlying the different behavior between those who did and did not complete the survey. It is likely to reflect self-selection of the more determined students into the set of survey takers, but it could also be the result of an endowment effect created by the time cost of filling out the survey. If indeed it is the act of engaging in this simple activity that reduces attrition, this suggests that even small changes in interaction with the students could greatly improve the effectiveness of MOOCs.

² The proportional hazards assumption fails in both specifications with Chi2 of 39.08 for video retention and 15.55 for quiz retention.

Table 4: Video and Quiz Retention by Survey Participation

	Video		Quiz	
	(1)	(2)	(3)	(4)
Survey = 1	0.732***	0.660***	0.649***	0.757***
	0.017	(0.019)	0.022	(0.040)
Survey * time		1.006***		0.930***
		(0.001)		(0.017)
No. obs	13,483	13,483	4,571	4,571

Hazard ratios (Std Errors in Parentheses)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The selection issue involving the survey participation does not imply that the demographic information is unusable; rather, we should understand that survey participants are positively self-selected so any result in this sense applies better to the population of “Committed Learners” rather than “Browsers.”

Subject to this caveat, Tables 5a and 5b summarize the impact of demographic characteristics on the hazard of dropping the course as measured by video watching. We start with Table 5a by exploring the impact of motivation on retention. We run two specifications: one including age and occupation dummies and one without. Interestingly, we find that none of the motivation categories is associated with retention in a statistically significant way. Moreover, the F-test cannot reject the null hypothesis that all the categories have a coefficient equal to zero³.

In Table 5b we explore more carefully the impact of age and occupation. Because age and occupation categories tend to overlap, we run three separate specifications. Specification (1) includes only prior Coursera experience, gender and age. Specification (2) includes only prior Coursera experience, gender and occupation and Specification (3) includes prior Coursera experience, gender, age and occupation. In each case the omitted age category is 22-35 and the omitted occupation is full time employment. All the specifications still include the motivation dummies but the coefficients are not reported for sake of brevity.

³ The corresponding Chi2 are 5.18 and 6.75 for columns 1 and 2 respectively.

Table 5a: Video Retention by Demographics (Focus on Motivation)

	(1)	(2)
Experienced Coursera = 1	0.787*** (0.032)	0.809*** (0.033)
Female = 1	1.159*** (0.050)	1.154*** (0.050)
Learn the Principles	1.072 (0.058)	1.067 (0.058)
Preparation for Future Course	0.988 (0.068)	0.918 (0.065)
Help with my Job	0.929 (0.065)	0.965 (0.069)
Help with my life skills	1.059 (0.068)	1.081 (0.070)
Interested in a course with Upenn/ Dr Stein	1.092 (0.125)	1.103 (0.127)
Age Dummies	No	Yes
Occupation Dummies	No	Yes
No. obs	3,177	3,167

Hazard ratios (Std Errors in Parentheses)

* p<0.1, ** p<0.05, *** p<0.01

Omitted motivation: Expand my Knowledge

Table 5b: Video Retention by Demographics (Focus on Age and Occupation)

	(1)	(2)	(3)
Experienced Coursera =1	0.805*** (0.033)	0.803*** (0.033)	0.809*** (0.033)
Female = 1	1.136*** (0.049)	1.159*** (0.050)	1.154*** (0.050)
Under 18	1.383*** (0.140)		1.100 (0.189)
18-21	1.296*** (0.080)		1.099 (0.097)
36-50	0.837*** (0.048)		0.852*** (0.050)
51-70	0.907 (0.073)		0.970 (0.081)
71-90+	1.095 (0.261)		1.499 (0.384)
Student Pre-College		1.406*** (0.128)	1.237 (0.196)
Student Undergraduate		1.287*** (0.073)	1.169** (0.093)
Student Graduate		0.957 (0.068)	0.926 (0.067)
Employed Part-Time		0.964 (0.114)	0.962 (0.114)
Self-Employed		0.904 (0.075)	0.925 (0.079)
Unemployed		0.865 (0.078)	0.862 (0.078)
Other		0.717*** (0.085)	0.663*** (0.085)
Motivation Dummies	Yes	Yes	Yes
No. obs	3,171	3,172	3,167

Hazard ratios (Std Errors in Parentheses)

* p<0.1, ** p<0.05, *** p<0.01

Omitted age: 22-35. Omitted occupation: Employed Full-Time

In all three specifications, we can see that prior experience reduces the hazard of dropping out of the course by about 20% at any time. Females are between 13 and 15% more likely to drop out of the course in any given period. Younger participants (specification 1) and those who are full time pre college or undergraduate students (specification 2) are also more likely to drop out of the course. The likelihood to stick with the course is increasing with age. Those between the ages of 36-50 are 15% more likely to continue in any given period compared to those 22-35, while those younger than 22 are about 30% less likely to. Once we include both age and occupation dummies in the same regression (specification 3), the coefficients for the age category “Under 18” and the occupation category “Student Pre-College” lose significance. This is not surprising since the two categories highly overlap. The joint test rejects the null hypothesis, confirming our intuition⁴.

We run also a regression including interaction terms between all the dummy variables and gender. For sake of brevity, we do not report the complete results since they are qualitatively and quantitatively similar to the previous ones. However, we highlight some interesting facts. While male students who enrolled because interested in attending a course with Upenn and/or Dr Stein are 30% **more** likely to drop-out than male students who enrolled because willing to expand their knowledge about Microeconomics, we find that female students who enrolled because interested in attending a course with Upenn and/or Dr Stein are 44% **less** likely to drop-out than female students who enrolled because willing to expand their knowledge about Microeconomics. A future study that includes a range of courses could pursue this result further to test whether female students are motivated to stick with a course that is presented by a female instructor. We find another interesting gender effect in the interaction coefficient with part time work. There is no statistically significant difference between part-time and full time workers within male students, but among female students, those who work part-time are 41% less likely to drop-out than those who work full-time (the coefficient is statistically different from zero at the 5% level of significance). This is consistent with many full-time working females having more time constraints than male counterparts: for the former adding an on-line course is clearly a bigger challenge than for the latter.

Table 6 includes results from a similar specification except that the dependent variable is quiz retention⁵. Since many of the coefficients that were significantly different from zero for video retention are not for quiz retention, taking the first quiz appears to be a strong signal of commitment to the course. Specifically, the coefficient on females is no longer statistically significant, suggesting that female students who engaged in the first quiz are

⁴ The Chi2(2) of the F-test on coefficients “Under 18” and “Student Pre-College” is 9.21

⁵ Here again we looked at motivation but none of the categories is statistically significant and we fail to reject the null hypothesis of the coefficients being all equal to zero.

not different from those who did not. Still, young students, by this measure too, are more likely to drop out as are, in this case, the very old.⁶

⁶ The regression including the interaction terms between each dummy and gender confirms the previous results. The only coefficient on the interaction terms that is significantly different from zero is the one relating gender and motivation “Help with life skills”. Female students who enrolled because they thought the course could be helpful with their life skills are 65% more likely to drop-out in terms of quiz retention than the female students who enrolled because willing to expand their knowledge about Microeconomics.

Table 6: Quiz Retention by Demographics (Focus on Age and Occupation)

	(1)	(2)	(3)
Experienced Coursera =1	0.940 (0.051)	0.947 (0.051)	0.949 (0.051)
Female = 1	1.075 (0.061)	1.088 (0.062)	1.091 (0.063)
Under 18	1.226 (0.166)		1.038 (0.234)
18-21	1.246*** (0.101)		1.169 (0.138)
36-50	0.938 (0.069)		0.949 (0.071)
51-70	0.937 (0.099)		1.006 (0.110)
71-90+	1.526 (0.466)		2.220** (0.732)
Student Pre-College		1.243* (0.149)	1.139 (0.236)
Student Undergraduate		1.118 (0.084)	1.005 (0.107)
Student Graduate		0.892 (0.085)	0.878 (0.085)
Employed Part-Time		1.028 (0.159)	1.025 (0.159)
Self-Employed		0.823* (0.090)	0.824* (0.092)
Unemployed		0.846 (0.098)	0.841 (0.097)
Other		0.698** (0.101)	0.615*** (0.098)
Motivation Dummies	Yes	Yes	Yes
No. obs	1,989	1,992	1,987

Hazard ratios (Std Errors in Parentheses)

* p<0.1, ** p<0.05, *** p<0.01

Omitted age: 22-35. Omitted occupation: Employed Full-Time

To see whether engagement during the first week impacts the hazard of dropping out we run a similar model where the independent variables are measures of the first week activity including quiz participation, completion of peer-assessed assignment and participation in a forum. Table 7 summarizes the results of these specifications. Once again we must reject the proportional hazard assumption⁷ so we provide both time invariant and time variant coefficients.

We can see that students taking the first week's quiz, whether or not they participated in a forum, were more than 30% less likely to drop the course compared to students who only watched the videos during the first week. Students taking the first week's peer-assessed assignment, whether or not they participated in a forum, were more than 60% less likely to drop the course compared to students who only watched the videos during the first week. This suggests that those enrolled in a MOOC belong indeed to different types and that activity in the first few days indicates their commitment to the course.

The hazard ratios in column (2) can be interpreted as follows: at the beginning of the course, students who took the first quiz (with no participation in the forum and no peer assessment) are 50% less likely to drop out than students who only watched the first video. However, the difference shrinks by 2.1% in each period. This suggests that over time the first week's activity loses its importance as an indicator of attachment to the course, as we should expect since those least committed to the course have dropped out.

Looking at participation in the online forum, we observe in column (1) the curious result that students who posted in the forum were 28% more likely to drop out of the course. However, once we add time interactions, we see that this effect is due to their large drop off between the first and the second week, as indeed is reflected in Figure 1. Why are these students, who took the time to participate in the forum during the first week, dropping out of the course is worth pursuing further. Our conjecture is that they used the forum to voice their frustration about the platform or the peer assessment and thus the postings and their dropping out are both expressions of the same barriers to learning.

⁷ The proportionality assumption is rejected in the model of video retention by first week activity while it holds in case of quiz retention by first week activity. The test on each separate covariate is consistent with the global test. Results are available from the authors.

Table 7: Video and Quiz Retention by First Week Activity

	Video		Quiz	
	(1)	(2)	(3)	(4)
Video Only - Forum	1.278*** (0.102)	1.043 (0.099)		
Quiz1 - no PA1 - no Forum	0.690*** (0.018)	0.502*** (0.017)		
Quiz1 - no PA1 - Forum	0.689*** (0.066)	0.520*** (0.063)	0.908 (0.083)	0.945 (0.133)
PA1 - no Forum	0.359*** (0.013)	0.207*** (0.011)	0.384*** (0.015)	0.411*** (0.026)
PA1 - Forum	0.320*** (0.023)	0.169*** (0.019)	0.338*** (0.025)	0.328*** (0.040)
Video Only - Forum * t		1.019*** (0.005)		
Quiz1 - no PA1 - no Forum * t		1.021*** (0.001)		
Quiz1 - no PA1 - Forum * t		1.019*** (0.004)		0.979 (0.055)
PA1 - no Forum * t		1.027*** (0.001)		0.972 (0.020)
PA1 - Forum * t		1.029*** (0.003)		1.007 (0.035)
No. obs	13,483	13,483	13,483	4,571

Hazard ratios (Std Errors in Parentheses)

* p<0.05, ** p<0.01, *** p<0.001

Omitted: Video Only - no Forum; Quiz1 - no PA1 - no Forum

Conclusion and Future Work

This paper uses data from one Coursera course to try and understand better the characteristics of student course engagement. We find that patterns of engagement are set early, so that within the first week we can distinguish statistically between two groups of students: Browsers and Committed Learners. Students who engaged in the course during the first days, by either completing the first quiz or submitting a peer assessment exercise, were less likely to drop out of the course. Among students who took the first quiz, those who also completed the peer assignment were less likely to drop out of completing future quizzes. Interestingly, the students who participated in the forum, but did not complete the first quiz or assignment, were more likely to drop out of the course. Time interactions suggest that the result is driven by a sharp drop off in participation between the first and second week. We conjecture that these initial forum posting may be voicing frustrations about the Coursera platform or course specific requirements that are set early on.

The results show how raw completion rates cannot be the criterion to measure the performance of a MOOC. The population of interest is, in many cases, not the entirety of enrollees but only the Committed Learners, i.e. the students who actively engaged in some activity other than video watching during the first week, showing their commitment to the course. This group is a better basis for comparison with the students in the traditional classroom. In the subsample of enrollees for whom we have demographic data, we find that younger participants (under age 25), full time college students and females are more likely to drop out of the course while those between the ages of 36-50 as well as those with prior experience with Coursera are less likely to drop out. We do not find any differences across the various motivation categories of taking the course.

The self-reported demographic data was collected by means of a voluntary survey. We observe that the participation in the survey is in and of itself a strong indicator of commitment to the course. A future study, that makes the survey available only to a random group of students, may allow us to distinguish between self-selection of those who completed the survey and an endowment effect created by the time cost of filling out the survey.

Many additional questions still remain for future work. We would like to explore further why students who participated in a forum were more likely to drop out of the course. As this study reflects data from only one course, we would also like to extend the analysis to other courses. This will give us not only more statistical power but also the opportunity to explore the importance of subject matter, course format and instructor characteristics (e.g., gender) and their impact on course retention.

Bibliography

2013 Doug Clow, MOOCs and the Funnel of Participation, *presented at the third conference on learning analytics and knowledge, Leuven, Belgium.*

2012 Carolyn Hart, Factors Associated with student Persistence in an Online Program of Study: A review on the Literature. *Journal of Interactive online Learning.* Vol 11, no 1

2014 Katy Jordan, Initial Trends in Enrolment and Completion of Mass Open Online Courses *The International Review of Research in Open and Distance Learning* Vol 15 no 1

2013 Daphne Koller, Andrew Ng, Chuong Do, and Zhenghao Chen. Retention and Intention in Massive Open Online Courses: In Depth. *Educause Review.*

2014 Laura Perna, Alan Ruby, Robert Boruch, Nicole Wang, Janie Scull, Seher Ahmad and Chad Evans. The Life Cycle of a Million MOOC Users. *Working paper.*