Types of Participant Behavior in a Massive Open Online Course

Tali Kahan, Tal Soffer, and **Rafi Nachmias**
Tel Aviv University

Abstract

In recent years there has been a proliferation of massive open online courses (MOOCs), which provide unprecedented opportunities for lifelong learning. Registrants approach these courses with a variety of motivations for participation. Characterizing the different types of participation in MOOCs is fundamental in order to be able to better evaluate the phenomenon and to support MOOCs developers and instructors in devising courses which are adapted for different learners' needs. Thus, the purpose of this study was to characterize the different types of participant behavior in a MOOC. Using a data mining methodology, 21,889 participants of a MOOC were classified into clusters, based on their activity in the main learning resources of the course: video lectures, discussion forums, and assessments. Thereafter, the participants in each cluster were characterized in regard to demographics, course participation, and course achievement characteristics. Seven types of participant behavior were identified: *Tasters* (64.8%), *Downloaders* (8.5%), *Disengagers* (11.5%), *Offline Engagers* (3.6%), *Online Engagers* (7.4%), *Moderately Social Engagers* (3.7%), and *Social Engagers* (0.6%). A significant number of 1,020 participants were found to be engaged in the course, but did not achieve a certificate. The types are discussed according to the established research questions. The results provide further evidence regarding the utilization of the flexibility, which is offered in MOOCs, by the participants according to their needs. Furthermore, this study supports the claim that MOOCs' impact should not be evaluated solely based on certification rates but rather based on learning behaviors.

Keywords: massive open online course, types of participant behavior, educational data mining, cluster analysis

Introduction

In today's information society, knowledge has become a central resource and a major parameter in the labor market, thus it is of great importance for the individual. Acquiring education and knowledge is a fundamental human right, as declared by the United Nations' Universal Declaration of Human Rights:

"Everyone has the right to education" (United Nations [UN], 1948). Unfortunately, higher education is not yet a public domain. Moreover, studying in the higher education usually constitutes a relatively short period within one's lifetime. Thus, lifelong learning becomes an important aspect in the twenty-first century.

 In 2012, the Paris OER Declaration of UNESCO called governments to promote the use of open educational resources to widen access to education in a perspective of lifelong learning (UNESCO, 2012). Alongside, a new means for engaging in lifelong learning has emerged in the form of massive open online courses (MOOCs). Initiatives like Coursera, edX, and Udacity provide platforms for higher education institutions to develop and deliver online courses to the general public. The courses are usually offered free of charge, with no preconditions or commitment. From a pedagogical perspective, the courses (sometimes referred to as xMOOCs) usually follow a cognitive-behaviorist approach (Daniel, 2012; Rodriguez, 2012) which is exemplified by content-based learning delivered at scale (Anders, 2015). The courses consist of diverse learning resources, including video lectures, discussion forums, and assessments. Some courses give participants a statement of accomplishment upon successful completion of course requirements. Hence, MOOCs provide unprecedented opportunities for lifelong learning, by enabling the delivery of knowledge from well-known institutions to people worldwide.

Since their appearance, the MOOCs have attracted a massive number of registrants. However, a central criticism in the popular discourse about MOOCs refers to the relatively low completion rates of participants, with 10% or less of the course registrants earning a statement of accomplishment (Daniel, 2012; Kizilcec, Piech, & Schneider, 2013). It is important to note, however, that unlike the higher education arena, in which the vast majority of students enroll to courses with the explicit intent of earning a credential, students approach MOOCs with a variety of motivations for participation (Koller, Ng, Chuong, & Chen, 2013; Wang & Baker, 2015). These may include a drive for intellectual stimulation, fun and enjoyment, social experience, trying out learning online, and so on (Belanger & Thornton, 2013; Ferguson & Clow, 2015). The range of motivations leads to diverse learners' needs, behaviors, and persistence in the courses. Thus, as indicated by Ho et al. (2014) certification rates is a misleading representation of the diverse ways in which registrants are engaging with MOOCs; earning a certificate is only one possible pathway, while others may include watching videos, reading texts, focusing on assessments, and so on.

Understanding the different ways in which registrants are engaging with MOOCs is fundamental in order to be able to evaluate the phenomenon and its impact in delivering lifelong learning on a large-scale. Moreover, such an understanding is essential for MOOC developers and instructors in order to be able to devise courses, which are adapted for different learners' needs. Subsequently, the purpose of this study was to identify the different types of participant behavior in a MOOC and to characterize each type according to demographics, course participation, and course achievements. The data collected from participant interactions with MOOCs open up opportunities for studying students' engagement on a large scale (Ramesh, Goldwasser, Huang, Daume, & Getoor, 2014). Hence, the study was conducted using educational data mining approach (Baker & Siemens, 2014). The participants of one Coursera MOOC were classified into types, based on a detailed description of their activity in the main learning resources of the course: video lectures, discussion forums, and assessments. Thereafter, the participants in each

type were characterized in regard to demographics, course participation, and course achievement characteristics.

Background

MOOCs Learning Resources

Video lectures, reading materials, discussion forums, and assessments are core learning resources in MOOCs (Glance, Forsey, & Riley, 2013; Nkuyubwatsi, 2013). The videos are central to the student learning experience (Diwanji, Simon, Marki, Korkut, & Dornberger, 2014; Guo, Kim, & Rubin, 2014). MOOCs usually consist of short video lectures interspersed with interactive assessment items (Seaton, Nesterko, Mullaney, Reich, & Ho, 2014). Several studies have examined video usage by MOOC participants. Breslow et al. (2013), for example, found that certificate earners of the first edX MOOC spent the majority of their time watching videos. Seaton et al. (2014) found two modes of video usage by certificate earners in MITx courses: bimodal and high use (characterized via unique lecture video accesses). Other studies examined patterns of student interaction with the videos, and their relation to student performance (Li, Kidzi´nski, Jermann, & Dillenbourg, 2015; Sharma, Jermann, & Dillenbourg, 2014).

The discussion forums provide a platform for asynchronous communication that facilitates interactions among students and instructors (Wong, Pursel, Divinsky, & Jansen, 2015). The forums support the integration of social-constructivist pedagogies by enabling collaborative and social learning (Anders, 2015; Brinton et al., 2013) and help to create a learning community through which learners generate knowledge (Li, 2004). Due to the large scale participation versus the low number of instructors in MOOCs, peer communication and support become central (Onah, Sinclair, & Boyatt, 2014), and the discussion forums constitute a primary means of interaction among the participants. Nevertheless, studies described the usage of discussion forums in MOOCs as quite low in general, often involving a minority of course participants (Breslow et al., 2013; Ho et al., 2014; Onah et al., 2014). In addition, several studies indicated that certificate earners are significantly more active in the forums than non-certificate earners (Breslow et al., 2013; Ho et al., 2014; Kizilcec et al., 2013; Mustafaraj & Bu, 2015).

The assessments in MOOCs may serve different goals, due to the open nature of the courses, such as self-testing or formal assessment for course credit (Woodgate, Macleod, Scott, & Haywood, 2015). With the massive number of participants in a course, conducting assessments by the instructors is impossible (Glance et al., 2013; Sandeen, 2013; Suen, 2014) and different models of assessments are evolving, including: automated quizzes, peer assessments, and self-assessments. Nevertheless, studies indicated that with free and easy registration for MOOCs, the courses include a large number of participants who may not have any interest in completing the assessments (Breslow et al., 2013; Ho et al., 2014).

Types of Participant Behavior in MOOCs

Several studies have examined types of participant engagement in MOOCs, based on different criteria. Kizilcec et al. (2013), for example, examined patterns of engagement in three Coursera MOOCs, based on participant actions in regard to videos and assessments, and identified four prototypical trajectories of engagement: *Completing* - learners who completed the majority of the assessments; *Auditing* - learners who did assessments infrequently, if at all, and engaged by watching video lectures, without obtaining course credit; *Disengaging* - learners who did assessments at the beginning of the course but then had a marked decrease in engagement; and *Sampling* - learners who watched video lectures for only one or two assessment periods.

Ferguson and Clow (2015) investigated whether the patterns of engagement, identified in the work of Kizilcec et al. (2013), are found in MOOCs that employ a social constructivist pedagogy. They examined four FutureLearn MOOCs and added a third component to the analysis: participation in course discussions. Seven distinct patterns of engagement were identified: *Samplers* - learners who visited, but only briefly; *Strong Starters* - learners who completed the first assessment of the course, but then dropped out; *Returners* - learners who completed the assessment in the first two weeks, and then dropped out; *Mid-way Dropouts* - learners who completed three or four assessments, but then dropped out about halfway through the course; *Nearly There* - learners who consistently completed assessments, but then dropped out just before the end of the course; *Late Completers* - learners who completed the final assessment and submitted most of the other assessments, but were either late or missed some out; and *Keen Completers* - learners who completed the course diligently, engaging actively throughout.

Ho et al. (2014) examined the first 17 MOOCs of the edX platform and presented a classification of the registrants, which was comprised of four categories: *Only Registered* - registrants who never accessed the courseware; *Only Viewed* - non certified registrants who accessed less than half of the available chapters; *Only Explored* - non certified registrants who accessed more than half of the available chapters; and *Certified* - registrants who earned a certificate. The researchers also examined the demographic profile of the participants. The most typical course registrant was found to be male, with a bachelor's degree, and 26 years old or older. Yet, they found considerable differences in the average demographics across courses, in terms of gender, college degree attainment, median age, and percentage from the US.

Finally, Halawa, Greene, and Mitchell (2014) identified four common patterns of persistence in MOOCs, based on the participants' frequency of course visits: *Continuous Persistence* - students who visited the course once every few days, at most; *Continuous Persistence with Extended Absences* - students who followed a similarly smooth trajectory, except that there were one or more extended absences, after which the student continued from where he/she stopped previously; *Bursty Persistence* - students who only visited the course occasionally, and usually sampled content from different units each day they visited; and *Drop Out* - students who started off as *Continuous* or *Bursty* visitors, but disappeared totally after a certain point before the end of the course.

As emerged from the literature review, video lectures, discussion forums, and assessments are fundamental learning resources in MOOCs. Thus, analyzing a participant's activity in these components reflects their behavior in the course. In this study, we characterized types of participant behavior in a MOOC by classifying the participants of a course into clusters, based on a detailed description of their

activity in these main learning resources. Thereafter, the participants in each cluster were characterized in regard to demographics, course participation, and course achievement characteristics. This study uses a holistic approach which adds to the existing literature in two ways: it examines the overall participant activity throughout the entire course, while other studies tended to examine periodical participant activity in the course (Ferguson & Clow, 2015; Kizilcec et al., 2013). In addition, it uses a wide set of different variables to describe the basic participant activity in the course. Specifically, it refers to the participants' activity in: watching video lectures (online or offline), answering in-video questions, participating in discussion forums (actively or passively) and submitting course assessments (quizzes and exam).

Research Questions

The purpose of this study was to characterize the different types of participant behavior in a MOOC. The research questions were:

1. What are the types of participant behavior in the course, based on participant activity in the video lectures, discussion forums, and assessments?

2. What are the characteristics of each type of participant, in regard to: demographics, course participation, and course achievements?
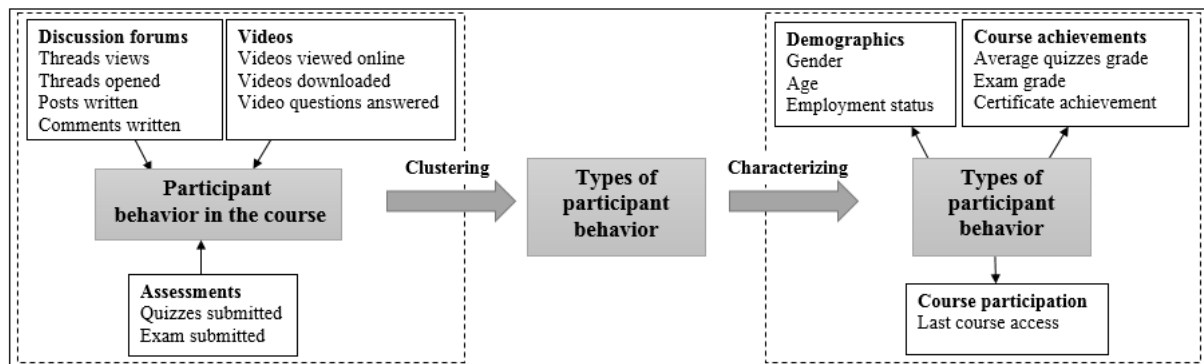
Figure 1 presents the research framework.

Figure 1. The research framework.

Method

The Course

The study examined one Coursera MOOC on biology, which lasted seven weeks and consisted of diverse learning resources, including: professor announcements, reading recommendations, 50 short video lectures, 39 interactive in-video questions, seven discussion forums, six quizzes, and a final exam. The video lectures were uploaded to the course website on a weekly basis and the participants could watch them online and/or download them to watch offline. The in-video questions were presented in online

mode only. The discussion forums were arranged into seven sub forums according to topics. The course assessments were based on weekly quizzes and a final exam. The participants were given several attempts to complete each quiz, and were required to submit it within one week of its release, in order to receive credit for it. The quizzes remained accessible throughout the course, enabling the participants to use them for practice. The final exam was published during the last week of the course and consisted of close-ended questions. Participants who completed the course with a grade of at least 78%, comprised of 50% quizzes and 50% final exam, received a statement of accomplishment.

Population

Out of a total of 32,007 people who registered for the course, 68.4% of them (21,889) started it. The study examined the behavior of all the participants who started the course. Of this participant group, 10.6% (2,319) completed the course and achieved a statement of accomplishment. According to Coursera's demographic survey (N=4,778), the course registrants consisted of 54% females and 46% males. The average respondent's age was 39 with a standard deviation of 14 years. Sixty-three percent of the respondents stated that they were working.

Procedure

The study was conducted using educational data mining and statistics methods. The data was recorded by Coursera during the course and was received from the company after the course ended. Three data sources were used: (1) Log data in SQL tables that documented participant actions during the course (e.g., video lecture views, quiz submissions, discussion forum views). The data contained over 1 million records in total; (2) Log data in Excel files that documented participant responses to in-video questions; and (3) An Excel file with participant responses to Coursera's demographic survey.

The study was executed in several stages. First, a set of SQL queries and Excel functions was written in order to compute a set of variables from the log data, for each participant, describing his/her activity in the course, in regard to learning resources usage, course participation, and course achievements. The demographic variables were extracted from the demographic survey responses file. All the variables were then merged into a unified table, consisting of one row per participant and one column per variable. The variables are described in Table 1.

Table 1

The Variables That Were Computed per Participant

Variable name	Variable description
unique video lectures viewed online	The number of different video lectures viewed by the participant online.
unique video lectures downloaded	The number of different video lectures downloaded by the participant.
unique video questions answered	The number of different in-video questions answered by the participant.
total threads views	The number of times the participant viewed threads in the discussion forums.

total threads opened	The number of threads opened by the participant in the discussion forums.
total posts written	The number of posts written by the participant in the discussion forums.
total comments written	The number of comments written by the participant in the discussion forums.
unique quizzes submitted	The number of different quizzes submitted by the participant.
exam submitted	Whether or not the participant submitted the exam.
last course access	The last day in which the participant entered the course website (represented as the number of days from the beginning of the course).
average quiz grade	The average grade of all quizzes submitted by the participant (referring to the highest grade achieved by the participant for each quiz).
exam grade	The participant's grade on the exam.
achieved certificate	Whether or not the participant achieved a certificate in the course.
gender	
age	
employment status	Whether or not the participant is working ("Not working" employment status refers to registrants who selected one of the following: homemaker, taking care of a family member, on maternity/paternity leave; retired; unable to work; unemployed and looking for work; unemployed and not looking for work).

Next, in order to identify types of participant behavior in the course, a cluster analysis was applied. Cluster analysis is an exploratory data mining approach, which enables the discovery of structure in data without an a priori idea of what should be found. The analysis enables to find data points that naturally group together, splitting the data set into a set of clusters (Baker & Siemens, 2014). Hence, in this study a cluster analysis was used to identify groups of participants, such that all the participants in the same group exhibit similar behavior in the learning resources, than the participants in the other groups. The cluster analysis included nine variables that describe the basic participant activity in the main learning resources: video lectures - *unique video lectures viewed online, unique video lectures downloaded, unique video questions answered,* discussion forums - *total threads views, total threads opened, total posts written, total comments written,* assessments - *unique quizzes submitted, and exam submitted*. Specifically, the Two-Step clustering procedure was used. The first step of this procedure is the formation of preclusters. In the second step, a hierarchical clustering algorithm is used on the preclusters (Norusis, 2012). The Two-Step clustering method was selected because it is the suitable method for handling large data files with a mixture of continuous and categorical variables (Norusis, 2012), as was the case in this study (*exam submitted* being a categorical variable). The analysis was executed using the log-likelihood distance measure, which is the only criterion that can be used when the data contains a mixture of continuous and categorical variables (Norusis, 2012). The silhouette coefficient score, a measure of the cohesion within a cluster and separation between the clusters, was used to quantify the "goodness" of the clustering. The coefficient ranges from -1 to +1, such that a score is considered "good" if it is over 0.5,

"fair" if it is between 0.2 and 0.5, and "poor" if it is smaller than 0.2 (Kaufman & Rousseeuw, 2009; Norusis, 2012).

Descriptive statistics were then computed per cluster in regard to demographics (*gender, age, employment status*), course participation (*last course access*), and course achievement variables (*average quiz grade, exam grade, achieved certificate*). In both questions, ANOVA tests were used to test for differences between the clusters in regard to the interval variables and a chi-squared test was used in order to test for dependence between the clusters and the categorical variables. Since the clustering variables were non-normally distributed, the ANOVA analyses were conducted using a bootstrapping procedure with 1000 samples drawn from the data set and 95% confidence intervals (Mooney & Duval, 1993). The effect size was measured via Partial Eta-Squared for the ANOVA tests and Phi and Cramer's V for the chi-squared test.

Results

Types of Participant Behavior in the Course

The cluster analysis was applied in order to identify types of participant behavior in the course. After running the analysis several times and exploring different number of clusters to be formed, a model with seven clusters was selected. This model achieved a good silhouette coefficient score of 0.6. Moreover, the resulting model seemed to be the most informative and exhaustive model from an educational perspective – distinguishing between participants behaviors in the course in high level. The ANOVA tests showed statistically significant differences between the clusters in regard to the interval variables ($ps < 0.001$), and the chi-squared test showed statistically significant dependence between the clusters and the categorical variable ($p < 0.001$). Table 2 presents descriptive statistics of the variables that were used for the clustering, for each cluster, and the results of the ANOVA and the chi-squared tests.

Table 2

The Clusters Obtained From the Cluster Analysis

	Population	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	
number of participants (percentage in the population)	21,889 (100%)	14,186 (64.8%)	1,857 (8.5%)	2,507 (11.5%)	778 (3.6%)	1,627 (7.4%)	799 (3.7%)	135 (0.6%)	
				Average (standard deviation)					*F*(6, 21,882) (effect size)
unique video lectures downloaded	6.89 (15.73)	0.95 (3.32)	46.07 (7.44)	2.87 (9.30)	31.39 (22.17)	1.87 (5.57)	18.92 (23.00)	15.64 (21.59)	10,530.65** (0.74)
unique video lectures viewed online	12.99 (17.15)	4.47 (5.55)	4.17 (8.09)	34.63 (12.22)	13.07 (15.53)	48.22 (4.46)	40.13 (15.50)	41.59 (14.59)	13,452.62** (0.78)

								X²(6) (effect size)
unique video questions answered	8.53 (12.27)	2.57 (3.62)	1.81 (4.49)	25.23 (8.64)	4.80 (7.35)	32.98 (6.41)	28.10 (12.40)	28.78 (12.47)	14,450.73** (0.79)
total threads views	5.80 (29.74)	0.97 (2.95)	1.37 (5.15)	4.93 (8.47)	8.83 (14.06)	12.93 (14.49)	45.95 (44.45)	249.68 (231.14)	3,720.03** (0.50)
total threads opened	0.04 (0.32)	0.00 (0.06)	0.00 (0.05)	0.00 (0.28)	0.00 (0.00)	0.00 (0.00)	0.66 (0.71)	2.31 (2.35)	3,081.73** (0.45)
total posts written	0.31 (1.81)	0.05 (0.27)	0.04 (0.24)	0.20 (0.57)	0.19 (0.52)	0.45 (0.88)	3.05 (2.63)	15.50 (13.66)	4,104.59** (0.53)
total comments written	0.25 (3.46)	0.05 (0.28)	0.03 (0.23)	0.19 (0.61)	0.13 (0.49)	0.34 (0.80)	1.49 (2.56)	18.01 (39.58)	742.00** (0.16)
unique quizzes submitted	1.38 (2.21)	0.30 (0.67)	0.61 (1.34)	2.44 (2.13)	5.41 (1.65)	5.90 (0.64)	5.28 (1.62)	5.62 (1.22)	11,156.98** (0.75)

	Mode (frequency)							X²(6) (effect size)
exam submitted	No (86%)	No (100%)	No (100%)	No (100%)	Yes (100%)	Yes (100%)	Yes (78%)	Yes (85.2%)	20,634.48** (0.97)

Seven types of participant behavior emerged from the analysis. The first type appeared in cluster 1 that consists of 64.8% of the course participants (14,186 participants). This cluster is characterized by very low average values in all variables, which indicates very low activity in all the main learning resources of the course. The participants in this cluster were thus named the *Tasters*.

The second type appeared in cluster 2 that consists of 8.5% of the course participants (1,857 participants). Similarly to the first cluster, it is characterized by very low average values in all variables, except for the variable *unique video lectures downloaded*. The participants in this cluster were mostly inactive in the course, but they downloaded a very large portion of the video lectures (around 92% of the videos on average). Thus, they were named the *Downloaders*.

The third type appeared in cluster 3 that consists of 11.5% of the course participants (2,507 participants). These participants watched approximately 70% of the video lectures online, answering around 65% of the in-video questions, and submitted around 40% of the quizzes - on average. They entered the discussion forums a few times on average (4.93), mostly for observation, and none of them submitted the final exam. They were thus named the *Disengagers*.

The fourth type appeared in cluster 4 that consists of 3.6% of the course participants (778 participants). These participants demonstrated high levels of engagement in the course. They tended to download the video lectures (60% of the videos on average) rather than to watch them online (25% of the videos on average), they submitted almost all the assessments (5.41 on average), and entered the discussion forums several times on average (8.83), mostly for observation. All of them submitted the final exam. They were named the *Offline Engagers*.

The fifth type appeared in cluster 5 that consists of 7.4% of the course participants (1,627 participants). Similarly to the *Offline Engagers*, they demonstrated high levels of engagement in the course, but as opposed to them, they tended to watch most of the video lectures online (96% of the videos on average) and to answer most of the in-video questions (85%) - on average. They submitted almost all the assessments (5.90 on average), and entered the discussion forums threads a few more times on average (12.93), mostly for observation. All of them submitted the final exam. They were thus named the *Online Engagers*.

The sixth type appeared in cluster 6 that consists of 3.7% of the course participants (799 participants). Similarly to the *Offline* and the *Online Engagers*, these participants demonstrated high levels of engagement in the course. They watched around 80% of the video lectures online and answered around 72% of the in-video questions – on average. In addition, they submitted almost all the quizzes (5.28 on average). However, most prominent in this cluster is the participants' activity in the discussion forums, which is higher than the previous clusters, yet still moderate: they viewed the discussions 45.95 times, posted 3.05 messages and wrote 1.49 comments - on average. Most of them (78%) submitted the final exam. They were thus named the *Moderately Social Engagers*.

Finally, the seventh type appeared in cluster 7 that consists of 0.6% of the course participants (135 participants). Similarly to the previous cluster these participants demonstrated high levels of engagement in the course; they watched around 83% of the videos online, answered around 74% of the in-video questions, and submitted almost all the quizzes (5.62) - on average. Notably, these participants demonstrated the highest levels of activity in the discussion forums: they viewed the discussions 249.68 times, posted 15.50 messages and wrote 18.01 comments - on average. Most of them (85.2%) submitted the final exam. They were named the *Social Engagers*.

Characterization of the Participants in Each Cluster

The participants in each cluster were characterized in regard to demographics, course participation, and course achievement. The ANOVA tests showed statistically significant differences between the clusters in regard to the interval variables ($ps < 0.001$), and the chi-squared tests showed statistically significant dependence between the clusters and the categorical variables ($ps < 0.001$). Table 3 presents the characterization of the participants in each cluster. The table displays descriptive statistics of the variables that were used for characterizing the clusters, and the results of the ANOVA and the chi-squared tests. As can be seen, the participants in most of the clusters consisted of a majority of females, ranging between 52.3% and 76.1%, except for the *Offline Engagers* and the *Downloaders*, who consisted of a majority of males (56.9% and 65.8%, correspondingly). The *Offline Engagers, the Tasters and* the *Downloaders* were the youngest participants, ranging between 37.27 and 38.31 years old on average (with no significant difference between these clusters), whereas the *Moderately Social Engagers,* the *Disengagers,* and the *Social Engagers* were the oldest participants, ranging between 47.5 and 54.66 years old - on average (with no significant difference between these clusters). The differences between the youngest clusters and the oldest clusters were statistically significant ($ps < 0.001$). All clusters consisted of a majority of working people, ranging between 50% and 71.3%. The *Tasters* were the first to leave the course, they last accessed it 20.48 days after it began on average, whereas the four *Engagers* clusters (clusters 4 to 7) were the last who accessed the course, between 8 and 15.93 days after the course ended - on average. Only the

10

participants in the *Engagers* types achieved a certificate, ranging between 62% and 78.5% of the participants per cluster. Among them, the *Social Engagers* achieved the highest grades on average in the quizzes (95.52%) and in the exam (86.33%), whereas the *Online Engagers* achieved the lowest average grade in the quizzes (92.6%), and the *Offline Engagers* achieved the lowest grade on the exam (75.3%).

Table 3

Additional Characterization of the Participants in Each Cluster

	Population	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	
name		*Tasters*	*Down-loaders*	*Dis-engagers*	*Offline Engagers*	*Online Engagers*	*Moderately Social Engagers*	*Social Engagers*	
number of participants (percentage in the population)	21,889 (100%)	14,186 (64.8%)	1,857 (8.5%)	2,507 (11.5%)	778 (3.6%)	1,627 (7.4%)	799 (3.7%)	135 (0.6%)	
	Average (standard deviation)								**F (effect size)**
last course access (days from course beginning)	33.37 (26.95)	20.48 (21.68)	56.25 (17.24)	46.18 (18.91)	65.45 (10.82)	66.46 (9.63)	63.76 (17.36)	71.93 (15.42)	$F_{(6, 21,882)}$ = 3067.17** (0.45)
average quiz grade	87.06% (12.76%)	81.94% (13.92%)	85.34% (12.44%)	85.15% (12.11%)	93.31% (9.69%)	92.86% (8.59%)	92.98% (9.06%)	95.52% (8.34%)	$F_{(6, 8,485)}$ = 249.01** (0.15)
exam grade	80.00% (12.42%)	-	-	-	75.30% (14.78%)	80.79% (10.76%)	83.00% (11.86%)	86.33% (9.53%)	$F_{(3,3,139)}$ = 65.13** (0.05)
age	40.27 (16.35)	37.37 (15.55)	38.31 (14.13)	47.64 (17.42)	37.27 (17.51)	42.93 (15.83)	47.50 (15.04)	54.66 (12.04)	$F_{(6,3,547)}$ = 45.04** (0.07)
	Mode (frequency)								**X^2 (effect size)**
gender	Female (53.7%)	Female (56.0%)	Male (65.8%)	Female (61.2%)	Male (56.9%)	Female (52.3%)	Female (57.8%)	Female (76.1%)	$X^2_{(6)}$ = 88.97** (0.15)
employment status	Working (62.6%)	Working (63.6%)	Working (71.3%)	Working (58.0%)	Working (63.8%)	Working (61.6%)	Working (54.8%)	Working (50.0%)	$X^2_{(12)}$ = 47.15** (0.11)
achieved certificate	No (89.4%)	No (100%)	No (100%)	No (100%)	Yes (62.0%)	Yes (75.0%)	Yes (63.8%)	Yes (78.5%)	$X^2_{(6)}$ = 14,548.29** (0.81)

Discussion and Conclusion

MOOCs provide a flexible learning environment, which enables the learner to choose the suitable learning pathway according to his own motivations and needs. This was clearly demonstrated in this study, which identified seven types of participant behavior in a MOOC. The *Tasters* and the *Downloaders* exhibited low levels of engagement in the course and constituted 73.3% of course participants. The *Disengagers* were moderately engaged in the course and constituted 11.5% of course participants. The *Online Engagers*, the *Offline Engagers*, the *Moderately Social Engagers,* and the *Social Engagers* exhibited high levels of engagement in the course and constituted 15.3% of course participants.

The *Tasters* sampled a few learning resources and dropped out approximately three weeks after the course began, on average. This type of participation resembles types identified by Kizilcec et al. (2013) and Ferguson and Clow (2015), who referred to *sampling* participants, and Ho et al. (2014) who referred to *only viewed* participants. The existence of this group may, at least partially, be explained by the fact that registering to a MOOC is fast and easy, and entails no monetary costs, thus a major group of people may sign up for the course without the intention of finishing it (Fischer, 2014; Ho et al., 2014; Mustafaraj, 2014; Onah et al., 2014).

The *Downloaders* engaged in the course by downloading the vast majority of video lectures. They last entered the course close to its end, which may be due to the fact that they focused on downloading the videos that were uploaded on a weekly basis. Some possible reasons for downloading MOOC content by students may be: in order to finish the course on their own time (Belanger & Thornton, 2013; Khalil & Ebner, 2014), due to poor Internet access, culture, or preferences (Seaton et al., 2014). With the absence of data regarding their offline activity, it is unknown whether they watched the videos offline or perhaps stored them for future engagement with the content.

The *Tasters* and the *Downloaders* constituted the majority of course participants. They were among the youngest participants on average and among the groups with the highest percentage of working participants. It is hypothesized that younger participants may be busier in career, having less available time for engaging with the course, but this requires further investigation. These participants used the standard MOOC format, but perhaps could be better served by other formats, such as: shorter courses or a centralized location for downloading content. More qualitative data is required in order to better understand their needs.

The *Disengagers* watched some of the video lectures and submitted some quizzes. They disengaged from the course around 10 days before it ended, on average, without submitting the exam. This type of participation resembles Kizilcec et al.'s (2013) *Auditing* or *Disengaging* participants, and Ho et al.'s (2014) participants who *only viewed* or *only explored* the course. These participants may have disengaged from the course since they met their learning objectives, or for other reasons, such as losing interest or motivation, lack of time, course workload and so on (Kizilcec et al., 2013; Onah et al. 2014; Padilla Rodriguez, Bird, & Conole, 2015). It is essential to identify the reasons for participants' disengagement from MOOC and to develop appropriate interventions for supporting participants who wish to stay engaged but fail to do so.

The four *Engagers* types demonstrated high levels of engagement in the course by using the video lectures and the assessments thoroughly. These types may resemble Kizilcec et al.'s (2013) and Ferguson and Clow's (2015) *completing* participants, as well as Ho et al.'s (2014) *only explored* or *certified* participants. However, the current classification further distinguishes between them by their videos and discussion forums usage, indicating that there were different ways for highly engaging with the course (e.g., watching the videos online versus offline, extent of participation in the forums). Most of the *Engagers* achieved a certificate. However, there were significant differences between them in course achievement. The *Social Engagers* achieved the highest average grades in the assessments, which is in line with previous studies that found that students who engage explicitly in the discussion forums are often higher performing than

students who do not (Gillani & Eynon, 2014; Phan, McNeil, & Robin, 2016). The *Social Engagers* was the cluster with the highest percentage of non-working people. It is hence hypothesized that they may have had more time for higher engagement with the course, but this requires further investigation.

It is noteworthy that a significant number of *Engagers*, 1,020 participants, accessed substantial amounts of the course content without achieving a certificate. With reference to the criticism regarding relatively low completion rates in MOOCs, as noted by Kizilcec et al. (2013), under a monolithic view of course completion these participants would have been considered to be just *non-completing*. However, the closer analysis of their behavior reveals that they were highly engaged in the course.

Overall, the analysis of the participants' behaviors across the clusters revealed several types of video usage: watching the videos online, downloading them, and a combination of the two. The different types of video usage may have significant implications on the learning process, as they facilitate different pedagogies. Watching the videos offline enables greater flexibility in learning time and place, whereas watching online provides access to online scaffoldings, such as the in-video questions. In order to support participants who download the videos MOOCs designers should consider the addition of alternative scaffoldings (e.g., download the video questions separately).

Furthermore, regarding the discussion forums, the results of this study indicated that the vast majority of participants mainly viewed the forums to varied degrees, whereas only a small group of participants was active in the discussions. The most active participants, the *Social Engagers*, were the oldest participants on average and most of them achieved a certificate. These findings are consistent with the literature (Breslow et al., 2013; Guo & Reinecke, 2014; Ho et al., 2014; Huang, Dasgupta, Ghosh, Manning, & Sanders, 2014; Kizilcec et al., 2013; Onah et al., 2014). According to the social-constructivism approach, the social context and relationships with others are crucial to a process of negotiating meaning and developing new skills (Anders, 2015). Hence, these findings raise questions regarding the extent to which the social learning potential of MOOCs is realized, as well as regarding the feasibility of holding effective discussions in a course that contains thousands of participants. There is a need to consider the development of social learning mechanisms, which are more adapted for massive courses.

Finally, most of the participants did not use the assessments thoroughly. The low usage level raises the need to examine alternative forms of assessments and practice, which will result in higher engagement. Freire, Martínez-Ortiz, Moreno-Ger, and Fernández-Manjón (2015), for example, suggested the integration of educational games to improve interaction and assessment in MOOCs.

To conclude, MOOCs are characterized by offering high flexibility in learning, which enables different ways of participating in the course according to one's motivations and needs. This study provided further evidence regarding the utilization of this flexibility, by identifying seven different types of participants' behaviors in the course. It should be noted that a significant number of participants were engaged in the course (e.g., by watching videos, submitting the assessments), and may have been contributed from it according to their needs, despite the fact that they did not achieve a certificate. These results further support the literature which claimed that MOOCs' impact should not be evaluated solely based on certification rates (Ho et al., 2014), but rather based on learning behaviors. Understanding the

participants' behaviors and characteristics will enable to better adapt the courses to different learners' needs, thus maximizing the MOOCs impact in delivering lifelong learning on a large-scale.

Future Study and Limitations

It should be noted that the study examined one MOOC. More research is required to examine if the types of participation identified in this study are found in other MOOCs from varied disciplines, course structures, target audiences, platforms, and times. In addition, future research can further deepen the analysis by adding variables that describe learning process evolution, such as: changes in videos views and downloads, changes in active and passive participation in the forums over the course and so on. Finally, the study was conducted using educational data mining, which is an objective research methodology that helps to ground research in real data (Siemens et al., 2011). However, it lacks direct contact with the research population. Future research should combine other research approaches such as surveys and interviews, in order to shed more light on participants' motivations, goals and needs.

References

Anders, A. (2015). Theories and applications of massive online open courses (MOOCs): The case for hybrid design. *The International Review of Research in Open and Distributed Learning, 16*(6). doi: http://dx.doi.org/10.19173/irrodl.v16i6.2185

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In K. Sawyer (Ed.) *Cambridge handbook of the learning sciences* (2nd ed., pp. 253-274). New York, NY: Cambridge University Press.

Belanger, Y., & Thornton, J. (2013). Bioelectricity: A quantitative approach, Duke University's first MOOC [Electronic document]. Retrieved from http://dukespace.lib.duke.edu/dspace/handle/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.

Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z., & Wong, F.M.F. (2014). Learning about social learning in MOOCs: From statistical analysis to generative model. *IEEE Transactions on Learning Technologies, 7*(4), 346-359. doi:10.1109/TLT.2014.2337900

Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in* Education, *3*. doi: http://doi.org/10.5334/2012-18

Diwanji, P., Simon, B. P., Marki, M., Korkut, S., & Dornberger, R. (2014). Success factors of online learning videos. In *International Conference on Interactive Mobile Communication Technologies and Learning* (pp. 125-132). IEEE. doi: 10.1109/IMCTL.2014.7011119

Ferguson, R., & Clow, D. (2015). Examining engagement: Analysing learner subpopulations in massive open online courses (MOOCs). In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 51-58). ACM. doi: 10.1145/2723576.2723606

Fischer, G. (2014). Beyond hype and underestimation: Identifying research challenges for the future of MOOCs. *Distance Education, 35*(2), 149-158. doi: http://dx.doi.org/10.1080/01587919.2014.920752

Freire, M., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2015). Requirements for educational games in MOOCs. In *IEEE Global Engineering Education Conference (EDUCON)* (pp. 993-997). IEEE. doi: 10.1109/EDUCON.2015.7096094

Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education, 23*, 18-26. doi: 10.1016/j.iheduc.2014.05.004

Glance, D. G., Forsey, M., & Riley, M. (2013). The pedagogical foundations of massive open online courses. *First Monday, 18*(5). doi: http://dx.doi.org/10.5210/fm.v18i5.4350

Guo, P.J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the First ACM Conference on Learning @ Scale* (pp. 41-50). ACM. doi: 10.1145/2556325.2566239

Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. In *Proceedings of the First ACM Conference on Learning @ Scale* (pp. 21-30). ACM. doi: http://dx.doi.org/10.1145/2556325.2566247

Halawa, S., Greene, D., & Mitchell, J. (2014). Dropout prediction in MOOCs using learner activity features. In *Proceedings of the European MOOC Stakeholder Summit* (pp. 58-65). Retrieved from https://www.emoocs2014.eu/sites/default/files/Proceedings-Moocs-Summit-2014.pdf

Ho, A. D., Reich, J., Nesterko, S., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014, January 21). *HarvardX and MITx: The first year of open online courses in HarvardX and MITx*. HarvardX Working Paper No. 1. doi: http://dx.doi.org/10.2139/ssrn.2381263

Huang, J., Dasgupta, A., Ghosh, A., Manning, J., & Sanders, M. (2014). Superposter behavior in mooc forums. In *Proceedings of the First ACM Conference on Learning @ Scale* (pp. 117–126). ACM. doi: 10.1145/2556325.2566249

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis* (Vol. 344). Hoboken, NJ: John Wiley & Sons.

Khalil, H., & Ebner, M. (2014). MOOCs completion rates and possible methods to improve retention - A literature review. In *Proceedings of the W*orld Conference on Educational Media and Technology (Vol. 2014, No. 1, pp. 1305-1313). Tampere, Finland: Association for the Advancement of Computing in Education (AACE). Retrieved from https://www.learntechlib.org/p/147656/

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 170–179). ACM. doi: 10.1145/2460296.2460330

Koller, D., Ng, A., Chuong, D., & Chen, Z. (2013, June 3). Retention and intention in Massive Open Online Courses. *Educause Review*, *48*(3), 62–63.

Li, Q. (2004). Knowledge building community: Keys for using online forums. *TechTrends, 48*(4), 24–29. doi: 10.1007/BF02763441

Li, N., Kidzi´nski, L., Jermann, P., & Dillenbourg, P. (2015). MOOC video interaction patterns: What do they tell us? In *Proceedings of the 10th European Conference on Technology Enhanced Learning* (pp. 197-210). Cham, Switzerland: Springer. doi: 10.1007/978-3-319-24258-3_15

Mooney, C. Z., & Duval, R. D. (1993). Bootstrapping: A nonparametric approach to statistical inference. In M. S. Lewis Beck (Ed.), *Quantitative applications in the social sciences* (Vol. 95, pp. 1–72). Thousand Oaks, CA: Sage.

Mustafaraj, E. (2014). What does enrollment in a MOOC mean?. In *Proceedings of the First ACM conference on Learning @ Scale* (pp. 203-204). ACM. doi:10.1145/2556325.2567882

Mustafaraj, E., & Bu, J. (2015). The visible and invisible in a MOOC discussion forum. In *Proceedings of the Second ACM Conference on Learning @ Scale* (pp. 351-354). ACM. doi: 10.1145/2724660.2728691

Nkuyubwatsi, B. (2013). Evaluation of massive open online courses (MOOCs) from the learner's perspective. In *European Conference on e-Learning* (p. 340). Kidmore End, UK: Academic Conferences International Limited. Retrieved from http://hdl.handle.net/2381/28553

Norusis, M. J. (2012). *IBM SPSS statistics 19 statistical procedures companion*. Upper Saddle River, NJ: Prentice Hall.

Onah, D. F. O., Sinclair, J. E., & Boyatt, R. (2014). Exploring the use of MOOC discussion forums. In *Proceedings of London International Conference on Education* (pp. 1-4). LICE. doi: 10.13140/RG.2.1.3319.5042

Padilla Rodríguez, B.C., Bird, T., & Conole, G. (2015). Evaluation of Massive Open Online Courses (MOOCs): A case study. In *Proceedings of the Global Conference on Learning and Technology*

(Vol. 2015, No. 1, pp. 527-535). Berlin, Germany: Association for the Advancement of Computing in Education (AACE). Retrieved from https://www.learntechlib.org/p/150900/

Phan, T., McNeil, S. G., & Robin, B. R. (2016). Students' patterns of engagement and course performance in a Massive Open Online Course. *Computers & Education, 95,* 36-44. doi: 10.1016/j.compedu.2015.11.015

Ramesh, A., Goldwasser, D., Huang, B., Daume III, H., & Getoor, L. (2014). Learning latent engagement patterns of students in online courses. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* (pp. 1272-1278). Palo Alto, CA: AAAI Press. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8571

Rodriguez, C. O. (2012). MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. *European Journal of Open, Distance and E-learning, 15*(2). Retrieved June 12, 2016 from http://www.eurodl.org/index.php?p=archives&year=2012&halfyear=2&article=516

Sandeen, C. (2013). Assessment's place in the new MOOC world. *Research & Practice in Assessment, 8*(1), 5-12.

Seaton, D. T., Nesterko, S., Mullaney, T., Reich, J., Ho, A., & Chuang, I. (2014). Characterizing video use in the catalogue of MITx MOOCs. In *Proceedings of the European MOOC Stakeholder Summit* (pp. 140-146). Retrieved from https://www.emoocs2014.eu/sites/default/files/Proceedings-Moocs-Summit-2014.pdf

Sharma, K., Jermann, P., & Dillenbourg, P. (2014). "With-me-ness": A gaze-measure for students' attention in MOOCs. In *Proceedings of the International Conference of the Learning Sciences* (Vol. 2014, No. 2, pp. 1017-1021). Boulder, CO: International Society of the Learning Sciences. Retrieved from http://infoscience.epfl.ch/record/201918

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., … Baker, R. S. J. D. (2011, July 28). *Open learning analytics: An integrated & modularized platform.* Retrieved from http://solaresearch.org/OpenLearningAnalytics.pdf

Suen, H. K. (2014). Peer assessment for massive open online courses (MOOCs). *The International Review of Research in Open and Distance Learning, 15*(3), 313-327. doi: http://dx.doi.org/10.19173/irrodl.v15i3.1680

United Nations. (1948). The universal declaration of human rights. Retrieved from http://www.unesco.org/education/information/50y/nfsunesco/doc/hum-rights.htm

UNESCO. (2012). Paris OER declaration. Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/WPFD2009/English_Declaration.html

Wang, Y., & Baker, R. (2015). Content or platform: Why do students complete MOOCs? *MERLOT Journal of Online Learning and Teaching, 11*(1), 17-30.

Wong, J. S., Pursel, B., Divinsky, A., & Jansen, B. J. (2015). An analysis of MOOC discussion forum interactions from the most active users. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 452-457). Cham, Switzerland: Springer. doi: 10.1007/978-3-319-16268-3_58

Woodgate, A., Macleod, H., Scott, A. M., & Haywood, J. (2015). Differences in online study behavior between sub-populations of MOOC learners. *Education XX1, 18*(2), 147-163. doi: http://dx.doi.org/10.5944/educxx1.13461