

February – 2019

Understanding Participant's Behaviour in Massively Open Online Courses

Bruno Poellhuber, Normand Roy, and Ibthihel Bouchoucha
Université de Montréal

Abstract

As the offer of Massive Open Online Courses (MOOCs) continues to grow around the world, a great deal of MOOC research has focused on their low success rates and used indicators that might be more appropriate for traditional degree-seeking students than for MOOC learners, who, because of the openness of MOOCs, represent a more diverse clientele who exhibit different characteristics and behaviours. In this study, conducted in a French MOOC that is part of the EDUlib initiative, we systematically classified MOOC user profiles based on their behaviour in the open-source learning management system (LMS) – in this case, Sakai – and studied their survival in the MOOC. After formatting the logs in ordinal variables in order to reflect a continuum of participation central to the behavioural engagement concept (Fredricks, Blumenfeld, & Paris, 2004), we incrementally executed a two-step cluster analysis procedure that led us to identify five different user profiles, after having manually excluded Ghots : Browser, Self-Assessor, Serious Reader, Active-Independent, and Active-Social. These five profiles differed both qualitatively and quantitatively on the continuum of engagement, and a significant proportion of the less active profiles did not drop out of the MOOC. Our results confirm the importance of social behaviours, as in recent typologies, but also point out a new Self-Assessor category. The implications of these profiles for MOOC design are discussed.

Keywords: MOOC, participant profiles, cluster analysis, survival analysis, behavioural engagement

Introduction

Massive Open Online Courses (MOOCs), a recent form of distance courses essentially based on short videos and quizzes, have received a great deal of public attention lately, and the debate about the future of MOOCs and their impact on education continues. While some argue that MOOCs are now in the disillusionment phase of the Gartner Hype Cycle (White, 2014), i.e., the point in time when “interest wanes as experiments and implementations fail to deliver” (Gartner, 2019), data from the Open Education Scoreboard (European Commission, 2015), a European site that monitors MOOC offerings around the world, indicate that global MOOC offerings have been picking up since May 2015, corresponding, rather, to the “plateau of productivity” in the Gartner Cycle, referring to a certain maturity of the innovation. In fact, as demonstrated by Peters (2018), the MOOC offering is now much more diverse, with small MOOCs, private MOOCs and credential MOOCs, suggesting that students who register for MOOCs have diverse needs. While MOOCs contribute to wider accessibility of higher education and have a place in the distance education offering, many questions remain on persistence rates and on who registers, what their profiles are, how they use MOOCs and their motivation to do so.

A body of scientific literature on MOOC research is now emerging, with many journals publishing special issues on the topic. While MOOCs have drawn substantial attention by reaching millions of students, many of their biggest drawbacks relate to the near-catastrophic completion rate reported by many (Jordan, 2013, 2014). This is unsurprising, given that poor completion rates have been the subject of many articles on distance education (Simpson, 2003), and that MOOCs are a particular instance of distance courses. Although completion and success rates have been the golden standards for assessing the quality of both face-to-face and online courses, we question whether such standards are representative of the main objectives of MOOCs. Moreover, do they correspond to the expectations of MOOC learners?

The first O in the MOOC acronym stands for Open, in the original sense of accessibility and absence of barriers to instruction, the meaning also conveyed by the designation of Open Universities (Anderson, 2013). Registering for a for-credit course, whether distance or in-person, at any postsecondary institution requires a considerable investment of time and often money: applying to a program, supplying documentation of prior learning (e.g., transcripts and diplomas), waiting for an acceptance which may be conditional, paying tuition fees, choosing and registering for courses, doing course readings and activities, and taking exams. Barriers are present at many of these steps. Simply registering in a course demonstrates a high level of commitment. In contrast, registering for a MOOC is a low-barrier process that often requires no more than entering your name, email, and password, at first. Furthermore, the flexibility offered by MOOCs in terms of time and logistics is the main reason learners opt for them (Roy, Poellhuber & Bouchoucha, 2015), followed by the fact that MOOCs are free at registration. However, this is nowadays debatable since most MOOCs platforms offer a paying form of authenticated certificate or monthly subscription, and even integration in for credits programs that are quite expensive, but still quite less than the same programs offered on campus.

The argument we present in this study is that the openness and accessibility of MOOCs attract or at least used to attract a clientele who may, compared to traditional degree-seeking students who register in for

credit courses in universities, have a far more diverse profile in many dimensions: sex, age, occupation, country of residence, motivations, reasons for taking the MOOC, etc. (Poellhuber, Roy & Levasseur, 2017). The openness concept has evolved significantly and taken on new meanings. In the context of Open Distance Education, openness in Europe has expanded to include flexibility and adaptation to learners' needs, as in the French definition of Open and Distance Education ("formations ouvertes et à distance"; Carré, 2001). While MOOCs have been criticized for being not so free or open (Anderson, 2013), they still remain an important vector of openness in the original sense of removing barriers. With more than 100 million learners (Shah, 2018), they are now a trend in the distance education movement and they make eLearning available to new types of learners. Analysis of the age curve of the learners in the EDUlib initiative indicates that learners are older than regular HEC (Hautes études commerciales) Montreal (a business school) students and that fewer than 1% are actually registered students. MOOC designers and authors tend to view MOOC learners as students, anticipating that their behaviour will conform to what is expected from students, but this is far from certain. Much of the research in MOOC literature use implicitly success rates as the golden standard of MOOC quality, but this assumption may be false. Indeed, a large number of MOOC registrants never log in once the MOOC is opened (Hill, 2013). Furthermore, emerging research on MOOC participant typologies shows that the behaviour of a large number of MOOC learners differs substantially from what is expected from more traditional students.

MOOC Participant Typologies

Work on MOOC participant typologies began with manual and rational classifications (Hill, 2013; Ho et al., 2014; Milligan, 2012) and gradually evolved toward a more robust and systematic classification schemes that rely on cluster analysis. From very small studies to large-scale studies, a variety of attempts have been made to classify student engagement behaviour patterns in MOOCs, as synthesized in Table 1.

Table 1

Literature Review on MOOC Participant Typologies in Chronological Order

Authors	Sample	Type of analysis	Variables	Typologies
Milligan (2012) and Hill (2013)	1 connectivist MOOC (Change 11)	Rational analysis	Questionnaire (self-regulated learning, Motivated Strategies for Learning Questionnaire) and interview data	(1) No-shows, (2) Lurkers, (3) Passive Participants, (4) Active Participants
Kizilcec et al. (2013)	3 Stanford University MOOCs	Cluster analysis (k-mean)	On-time or late assessments, video watching, quizzes	(1) Completing, (2) Auditing, (3) Disengaging, (4) Sampling
Ho et al. (2014)	17 edX MOOCs	Rational analysis	Registrations, certifications, resource viewing	(1) Only Registered, (2) Only Viewed, (3) Only Explored, (4) Certified
Ferguson & Clow (2015)	4 Future Learn MOOCs	Cluster analysis (k-mean)	Activities (visits or on-time or late comments) on content and assessments	(1) Samplers, (2) Strong Starters, (3) Returners, (4) Mid-way Dropouts, (5) Nearly There, (6) Late Completers, (7) Keen Completers
Tseng et al. (2016)	5 Taiwan Yuan Ze University MOOCs	Cluster analysis	Logins, video viewing, assignment submission	(1) Bystanders, (2) Passive Learners, (3) Active Learners
Kovanovic et al. (2016)	28 offerings of 11 Coursera MOOCs	Cluster analysis (k-mean)	Activities in videos, assignments, quizzes, wikis, discussion forums	(1) Enrollees, (2) Low Engagement, (3) Videos, (4) Videos & Quizzes, (5) Social
Khalil & Ebner (2017)	1 Gratz University MOOC	Cluster analysis (k-mean)	Reading and writing frequencies, videos watched, quiz attempts	(1) Dropouts, (2) Perfect Students, (3) Gaming the System, (4) Socials
Kahan, Soffer & Nachmias (2017)	1 Coursera MOOC	Cluster analysis	Video, discussion forums, quizzes and assessment activities, grades, certification, demographics	(1) Tasters, (2) Downloaders, (3) Disengagers (4) Offline Engagers, (5) Online Engagers, (6) Moderately Social Engagers, (7) Social Engagers

Milligan's work inspired Hill (2013), who reinterpreted his typology outside the original conceptual framework, but in a very user-friendly way, and developed a well-known image (see Figure 1) of a popular MOOC typology.

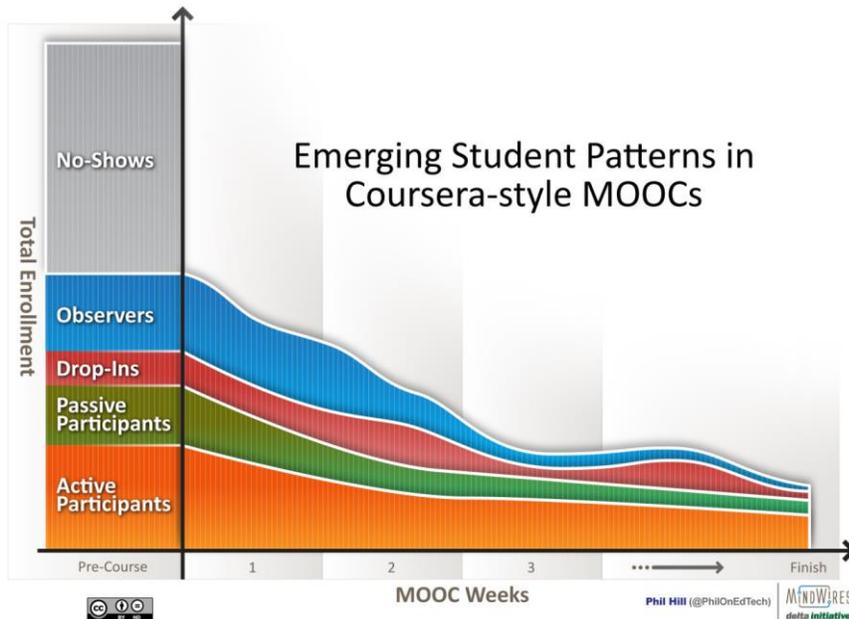


Figure 1. Hill's MOOC participant typology. From "Emerging patterns in MOOCs: A graphical view." P. Hill, *E-literate*, 2013. (mfeldstein.com/emerging_student_patterns_in_moocs_graphical_view/) CC/BY/ND

The first typologies used similar variables to understand MOOCs, mostly based on platform connection and the consultation of different types of resources, from lectures and documents (Milligan, 2012) to videos and assessments (Kizilcec, Piech, & Schneider, 2013). Although the terminology varies among the studies, they are mostly self-explanatory.

Most of the early typologies (pre-2016) center on the assumption that all learners should turn in assignments and attend the MOOC until the end. But can we really consider the "Auditing" and "Disengaging" to be disengaged? (Kahan, Soffer, & Nachmias, 2017). The strategy used to create these typologies relies on simple indicators or ones that are aligned with a classical view of what is expected of students, with assignment submissions used as a key indicator in the classification schemes. They suggest classification schemes that are both quantitative (focused on the level of engagement) and qualitative (focused on the type of engagement, as in Milligan, 2012). If we want to meaningfully classify students on both of these aspects, we need to ask what the most appropriate classification scheme is. Kovanović et al. (2016) shed light on an important category: the registered, who only register but do not log in. This represents a known fact on MOOCs and explains a large proportion of the dropouts (Jordan, 2014). Gradually, efforts have been made to integrate a wider level of activities in the classification variables, such as quiz attempts (Khalil & Ebner, 2017) and activity in discussion forums (Kovanović et al., 2016). Kahan, Soffer, and Nachmias (2017) went further, taking a novel approach in which activities in all resource categories were the subject of a cluster analysis, without preference to activities tied to assessment or completion.

In our study, the objective was to create MOOC participant profiles with an appropriate and systematic classification methodology based on participant interactions with MOOC resources, drawing on a conceptual framework of behavioural engagement. Rather than relying on pre-existing conceptual categories, we propose a sound and systematic strategy to characterize the different MOOC participant profiles, based on their pattern of engagement or interaction with any type of MOOC resources. We also wished to examine the dropout behaviours of these different profiles.

Conceptual Framework

The construct of engagement enjoys a wide variety of definitions, in games as well as in education, ranging from a broad acceptance of the term (e.g., any type of interaction) to precise types of engagement such as “gaming the system” (Baker et al., 2004). In educational research, engagement is closely related to motivation, especially in socio-cognitive expectancy-value motivational models (Eccles & Wigfield, 2002; Pintrich, 2003), inspired by Bandura’s triadic interaction model (1986, 1997). In these motivational models, engagement is sometimes seen as equivalent to motivation (Clark, 1999; Viau, 2003), while others (Pintrich, 2003) make a subtle distinction between engagement and motivation, with engagement viewed as a behavioural consequence of a favourable combination of expectancy beliefs and perceived value. Expectancy beliefs refer to the learner beliefs to succeed in a given task while value perception is the overall assessment of the utility, importance, or interest of the task (Pintrich, 2003).

Prominent authors on the construct of engagement distinguish among three aspects of engagement: behavioural engagement, cognitive engagement, and emotional engagement (Fredricks, Blumenfeld & Paris, 2004; Linnenbrink & Pintrich, 2003). Behavioural engagement refers to the observable indicators of participation, such as paying attention in class, avoiding distraction, and so on. Behavioural engagement can be conceptualized as a continuum: from the absence of disturbing behaviours, to respect for class rules, to active participation in class discussions and even participation in extracurricular activities. Emotional engagement refers to the display of positive or negative emotions involved in a learning context or the learning process itself. Cognitive engagement is rooted in the idea of mental effort and investment. This effort can be measured quantitatively (e.g., the intensity and duration of the cognitive resources invested in the task), but also qualitatively (e.g. the degree of appropriateness, sophistication, and efficiency of the cognitive investment) (Molinari et al., 2016). Cognitive engagement relates to cognitive strategies and to metacognition, which refers to the choice, monitoring, and regulation of these strategies. Pintrich, Smith, Garcia & McKeachie (1991) differentiate between cognitive strategies (repetition, elaboration, organization) and metacognitive strategies (critical thinking, regulation, resource management).

While behavioural engagement indicators are often also indicators of the cognitive investment that defines cognitive engagement, they are not always reliable. A student can look at the teacher while letting their mind wander. In this particular study, we submit that the learners’ actions that are logged in the learning management system (LMS) can be considered behavioural engagement indicators, or

manifestations of it. They are no more nor less reliable than behavioural engagement indicators that can be used in a face-to-face class setting. They focus on learner-content interactions, the most important type of interaction in distance education (Bernard & Amundsen, 2008). We submit the idea that if formatted properly, these logs reflect a continuum of participation, i.e., behavioural engagement. While they may be related to the level of cognitive investment (cognitive engagement), they may also be a poor indicator of it. For example, it is difficult to tell whether viewing a particular video five or six times is an indicator of cognitive engagement or of something else, such as connection problems or distractions at home. Whether we are observing merely behavioural engagement or cognitive engagement in traces is a matter of debate and interpretation; thus, interpreting traces as “observable behaviour” corresponding to behavioural engagement seems more appropriate. Furthermore, theory predicts links between behavioural engagement and cognitive engagement. To analyse computer traces from the perspective of behavioural engagement is, as observed in Table 2, a sound practice.

In summary, behavioural engagement pertains to visible manifestations of engagement and exists along a continuum. The more visible participation is, the more behaviourally engaged the participant is considered to be. Thus, we argue that in the MOOC context, traces of participant activities in the environment can be considered indicators of behavioural engagement.

Methods

This quantitative study relied on big data and learning analytics methodology based on traces of participant behaviour in the MOOC learning management system (LMS) and on classical statistical categorization techniques such as cluster analysis, multiple correspondence analysis and principal component analysis. The data pertaining to participant behaviour in the Sakai LMS were extracted, cleansed, and formatted using a combination of procedures in Stata and SPSS.

Context

The data used in this study come from a French-language MOOC (*Economic Problems and Policies* course), offered as part of the EDUlib initiative through the Sakai LMS, which was offered in the Spring 2013 semester. Four Thousand Eight Hundred and Fifty people registered for the course. The MOOC had six one-week modules. The course material for each week included, in general, six videos, six PDFs files corresponding to the lecture slides, and at least one required and one complementary reading. In this course, each week corresponded to a module that offered six short (10-12 min.) video lectures, the slides used in these lectures, a required reading, some suggested readings, a formative quiz, and a discussion board. Two discussion forums were offered (module content and general discussion). At the end of each module there was a summative test, and the students had one week to complete it. The same structure continued all through the MOOC.

Data

Learners' traces were first extracted in three different files: resources (document, video, etc.), events (discussion, communication, etc.), and visits (logging in the course webpage). We collected date and time, as well as counts, on three types of behaviours: the specific resources learners accessed (forums, tests, quizzes, resource, visits), events (new thread in forum, read thread, response thread, test and quiz assessment submit, test and quiz published assessment revise, read, site visit), and visits to the site. We then aggregated dates and times of traces by creating eight variables associated with each week of the six-week course, the final exam period, and the post-exam period.

Each activity carried out by the participant during each period was counted. Then, after data cleansing and integrity checking, the data were aggregated based on the activities and resources associated with each course module and week. For each week, we assigned variables, based on how many of the videos, PDF lecture slides, and required or optional readings learners accessed. We assigned a value of 0 if the person had not consulted any of the course material for the week, 1 if they consulted some, and 2 if they consulted all of the material.

Of the 4,850 registrants, 1,691 logged in at least once after Week 1 (this being our criterion to define them as learners). Of these, 185 learners dropped out of the course during the second week (their last login was in Week 2), and 323 others did no activities in Week 2, but did activities later in the course. These were considered to be "ghosts" and removed from our analysis because they had little or no trace of activity in Week 2 and to make the classification more precise. Of the 1,691 learners, 30% were female, with an overall median age of 35.8 years old. The learners were mostly full-time workers (65.2%) with a higher education diploma (post-graduate or PhD) (54.2%). Learners were mostly French-speaking from the province of Quebec in Canada (44%), but they also came from a variety of other French-speaking countries.

Data and Variables

At the end of Week 1, the Module 1 exam became available for a full week, and the material for Module 2 was made available at the same time (we thus have differentiated between Course week period and Examination period).

Prior to analyzing the data, we transformed it to align with the behavioural engagement concept and its central idea of a continuum of participation. We defined participation as any type of activity in the MOOC with any type of resource. Four types of resources were available: video lectures, quizzes and tests, the discussion forums, and PDF files consisting of the required reading (usually a book chapter), optional readings, or the lecture slides. For each of these resources, we created ordinal variables representing a continuum of participation. For example, instead of using the total weight of forum interactions, we differentiated among types of forum participation: no participation, reading, responding to a question asked by someone else, and asking a question, which we considered the highest level of engagement because it creates a new thread visible by all. While this choice entails some information loss, it is much more consistent with the underlying conceptual framework (behavioural engagement) than the weight of

forum interactions. Total time spent on activities would have been an obvious measure to include, but it was not available. Furthermore, the LMS had no automatic logout after a period of inactivity, a characteristic that would make that measure unreliable. We also introduce a categorization of the activities based on the timeframe to put into evidence if activities were completed “on time,” i.e. during the week dedicated to those activities, or “late,” in subsequent weeks. We conclude that this refinement added little to the model, and we ultimately chose seven different variables for each week’s activities, as displayed in Table 2. It is to be noted that there were no optional reading for the Module 2, for which we report the analysis in this paper.

Table 2

Description of the Variables Used for Cluster Analysis

Variable	Description
PDF lecture slides viewed (module 2)	Proportion of PDF files (lecture slides) viewed in the second week of the course.
Streamed videos viewed (module 2)	Proportion of the six videos viewed in streaming mode in the second week of the course (note: a link was made available for students who wanted to download a complete version of the videos, but activities on these links were not recorded).
Required readings downloaded (module 2)	Proportion of required reading downloaded.
Test & quiz submitted (module 2)	Number of tests and quizzes completed and submitted.
Test & quiz attempted, but un-submitted (module 2)	Number of tests and quizzes attempted (incomplete and un-submitted).
Forum questions asked (module 2)	Number of interactions in discussion forums to ask questions during the course week and the test week.
Forum answers contributed (module 2)	Number of interactions in discussion forums to answer questions during the course week and the test week.

Data Analysis

We used cluster analysis to place each participant in a group sharing similar characteristics. We used the traces from the beginning of Week 2 to the end of Week 3, corresponding to Module 2, which was the first module for which we had all the traces. In preliminary analysis, we distinguished three phases: the “on-time” activities carried out during the course week, Week 2 activities carried out later than Week 2, and activities carried out during the test week (which extended one week more). Exploratory analysis of all the variables for these three periods revealed that only the behaviour variables during the course week period helped distinguish among profiles. We therefore used only these variables for our final cluster procedure.

Two-step cluster analysis is a technique used to create homogeneous subsample groupings by identifying similar learners while minimize within-group variation and maximize between-group variation (Garson, 2014). First, the two-step procedure allows us to work with a continuous and categorical variable. Second, the method is effective with large datasets (Tuffery, 2011; Garson, 2014). Our main objective was to systematically divide the large number of learners into categories or patterns that would make sense from a behavioural engagement perspective. Performing this analysis on our dataset allowed us to group users who shared common behaviours on the MOOC platform, with obvious distinctions among the groups. We introduced the seven variables described earlier in a two-step cluster analysis using SPSS 21. Automatic sampling yielded three main categories, but to determine the optimal solution, we tested four, five, six, seven, and eight categories, comparing the quality of the different models and the meanings of the classes produced. We sought a classification model in which the profiles would be qualitatively and meaningfully different from a behavioural engagement perspective, while preserving the quality of the classification solution. This relates to the meaningfulness criterion in Garson's (2014) three ways to assess cluster validity: criterion (or variable) validity, distance (or proximities), and meaningfulness. The SPSS analysis showed that in the four-, five-, and six-group solutions, all included variables (criteria) provided useful information in all models, with a relative importance between 0.1 and 1.0. Based on cluster proximities, the results also indicated that all models presented a good solution. Therefore, while validity and distance did not provide enough evidence to select between the four-and six-factor solutions, the five-factor solution gave us a clearer qualitative portrait of each cluster by introducing forum activity as a variable separating the two most active profiles. We repeated the cluster analysis process for activities later in the MOOC (in Module 4) and the results (not presented here) replicated the clusters found in the present analysis.

While a great deal of MOOC research focuses on successful completion, the description of how long the learners of each profile remain active is particularly relevant to understanding these profiles. Survival analysis studies the time before some "death" event occurs (Tabachnick & Fidell, 2007) — here, dropping out or disappearing from the MOOC. We used it to estimate a survival function in relation to dropping out, which we defined as the week in which the participant last logged in. Because our dropout variable did not have a normal distribution and our variables were not perfectly continuous, we ran a discrete survival analysis, which gave the probability of dropping out at each week. Using function $F(x) = P(X \leq t)$, we computed the mortality ratio, i.e., the probability or risk that learners "surviving" a particular week will "die" the following week.

Results

The final results of the two-step cluster analysis procedure differentiated among five different groups that constitute a continuum in terms of the level of behavioural engagement and that also differ qualitatively. We do end up with seven groups because we did not take into consideration the participants who registered and never logged in after Week 1. These actually represent the largest group ($n = 3,159$), accounting for 65.1% of everyone who registered. Of the 1,691 learners, which we defined as those who

logged in after Week 1, 508 learners were members of the *Ghost* profile, described below, which was manually excluded from the cluster analysis procedure, and of whom we know little. Five additional profiles were created; *Browser*, *Self-Assessor*, *Serious Reader*, *Active-Independent* and *Active-Social*.

Table 3 presents the “on time” activities of the members of these five additional profiles for the second module of the course, on each modality of the seven variables selected for classification. The second line of Table 3 presents the number of learners in each particular profile and its proportion of the total. In this table, each variable is presented in bold in a row, and each variable modality is presented under the variable. The percentages, which must be read vertically, represent the proportion of the profile members that corresponds to the particular response modality for each variable. For example, if we look at the first variable presented in Table 3, we can see in the fourth row that 25.3% of the *Serious Readers* downloaded no PDF files, 24.9% downloaded some of them and 49.8% downloaded all of them. The column after the % column indicates the results of a column proportion test. This test is applied to see which variable profiles are significantly different from one another. For example, 66.7% of the *Active-Socials* downloaded all of the PDF files associated with Week 3. This differs significantly from the proportion of *Serious Readers* (column C) and *Active-Independents* (column D), of whom respectively 50% and 46% downloaded all files. It also differs from the *Browsers* (column A) and the *Self-Assessors* (column B), but since the proportion is 0% for these, they are not taken into account in the proportion test.

Table 3

Engagement Profile Based on Learners' Behavioural Engagement Variables

	A.	B.	C.	D.	E.	Total	χ^2
	Browser	Self-Assessor	Serious Reader	Active/Independent	Active-Social		
n	271	186	277	350	99	1183	
%	22.9%	15.7%	23.4%	29.6%	8.4%	100.0%	
PDF lecture slides viewed (module 2)							
None	99.3% ^{CDE}	99.5% ^{CDE}	25.3% ^{NS}	37.4% ^{CE}	16.2% ^{NS}	69.7%	671
Some	0.7% ^{NS}	0.5% ^{NS}	24.9% ^{AB}	16.3% ^{AB}	17.2% ^{AB}	8.6%	146
All	a	a	49.8% ^{NS}	46.3% ^{NS}	66.7% ^{AC}	21.6%	366
Total						1183	579.5 ***
Streamed videos viewed (module2)							
None	98.5% ^{CDE}	99.5% ^{CDE}	72.9% ^E	66.3% ^E	46.5% ^{NS}	69.7%	671
Some	1.1% ^{NS}	0.5% ^{NS}	18.1% ^{AB}	19.1% ^{AB}	31.3% ^{AB}	8.6%	146
All	0.4% ^{NS}	a	9.0% ^A	14.6% ^A	22.2% ^{CD}	21.6%	366
Total						1183	213.9 ***
Required readings downloaded (module 2)							
None	100% ^a	100% ^a	41.9% ^E	48.9% ^E	19.2% ^{NS}	75.2%	763
All	a	a	58.1% ^{NS}	51.1% ^{NS}	80.8% ^{CD}	24.8%	420
Total						1183	439.6 ***

Test & quiz submitted (module 2)												
None	100%	^a	42.5%	^E	52.7%	^E	^a	1.0%	^{NS}	42.0%	497	
1 or more		^a	57.5%	^{NS}	47.3%	^{NS}	100%	^a	99.0%	^{B C}	58.0%	686
Total											1183	709.0 ***
Test & quiz attempted, but un-submitted (module 2)												
None	100%	^a		^a	26.4%		^a		^a	29.1%	344	
1 or more		^a	100%	^a	73.6%		100%	^a	100%	^a	70.9%	839
Total											1183	922.3 ***
Forum questions asked (module 2)												
None	100%	^a	99.5%	^E	97.8%	^E	100%	^a	48.5%	^{NS}	95.1%	1125
1 or more		^a	0.5%	^{NS}	2.2%	^{NS}	^a		51.5%	^{B C}	4.9%	58
Total											1183	505.4 ***
Forum answers contributed (module 2)												
None	99.6%	^C	99.5%	^C	91.0%	^{NS}	100%	^a		^a	89.3%	1057
1 or more	0.4%	^{NS}	0.5%	^{NS}	9.0%	^{A B}	^a		100%	^a	10.7%	126
Total											1183	923.0 ***

Note. Results are based on two-tailed tests, $p < 0.05$. For each significant pair, the key of the category with the smaller column proportion appears under the category with the larger column proportion.

***: $p < 0.001$

^{ABCDE}Exponents represent "Comparisons of column proportions" tests.

^{NS}There is no significant relationship.

^aThis category is not used in comparisons because its column proportion is equal to zero or one, but it does not mean that there is no significant relationship.

The Six Profiles

In our analysis, we identified six distinct profiles of behavioural engagement: Ghost, Browser, Self-Assessor, Serious Reader, Active-Independent, and Active-Social.

Ghost (n = 508, 30.0%). We identify Ghosts as potential learners who had no or almost no activity during the second week of the course. They represent a little over 30% of MOOC learners. Because Ghosts engaged in no or almost no activity in the second and third weeks, we have little to say about them. After Week 2, most of them did very few activities. In this way, they resemble Browsers.

Browser (n = 271, 16.0%). Like Ghosts, Browsers' activity level was very low. Many Browsers did not consult any resources in Week 2. Small numbers of them viewed some videos or written files or participated in the forums. No member of this group attempted or submitted any tests. They were comparable to the "lurkers" in Milligan's classification (2012).

Self-Assessor (n = 186, 11.0%). Self-Assessors are a surprising emerging profile. As the name suggests, Self-Assessors' main activities were quizzes and tests. They engaged little, if at all, with any other

type of activity. Less than 1% of Self-Assessors viewed the streaming videos or PDF files, and only 0.5% of them interacted on the discussion forums. While they did not consult any resources, they all attempted quizzes and 57.5% of them completed and submitted quizzes or tests for the second module.

Serious Reader (n = 277, 16.4%). Serious Readers were much more active than Self-Assessors in terms of viewing and reading course materials; 75% of learners with this profile downloaded at least one PDF lecture slides, 58% downloaded the required reading, and 27% viewed at least one video in streaming mode. This level of activity is fairly high and video-watching activity is underestimated because only data from video streaming and not from video downloads were collected. A large proportion of Serious Readers (74%) made at least one attempt at a quiz without submitting it, and 47% completed and submitted at least one test. Serious Readers were not active on the forums, however, where their participation was almost zero.

Active-Independent (n = 350, 20.7%). Active-Independents were more active than the preceding profiles. They behaved similarly to what we would expect from students, except that they did not engage in the discussion forums. However, they engaged more with all of the other resources than Self-Assessors and Serious Readers did. Active-Independents actively engaged with quizzes and tests, with all of them attempting at least one and submitting at least one. A substantial proportion of these learners viewed and downloaded the course materials: 63% viewed at least one PDF lecture slides, 51% downloaded the required reading, and 34% watched at least one video.

Active-Social (n = 99, 5.6%). Representing only 5.6% of learners, the Active-Socials did everything that is expected from MOOC learners and resembled regular students in their behaviour. They closely resembled Active-Independents in their engagement with readings, videos, and assessments, albeit with a slightly higher level of activity. Almost all Active-Socials attempted at least one quiz or test and submitted at least one. Over 80% of Active-Socials viewed and/or downloaded at least one PDF lecture slides and one required reading, and 54% of them viewed at least one video. Active-Socials distinguished themselves from Active-Independents through their discussion forum activity. All learners in this profile answered at least one question in the discussion forums, and 52% of them asked at least one new question.

Results of Survival Analysis

Figures 2 and 3 present the results of the discrete survival analysis procedure. Figure 2 shows the percentage of learners in each profile who “survived” (did not drop out) through the entire regular MOOC period.

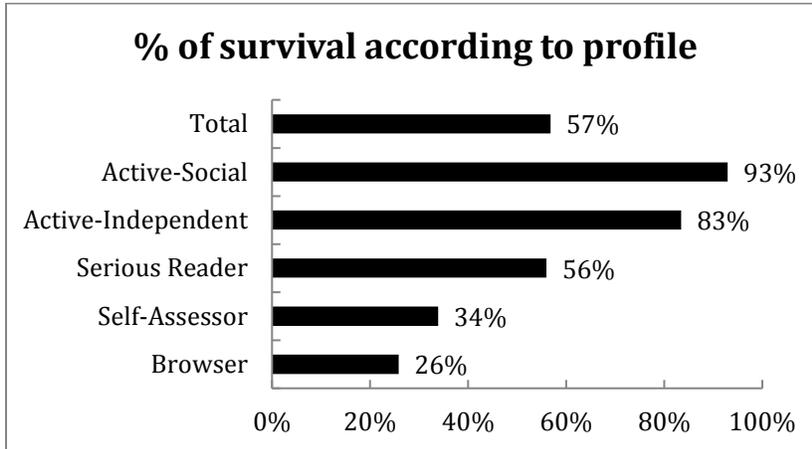


Figure 2. Results of survival analysis until week 7 for the five profiles obtained with cluster analysis (excluding Ghost).

Figure 3 shows the risk of dropping out in each week. The most active learners are represented by solid lines. The risk of dropping out is the highest in Week 2 for Self-Assessors and in Week 3 for Browsers. For Serious Readers, the proportional hazard is much lower.

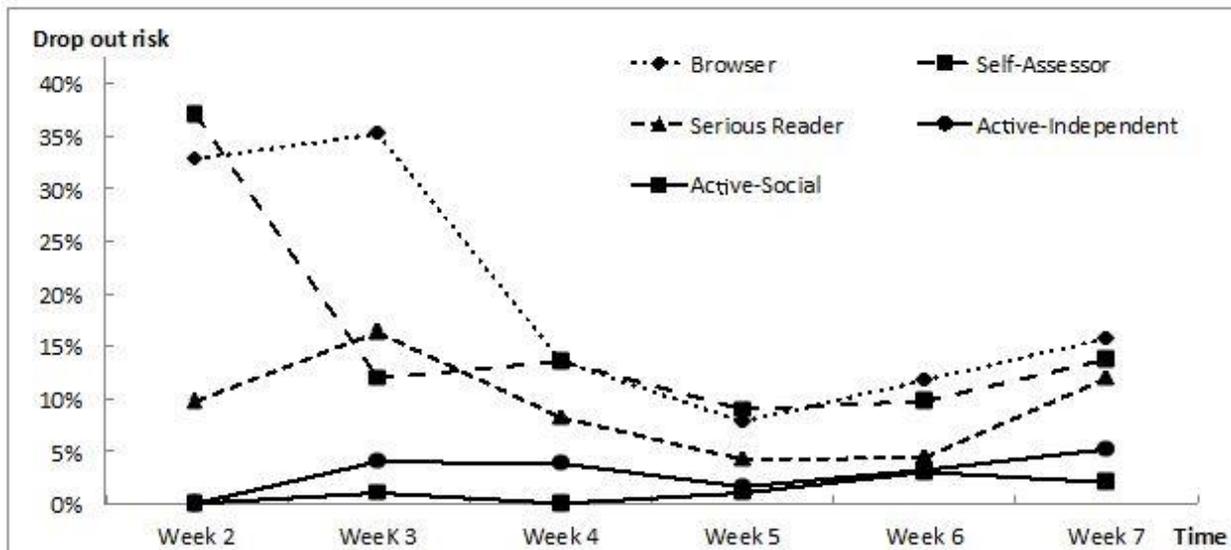


Figure 3. Time vs. dropout risk proportion for each of the participant profiles.

A log-rank test used to compare several survival curves that take into account all the follow-up time (Alberti, Timsit, & Chevret, 2005) shows significant differences between the different engagement profiles ($\chi^2(4) = 355.71, p < .001$).

Discussion

Our cluster analysis permitted us to differentiate among five different profiles of MOOC learners: *Browsers*, *Self-Assessors*, *Serious Readers*, *Active-Independents*, and *Active-Socials*, in addition to the *Ghosts* we excluded manually. Only two of our participant profiles, representing 38% of learners, engaged in ways that resembled student behaviour. Can we really therefore consider the other ones to be students? Many of those with less-engaged profiles (*Browsers*, *Self-Assessors*, and *Serious Readers*) survived the entire course, especially *Serious Readers*. Most *Serious Readers* stayed in the course for its whole duration but did not complete the MOOC because only a few of them took the weekly tests and the final exam (less than half actually completed the submission process for at least one quiz or one test).

In almost all of the literature on MOOC user profiles, we find some disengaged learners who start the activities but drop them very quickly, in addition to a very large number of registered learners who never show up in the course but are still included in the course statistics (Kizilcec et al., 2013; Whitmer, Schiorring, James, & Miley, 2015). This is confirmed in our analysis, but we were able to offer a more nuanced view of the profiles tagged in other research as “disengaged.”

What we refer to as *Browsers* resemble the “Observer” category in Hill’s typology (2013). These learners may not have had the intention to complete anything or do anything seriously. Even so, some *Browsers* may have come to the MOOC to get a glimpse of the content, see what a university course is like in that domain, or to access a subset of the content that was of interest to them.

We were puzzled by *Self-Assessors*. This profile represents learners who only accessed quizzes and tests, but who viewed no videos and read no files. Perhaps they used the MOOC for some sort of prior learning self-recognition. Alternatively, they might have used these profiles for cheating, i.e., trying the tests and quizzes on one account while completing the course on another. Their motivations are unclear and warrant further investigation.

The *Serious Reader* is an interesting profile and was quite common in our study, representing nearly one-fourth of the analysed profiles. *Serious Readers* were quite active compared to *Browsers* and *Self-Assessors*, and their participation seemed oriented towards print (PDF) materials. *Serious Readers* did not seem to care much about grades, since fewer than half of them completed and submitted tests. It is interesting to note, however, that more *Serious Readers* than *Active-Independents* (58.1% vs. 51.3%) consulted the required reading files. They may have been more intrinsically motivated than *Active-Independents*. *Serious Readers* resembled the “auditing” profile in Kizilcec et al. (2013). These users do not earn a grade and are likely to be considered failures by the institution, but are they really? Because MOOCs have fewer barriers, they permit new types of user behaviours. Whitmer et al. (2015) refer to a similar profile, which they named “Declining users.” This type of behaviour is not usually seen in presence or distance courses, for which admission and registration are costly. *Serious Readers*’ behaviour could be seen as an easier way to access the content in cases of low bandwidth, or it may simply represent a learning preference.

Users in the two most active profiles, Active-Independent and Active-Social, interacted with a variety of resources. Active-Independents look similar to the typical profile of independent learner in distance education who does most of the expected activities but has no time or interest in collaboration (Diaz & Carnal, 1999). Most of Active-Independents (83%) stayed in the MOOC for its duration.

Discussion forum activity is what distinguishes the Active-Social from the Active-Independent, the two profiles that are the most active and have the least risk of dropping out.

The Active-Social profile support results from other studies that found that learners who are socially active, interacting with others in the discussion forums, are the most engaged and the most successful (Jiang, Williams, Schenke, Warschauer, & O'Dowd, 2014). It may be that the people who are the most motivated to participate in MOOCs (those with the highest expectancy beliefs and value perceptions) engage with a maximum of course components, but it may also be that the peer communication they derived from the discussion forums, which were almost the sole means of communication available to the learners in our study, fostered greater engagement in the course.

The classification scheme we produced from our analysis suggests that these profiles should not be defined in terms of grades or success rates, as they typically have been in MOOC research. Our results suggest that we need to rethink the names and definitions used in MOOC research, especially concerning who is a student, what persevering means, and what success is from the MOOC participant's perspective.

Conclusion

In this study, we used a systematic classification procedure (two-step cluster analysis) to create five different user profiles, based on their pattern of interaction and behavioural engagement with MOOC resources for the second module of the course, after having manually removed the Ghosts (a sixth profile, not associated with the cluster analysis). In addition to the Ghost profile, of which we can say little due to the Ghosts' lack of activity in the period analysed, our classification procedures resulted in five different profiles, corresponding to an increasing level of engagement, in the following order:

1. *Browsers*, who consulted just a few resources.
2. *Self-Assessors*, whose main, almost sole activity was taking quizzes, and tests.
3. *Serious Readers*, who consulted a fairly high proportion of course materials, particularly those in print form (PDF), but who are not as active in taking quizzes and tests.
4. *Active-Independents*, who interacted with all course resources except discussion forums.

5. *Active-Socials*, who differed little from *Active-Independents* except for their discussion forum activities.

These profiles differed from those created with traditional course success or completion in mind (Hill, 2013; Kizilcec et al., 2013; Whitmer et al., 2015).

The implications of our findings for MOOC design are important. MOOCs attract a wide variety of learners, many of whom do not adopt typical student behaviours. It is difficult to know in advance the objectives and needs of different types of learners, and therefore difficult to apply a classical top-down instructional design approach such as ADDIE (Analysis, Design, Development, Implementation, and Evaluation; Gagne, Wager, Golas, Keller, & Russell, 2005). Survival analysis showed that an early understanding of learner behaviour can help us determine learning outcomes. Knowing the different patterns of participation may help us design MOOCs not only for students, but also for other types of learners, in a way that supports diverse participation styles and encourages transitions to more engaged patterns of behaviour. MOOCs may also be an ideal ground not only for personalized learning, but also for adaptive learning tailored to the needs and objectives of different types of learners.

These results raise questions about the general tendency to consider MOOC learners as equivalent to students, to design MOOC courses as if learners will only follow the path dictated by their instructors, and to judge the success of the MOOC by the standards of for-credit courses. They shed light on the characteristics, motivations, and perspectives of learners with profiles different from students' profiles. These results suggest that the way MOOCs are designed might have to be reconsidered in order to permit and encourage these alternative forms of participation in a MOOC, but a better understanding of these profiles would be needed for that.

Our analysis and classification were carried out in the context of a single MOOC, in a very specific French-Canadian context, the EDUlib initiative (www.edulib.org). The fact that the LMS data were not available for the first week of the course is also a limitation. Because the classification procedures rely solely on traces in the LMS, analysis of these profiles is subject to interpretation. Our understanding of them may benefit from both survey data and in-depth qualitative investigations of the reasons behind each of these characteristic behaviours. More research should be conducted to determine whether these profiles hold in a larger number of MOOCs, both within the EDUlib initiative and outside it. Further research is needed on the qualitative characteristics and motivations of these diverse MOOC users.

Acknowledgments

The initial phase of this research was funded by the MOOC Research Initiative, led by Georges Siemens at Athabasca University and funded by the Bill and Melinda Gates Foundation. The latest phase has been funded by SSHRC (Social Sciences and Humanities Research Council).

References

- Alberti, C., Timsit, J.-F., & Chevret, S. (2005). Analyse de survie: Le test du logrank [Survival Analysis: The log-rank test]. *Revue des Maladies Respiratoires*, 22(5), 829-832.
[https://doi.org/10.1016/S0761-8425\(05\)85644-X](https://doi.org/10.1016/S0761-8425(05)85644-X)
- Anderson, T. (2013, April 1). Promise and/or peril: MOOCs and open and distance education [Blog post]. Retrieved from <https://landing.athabasca.ca/file/view/274885/promise-andor-peril-moocs-and-open-and-distance-education>
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004, April). Off-task behavior in the cognitive tutor classroom: when students game the system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 383-390). ACM.
<https://doi.org/10.1145/985692.985741>
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman.
- Bernard, R., & Amundsen, C. L. (2008). Antecedents to dropout in distance education: Does one model fit all? *International Journal of E-Learning & Distance Education*, 4(2), 25-46. Retrieved from <http://www.ijede.ca/index.php/jde/article/viewFile/530/716>
- Carré, P. (2001). *Accompagner des formations ouvertes: conférence de consensus*. Paris: L'Harmattan.
- Clark, R. E. (1999). The CANE model of work motivation: A two-stage model of commitment and necessary mental effort. In J. Lowyck (Ed.), *Trends in corporate training*. Leuven, Belgium: University of Leuven Press.
- Diaz, D. P., & Cartnal, R. B. (1999). Students' learning styles in two classes: Online distance learning and equivalent on-campus. *College Teaching*, 47(4), 130-135.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109-132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- European Commission. (2015). Open education Europa. *Open Education Scoreboard*. Retrieved from <https://www.openeducationeuropa.eu/en>
- 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

- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109.
<https://doi.org/10.3102/00346543074001059>
- Gagne, R. M., Wager, W. W., Golas, K. C., Keller, J. M., & Russell, J. D. (2005). Principles of instructional design. *Performance Improvement*, 44(2), 44-46.
- Garson, G. D. (2014). *Cluster analysis*. Asheboro, NC: Statistical Associates Publishers.
- Gartner. (2019). *Hype cycle research methodology* (Blog post). Retrieved from www.gartner.com/en/research/methodologies/gartner-hype-cycle
- Hill, P. (2013, March 6). Emerging Patterns in MOOCs: A graphical view. [Blog post]. Retrieved from mfeldstein.com/emerging_student_patterns_in_moocs_graphical_view/
- Ho, A. D., Reich, J., Nesterko, S., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). HarvardX and MITx: The first year of open online courses. *HarvardX and MITx Working Paper No. 1*.
<http://dx.doi.org/10.2139/ssrn.2381263>
- Jiang, S., Williams, A., Schenke, K., Warschauer, M., & O' Dowd, D. (2014, July). Predicting MOOC performance with week 1 behavior. In *Proceedings of the 7th International Conference on Educational Data Mining 2014*. Retrieved from <http://educationaldatamining.org/conferences/index.php/EDM/2014/paper/viewFile/1444/1410/>
- Jordan, K. (2013). A research summary on MOOC completion rates [Blog post]. Retrieved from <http://edlab.tc.columbia.edu/index.php?q=node/899>
- Jordan, K. (2014). Initial trends in enrolment and completion of Massive Open Online Courses. *The International Review of Research in Open and Distance Learning*, 15(1), 133 – 160.
<http://dx.doi.org/10.19173/irrodl.v15i1.1651>
- Kahan, T., Soffer, T., & Nachmias, R. (2017). Types of participant behavior in a massive open online course. *The International Review of Research in Open and Distributed Learning*, 18(6), 1-18.
<https://doi.org/10.19173/irrodl.v18i6.3087>
- Khalil, M., & Ebner, M. (2017). Clustering patterns of engagement in Massive Open Online Courses (MOOCs): The use of learning analytics to reveal student categories. *Journal of Computing in Higher Education*, 29(1), 114-132. <https://doi.org/10.1007/s12528-016-9126-9>
- Kizilcec, R. F., Piech, C., & Schneider, E. (2013, April). 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).
<https://doi.org/10.1145/2460296.2460330>
- Kovanović, V., Joksimović, S., Gašević, D., Owers, J., Scott, A. M., & Woodgate, A. (2016). Profiling MOOC course returners: How does student behavior change between two course enrollments? In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale* (pp. 269-272).
- Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom, *Reading and Writing Quarterly: Overcoming Learning Difficulties*, 19(2), 119-137. <http://dx.doi.org/110.1080/10573560308223>
- Milligan, C. (2012). Change 11 SRL-MOOC study: Initial findings [Blog post]. Retrieved from <http://worklearn.wordpress.com/2012/12/19/change-11-srl-mooc-study-initial-findings/>
- Molinari, G., Poellhuber, B., Heutte, J., Lavoué, E., Widmer, D. S., & Caron, P. A. (2016). L'engagement et la persistance dans les dispositifs de formation en ligne: regards croisés. *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, (13).
- Peters, D. (2018, Feb 22). MOOCs are not dead, but evolving. [Blog post]. Retrieved from <https://www.universityaffairs.ca/news/news-article/moocs-not-dead-evolving/>
- Pintrich, P. R. (2003). Motivation and classroom learning. In Reynolds, W. M. and Miller, G. E. (Eds), *Handbook of Psychology, vol 7: Educational Psychology*, Hoboken, NJ: John Wiley & Sons, Inc., 2003, pp. 103-122. doi: <http://dx.doi.org/10.1002/0471264385.wei0706>
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*, Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning. <http://dx.doi.org/10.1080/10573560308223>
- Poellhuber, B., Roy, N., & Levasseur, C. (2017). MOOC: A vector of accessibility for higher education in developing countries? In *AERA Proceedings*, San Antonio, TX. Shah, D. (2018, Nov. 12). By The Numbers: MOOCs in 2018. In *Class-Central*. Retrieved from <https://www.class-central.com/report/>
- Roy, N., Poellhuber, B., & Bouchoucha, I. (2015). Différences régionales à travers le monde des étudiants inscrits dans un MOOC francophone: portrait d'un cas issu de l'initiative EDUlib [Worldwide regional differences among students enrolled in a francophone MOOC: A portrait of a case under the EDUlib initiative]. *Revue internationale des technologies en pédagogie universitaire/International Journal of Technologies in Higher Education*, 12(1-2), 75-92.
- Simpson, O. (2003). *Student retention in online, open and distance learning*. London: Routledge. <http://dx.doi.org/10.4324/9780203416563>

- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. New York, USA: Allyn & Bacon/Pearson Education.
- Tseng, S. F., Tsao, Y. W., Yu, L. C., Chan, C. L., & Lai, K. R. (2016). Who will pass? Analyzing learner behaviors in MOOCs. *Research and Practice in Technology Enhanced Learning*, 11(8). <https://doi.org/10.1186/s41039-016-0033-5>
- Tufféry, S. (2011). Cluster analysis. In S. Tufféry (Ed.). *Data mining and statistics for decision making* (pp. 685). Chichester: Wiley. doi:[10.1002/9780470979174.ch9](https://doi.org/10.1002/9780470979174.ch9)
- Viau, R. (2003). *La motivation en contexte scolaire [Motivation in school context]*. Louvain-la-neuve, Belgium: De Boeck Supérieur.
- White, B. (2014, August). Is “MOOC-Mania” over? In *International Conference on Hybrid Learning and Continuing Education* (pp. 11-15). Springer, Cham.
- Whitmer, J., Schiorring, E., James, P., & Miley, S. (2015). How students engage with a remedial English writing MOOC: A case study in learning analytics with big data. Retrieved from <https://library.educause.edu/~media/files/library/2015/3/elib1502-pdf.pdf>

