Student Engagement as a Predictor of xMOOC Completion: An Analysis from Five Courses on Energy Sustainability

Brenda Edith Guajardo Leal and Jaime Ricardo Valenzuela González
Tecnologico de Monterrey

John Scott
University of California, Berkeley

Abstract

MOOCs are characterized as being courses to which a large number of students enroll, but only a small fraction completes them. An understanding of students' engagement construct is essential to minimize dropout rates. This research is of a quantitative design and exploratory in nature and investigates the interaction between contextual factors (demographic characteristics), student engagement types (academic, behavioral, cognitive and affective), and learning outcomes, with the objective of identifying the factors that are associated with completion of massive and open online courses. Two logistic models were adjusted in two samples, general and secondary, with the binary dependent variable defined as completes the course yes/no. The results in the general sample (15% completion rate) showed that the probabilities of a participant completing the course are positively and significantly related to participation in the forum and the participant educational level, and negatively related to gender (female) and age. The results in the secondary sample (87% completion rate) showed that the probabilities of a participant completing the course are positively and significantly related to participation in the forum, gender (female), and the motivation and satisfaction indexes, and negatively related to age, having previous experience in other MOOCs, and self-efficacy and task strategies indexes. The results lead to ideas on how these variables can be used to support students to persist in these learning environments.

Keywords: engagement, learning analytics, xMOOC, self-regulated learning, completion

Guajardo Leal, B.E., Valenzuela González, J.R., & Scott, J. (2019). Student engagment as a predictor of xMOOC completion: An analysis from five course on energy sustainability. *Online Learning, 23*(2), 105-123. doi:10.24059/olj.v23i2.1523

Student Engagement as a Predictor of xMOOC Completion: An Analysis from Five Courses on Energy Sustainability

In education, the student engagement construct has grown popularity in recent decades as a result of a greater understanding of the role that certain cognitive, emotional, behavioral and social factors play in the process of learning and social development (Appleton, Christenson, & Furlong , 2008). In addition to the fact that the construct is considered one of the best predictors of learning and personal development, the attitude of engagement also adds to the development of essential skills to live a productive and satisfying life (Pekrun & Linnenbrink, 2012).

In massive, open, online learning courses (MOOCs) student engagement research is recent and challenging (Guajardo-Leal, Navarro-Corona & Valenzuela, 2019). MOOCs have gained reputation among academics for their impressive enrollment numbers but have also come under criticism for their poor completion rates. Research in recent years has tried to better understand these challenges around student retention in MOOCs, and what trainers and course designers could do to stop or minimize student dropout (de Barba, Kennedy, & Ainley, 2016; Halawa, Greene, & Mitchell, 2014; Guajardo-Leal & Valenzuela, 2017). Given that dropout is not an instantaneous event, but rather a gradual process that occurs over time (Appleton et al., 2008), researchers and educators in MOOC perceive student engagement as the main theoretical model to understand behavior, as its study is the basis for the design of interventions related to school desertion, attrition and success (Reschly & Christenson, 2012).

This research aims to investigate the interaction between contextual factors (demographic characteristics), student engagement (academic, behavioral, cognitive and affective), and learning outcomes, on five xMOOCs developed by the Binational Laboratory for the Intelligent Management of Energy Sustainability and Technological Training project, in order to understand the factors associated with the completion of these courses.

Student Engagement

Student engagement has been studied by professors and researchers for decades; however, there is no single definition or form of measurement. Some scholars define the concept as the beliefs and values that a subject has about the importance of learning; others state it as the effort to learn, and some more in terms of cognitive and self-regulatory strategies (Fredricks & Mccolskey, 2012). Newmann, Wehlage and Lamborn (1992) define engagement as the psychological inversion in which the student invests energy by making cognitive effort to understand something. Meanwhile, York, Gibson and Rankin (2015) indicate that engagement is a term generally used to refer to the student's psychological investment, his willingness to invest time in educational behaviors, or to a general reference of student involvement in educational activities.

A rather accepted framework is that of Reschly and Christenson (2012) who define student engagement as a process and a learning outcome that encompasses four domains: academic, behavioral, cognitive and affective. Academic and behavioral mastery implies easily observed behaviors and results in the teaching-learning process (e.g., time devoted to activities, participation in class, completion and delivery of tasks, activities or exercises, qualification in partial exams, and persistence in the course). In contrast, cognitive and affective engagement are internal domains that can hardly be observed; however, according to the authors, these domains can be accurately informed by the student (e.g., self-regulatory strategies, interest, effort, self-efficacy, belonging, and relationships with companions).

Student Engagement in MOOC

Student engagement can be conceptualized in a similar way in face-to-face education and in MOOCs; however, its operationalization in terms of the forms and processes of data collection, is totally different. Joksimović et al. (2018), based on the multidimensional model of Reschly and Christenson (2012), developed a re-operationalization of the student engagement model to explain learning in MOOCs, through an analysis of the constructs related to learning used in the prediction and measurement of student engagement (see model in Figure 1). The authors state that academic engagement in MOOCs consists of time spent on course activities, for example, participation in tests and exams, time spent in videos, participation in exercises and assignments and the completion rate (e.g., Kizilcec, Pérez-Sanagustín, & Maldonado, 2017). Behavioral engagement includes voluntary participation in academic, social or extracurricular activities; demonstrations of the behavioral dimension in MOOCs are participation in discussion forums and participation in groups and social networks (e.g., Joksimović, Gašević, Kovanović, Riecke, & Hatala, 2015). Cognitive engagement refers to students' motivational objectives and self-regulated learning skills (Reschly & Christenson, 2012); in the context of MOOCs, cognitive engagement is expressed in artifacts that students generate during the learning process, specifically in the production of texts, and it is measured with linguistic indicators of discourse, narration, cohesion and coherence (e.g., Joksimović et al., 2015). Finally, affective engagement is related to the reactions of the participants, school identification, appraisal of learning, sense of belonging, satisfaction, self-consciousness of the feelings, emotional regulation, and the abilities of resolution of conflicts (Reschly & Christenson, 2012); to measure it, Joksimović et al. (2015) rely on positive or negative language analysis.

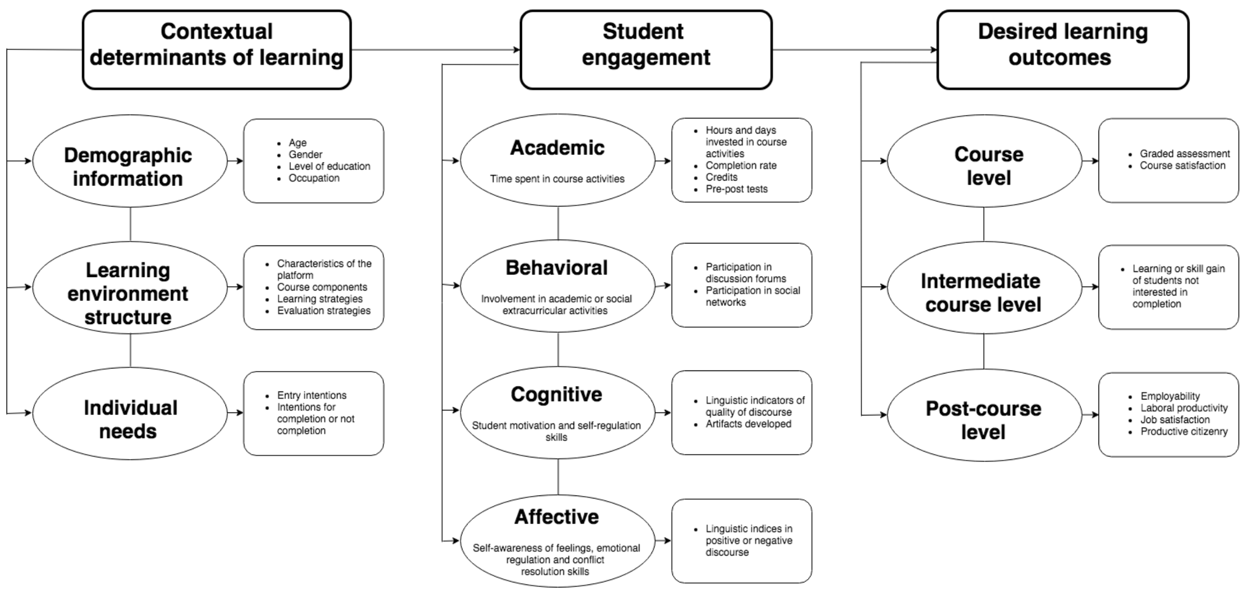

Figure 1. Student engagement model in MOOC (Adapted from Joksimović et al., 2018).

Factors Associated with MOOC Completion

Demographic variables have been commonly used to understand the factors that influence learning and / or completion in MOOC. Age, gender and level of education have been considered in several studies as predictors for student persistence or achievement, however, results differ across studies. For example, Goldberg et al. (2015) as well as Heutte et al. (2014) did not find significant differences in the probability of completing a course based on a student's level of education, while Greene, Oswald, and Pomerantz (2015) and Kizilcec and Halawa (2015) showed that more educated students are more likely to persist in a course and achieve better grades.

Pursel, Zhang, Jablokow, Choi, and Velegol (2016) examined student demographics, entry intentions, and course interactions to better understand the variables that are indicative of MOOC completion. Among their results they found that the previous online learning experience had no impact on completion; this result appears in contrast with findings from Yukselturk and Bulut (2007), who observed positive relationships between the past online learning experience and the performance in online learning environments. In addition, Pursel et al. (2016) found that the students who completed the MOOC had higher education levels, and also found that the number of times a participant watched a video and the number of posts in the forum were significant predictors of MOOC completion.

The motivation of the participants has also been studied because of its association with course completion. There is a consensus among research on the positive role of intrinsic motivation and persistence and / or achievement in MOOC (e.g., de Barba et al., 2016; Greene et al., 2015; Kizilcec & Halawa, 2015). For example, de Barba et al. (2016) performed a structural equation modelling to investigate the relationship between intrinsic motivation, participation, situational interest and performance in a sample of students that persisted until the end of a MOOC. Their results showed that motivation and participation are related to performance both directly and indirectly; motivation (value beliefs and domain focus) is mediated by participation, while participation (in videoconferences and activities) is mediated by motivation (situational interest).

The interaction and participation of students with the course materials and with their peers are also aspects that are strongly studied in MOOC research. Crossley, Dascalu, McNamara, Baker, and Trausan-Matu (2017) for example, conducted a network cohesion analysis to identify patterns related to the completion of a MOOC. Their findings showed that students who produce more quality publications in the forum are more likely to complete the course. These results are consistent with subsequent research by Engle, Mankoff and Carbrey (2015) and Goldberg et al. (2015) who demonstrated that students who collaborate most in the forum are more likely to complete the course.

Jiang, Williams, Schenke, Warschauer, and Dowd (2014) used a combination of student performance in Week 1, social interaction, and the role of external incentives to predict final performance in a MOOC. Using logistic regression as a classifier, they predicted the probability that students would obtain certificates in general, and certificates with or without distinction. Among their results they found that the average scores of the tests in the first unit strongly predict whether the students obtain the certificate; the activity of the students in the forum was not statistically significant in the predictive model. In a second model (certificate with or without distinction) they found that for each unitary increase in the number of evaluations between pairs (collaboration), the probabilities of obtaining a certificate with distinction were more than seven times greater.

xMOOCs on Energy Sustainability

The Binational Laboratory for the Intelligent Management of Energy Sustainability and Technological Training is a project financed by the energy sustainability fund of CONACYT-SENER and joined by the efforts of five higher education institutions: Tecnologico de Monterrey, National Technological Institute of Mexico (SEP), Electrical Research Institute, Arizona State University, and the University of California at Berkeley. Among the objectives of the Laboratory is the training of specialized talent in the electricity sector. To cover the training needs, a set of courses based on xMOOCs technology were developed. Five of the xMOOCs developed in this project were chosen as the scenario for the present investigation.

Courses Description

The courses are xMOOCs hosted on the MexicoX platform. Course design is a teacher-centered model focused on the delivery of high-quality content, computer-based evaluation mainly for student feedback purposes, and automation of all key transactions between participants and the learning platform, meaning all the activities and evaluations are self-contained and self-directed. Although in general terms xMOOCs are based on behavioral and cognitive learning theories, constructivist and andragogy theories were also promoted in the activities, for example, by designing a real-world challenge in energy sustainability that participants might solve.

The courses are designed for participants over the age of 17 with minimal high school studies. Each xMOOC is composed of four elements: (1) resources, (2) activities, (3) networking, and (4) evaluation. The resources available to the xMOOCs are videos (storytelling, problematization), PDFs (readings, articles, tables, processes, maps, definitions), HTML, infographics, and open resources. The course evaluation system comprises four types of assessments: (1) diagnostic, (2) progressive, (3) summative, and (4) subsequent to learning self-evaluation. The summative evaluation ranges from 0 to 100 points with a minimum pass of 60% of activities completed. A total of 10 weighted activities are evaluated: (1) six-grades in partial evaluations (quizzes), 30 points; (2) participation in the exercises, 2 points; (4) participation in a challenge, 20 points; participation in the practices with peer evaluation, 20 points; and (4) one-grade in the final exam, 28 points.

Purpose of the Study and Research Question

This study explores one central research question: What factors are associated with the completion of the energy xMOOCs? Specifically, the study attempts to understand the association between contextual factors (demographic characteristics), student engagement (academic, behavioral, cognitive, and affective), and learning outcomes at the course level, with the objective of identifying which factors are indicative xMOOCs completion. Although other investigations have taken into account demographic characteristics to predict an expected learning outcome in MOOC (e.g., de Barba et al., 2016; Greene et al., 2015; Halawa et al., 2014; Kizilcec & Halawa, 2015; Kizilcec et al., 2017), this research includes dimensions of student engagement and self-regulated learning, aspects that until now have been little used in research in the area (Joksimović et al., 2018).

Although the majority of the analysis is exploratory in nature, some specific hypotheses were considered: (1) students who exhibit higher social participation patterns, (in terms of the number of times they participated in discussion forums and practices with their peers), will be more likely to finish the course; (2) students with higher levels of education will be more likely to complete a course than those with lower levels of education; (3) students who have previous experience in a MOOC will be more likely to complete a course; (4) students who define the goal of completing the course, with or without obtaining the certificate, will be more likely to complete the course; and (5) students with high levels of self-regulation, measured in terms of self-report of self-motivation, self-efficacy, use of strategies for carrying out activities or tasks, satisfaction, and self-reaction, will be more likely to finish the course.

In the design of this study we used the student's engagement framework of Reschly and Christenson (2012) as the main conceptual base, as well as two additional frameworks: (1) the conceptual model of engagement in massive and open online learning environments of Joksimović et al. (2018), and (2) theories of self-regulated learning associated with the student engagement framework of Cleary and Zimmerman (2012). The operationalized theoretical framework used in this research can be consulted in Figure 2. An important adaptation to the model is the use of the self-regulated learning theories of Cleary and Zimmerman (2012) to operationalize the cognitive and affective dimensions of student engagement through self-reports before and during the course. The shaded parts of the model correspond to aspects to be taken into account in future research reports.

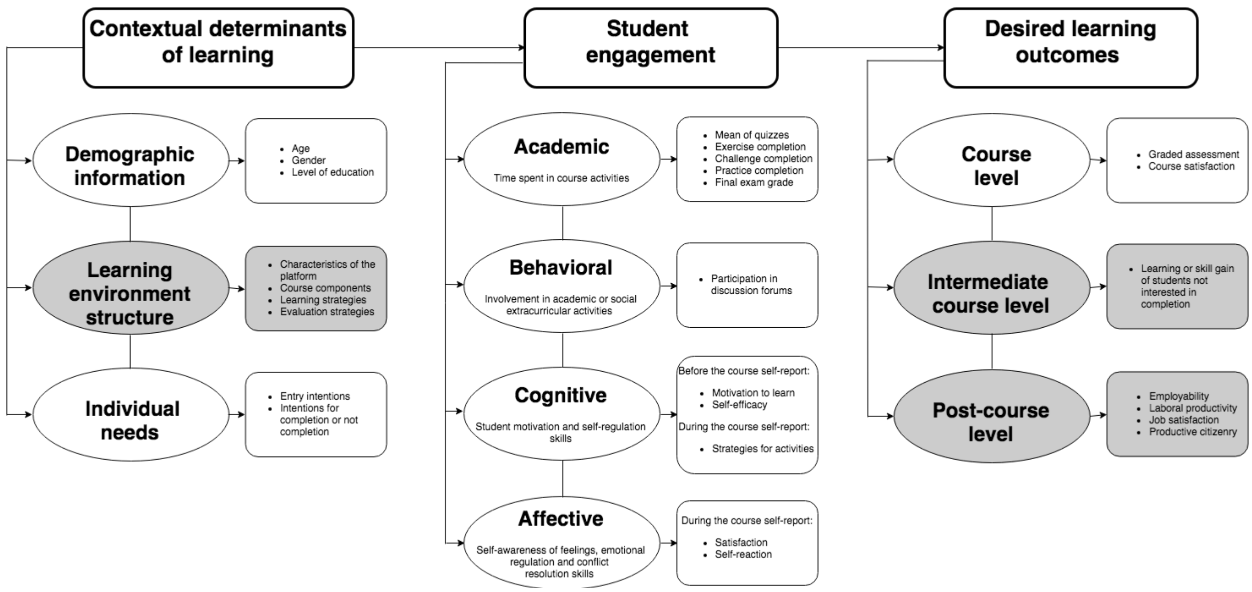

Figure 2. Operationalized model of engagement in xMOOCs for this study.

Methods

The present study is of a quantitative nature and has an explanatory approach. It uses the Binary Logistic Regression technique to address the degree to which a group of independent or explanatory variables contributes to the explained variance of a dependent dichotomous variable (French, Immekus & Yen, 2013). The dichotomous variable was defined as 1 = completed the course, and 0 = did not complete the course; in this research it is recognized that finalizing an xMOOC is not synonymous with "success" or "performance" (Breslow et al., 2013), however, the focus of this study provides an approach towards the understanding of the factors that may be associated with behavior of students who commit and complete this type of learning environments.

Participants

The main sample consisted of 50,244 participants in five xMOOCs offered by one institution, two to three times each, during August 2016 and December 2017. The secondary sample, corresponding to those of the main sample that answered two voluntary surveys, one at the beginning and one during the course, was composed of 808 participants. The breakdown of participants per course can be found in Table 1.

Table 1
Participants per Course

ID	Num. times offered	General sample	%	Secondary sample	%
xMOOC-1	2	10,746	21.4	237	29.3
xMOOC-2	3	9,475	18.9	196	24.3
xMOOC-3	3	13,657	27.2	174	21.3
xMOOC-4	3	6,343	12.6	47	5.8
xMOOC-5	3	10,023	19.9	155	19.1
Total	14	50,244	100.0	808	100.0

Instruments

Two surveys were developed: (1) The *Pre-course survey*, and (2) the *Cognitive and affective engagement of MOOC participants survey*. The *Pre-course survey* objective was to collect demographic questions (e.g., gender, age, level of studies, previous experience, and the plan to finish or just review some material and / or do some activities). The objective of the *Cognitive and affective engagement survey* was to know the level of cognitive and affective engagement of the student to plan, execute and evaluate activities with academic objectives. This second survey comprised the three phases of the engagement process in the self-regulation learning framework stated by Cleary and Zimmerman (2012): *forecast*, *monitoring*, and *self-reflection*. In each phase, some strategies that students select and use consciously to achieve their academic goals were chosen. It is important to mention that the *self-reflection phase* was carried out at the same time as the *monitoring phase* so as not to bias the results by only considering participants that completed the course.

The process of designing the items for the second survey was carried out through a review of theoretical and empirical research. From the literature, five studies related to the measurement of cognitive and affective engagement were selected: Kizilcec et al. (2017); Greene (2015); Christenson et al. (2008); Miller, Greene, Montalvo, Ravindran, and Nichols (1996); and Pintrich and DeGroot (1990). With the instruments of these sources, items that coincided with the constructs to be measured in this investigation were collected, after which they were translated and adapted to Spanish as statements in present tense. The scale of measurement was designed with values between zero and three, where a response of zero means the statement is not representative of the participant's attitude or behavior and three means the statement is very representative. The range zero-to-three was considered as a measure of the student's level of engagement on a continuum in which lower values of cognitive and affective participation indicate a superficial level of engagement, and high values indicate a deep level of engagement (Greene, 2015). The final instrument was composed of 33 items.

Before the course, the *forecast phase* of the instrument was implemented; it consisted of 14 items that measured the motivation to learn from the participants (beliefs of self-efficacy and interest [eight items] and motivation [five items]). During the course, the *monitoring and self-reflection phase* were implemented. The *monitoring phase* consisted of six items that measured cognitive and metacognitive strategies (self-instructions, image creation / self-registration, control of the work environment, search for help and cognitive monitoring). Finally, the *self-reflection* phase was composed of 13 items (self-satisfaction and satisfaction with the course [seven items], and adaptive and reactive inferences [six items]).

Content validity for the scales was established by a panel of six experts, three reviewers and three other independent experts. Then, exploratory factor analysis was performed using the principal component method and Varimax rotation. Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy for the analysis; the result was .92, well above the acceptable limit of .5 (Kaiser, 1974). Bartlett's test of sphericity, $\chi^2 = 9688.705$, $p < .00$, indicated that correlations between items were large enough for executing an exploratory factor analysis procedure. With this procedure the instrument was simplified by deleting complex items with loadings in more than two factors. After that, another factor analysis was run using a principal axis factoring method, in order to determine the underlined factor structure; Varimax was chosen as the rotation method. This time, the Kaiser-Meyer-Olkin (KMO) measure also verified the sampling adequacy for the analysis (.915), and Bartlett's test of sphericity, $\chi^2 = 8328.053$, $p < .00$, was large enough for executing an exploratory factor analysis procedure. Results, with only the remaining items from the rotated factor matrix can be seen in Table 2.

It is worth noticing that two factors (task strategies; adaptive and reactive inferences) had variables with low loadings (MG_M_ET4 y MG_E_AD6). Theoretically speaking, they were considered important by the experts and retained. Future usage of the instrument will need to re-check the items in both factors. Overall, the instrument proved to have high internal consistency $\alpha = .905$, Cronbach alphas for each scale were as follows: Self-efficacy and interest, five items, $\alpha = .879$; Motivation, five items, $\alpha = .851$; Task strategies, five items, $\alpha = .656$; Satisfaction, six items, $\alpha = .843$; and Adaptive and reactive inferences, five items, $\alpha = .638$.

Table 2

Rotated Factor Matrix Results

Scale	*Variable*	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*
Self-efficacy and interest	MG_P_AE4	.794				
	MG_P_AE5	.789				
	MG_P_AE3	.770				
	MG_P_AE2	.661				
	MG_P_AE1	.484				
Motivation	MG_P_Mot2		.748			
	MG_P_Mot3		.728			
	MG_P_Mot4		.669			
	MG_P_Mot5		.643			
	MG_P_Mot1		.594			
Task strategies	MG_M_ET5			.608		
	MG_M_ET3			.547		
	MG_M_ET2			.495		
	MG_M_ET1			.448		
	MG_M_ET4			.391		

Table 2 (Continued)

Rotated Factor Matrix Results

Scale	*Variable*	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*
Satisfaction	MG_E_S2				.791	
	MG_E_S1				.777	
	MG_E_S4				.742	
	MG_E_S6				.631	
	MG_E_S5				.545	
	MG_E_S3				.486	
Adaptive and reactive inferences	MG_E_AD2					.466
	MG_E_AD1					.391
	MG_E_AD6					.126

Note. Extraction method: Principal axis factoring; rotation method: Varimax with Kaiser normalization; six iterations.

Procedures

As a part of the instructions for each instrument, all participants were asked to complete a consent form regarding the use of data for research purposes. Responses to the instruments were organized across three databases, (1) pre-course results, (2) cognitive and affective engagement self-report results, and (3) student's profile and the interaction with the content and the course evaluations (academic and behavioral engagement). Data sources were combined using a numerical identifier and through the email of the participants. After the merging process, participants were only represented by the numerical identifier to respect confidentiality.

From the general sample (n = 50,244) the following data were obtained: Gender, educational level, country, final grade, average of partial grades (quizzes), performance of exercise, practice and challenge, final exam grade, number of times participated in the forum, whether or not the course was completed. It is important to notice that the variable *country* was collected; however, 98% of the sample belonged to a single country, for this reason the variable was discarded from the analysis. Likewise, the variable *type of course* was initially considered for inclusion in the analysis; however, after a thorough content analysis of the five courses, and after confirming our analysis with two interviews with the course developers, we found out that the five courses contained exactly the same instructional design structure: teaching strategies, type of evaluation, and the same video structure. The developers explained that the courses were developed with specific institutional templates. For this reason, we considered that there were not enough differences among the courses to consider them in the logistic models.

The secondary sample was totally voluntary, and thus only a small fraction completed both parts of the survey (n = 808). Data recovered from this sample were the following: previous experience in xMOOCs, reasons for enrollment, whether the participant planned to finish the course or not, and the motivation, self-efficacy, strategies for activities, satisfaction, self-reaction indexes. The last index of self-regulated learning was integrated with the five previous indexes. The descriptive statistics of the variables in the general and secondary samples can be found in Table 3.

Table 3
Descriptive Statistics of the Variables in the General and Secondary Samples

Name of the variable	General sample		Secondary sample	
	N	M (SD)	N	M (SD)
Completed the course	50,244	.15 (.36)	808	.87
Female	18,101	.36	270	.33
Bachelor's degree or above	31,418	.62	558	.70
Number of posts	50,244	.37 (1.58)	808	2.73 (3.67)
Participated in the challenge	50,244	.14	808	.78
Final grade	50,244	.14 (.30)	808	.80 (.23)
Mean of quizzes	50,244	.20 (.35)	808	.92 (.19)
Exercise grade	50,244	.20 (.40)	808	.90 (.29)
Practice grade	50,244	.04 (.20)	808	.46 (.49)
Final exam grade	50,244	.15 (.35)	808	.79 (.23)
With previous experience in MOOC			807	.43
Intention to finish the MOOC			798	.97
Motivation index			808	20.10 (3.13)
Self-efficacy index			808	13.09 (2.10)
Task strategies index			808	13.16 (2.62)
Satisfaction index			808	18.04 (2.42)
Self-reaction index			808	13.80 (2.65)
Self-regulated learning index			808	78.05 (10.00)

Note. For dummy variables, the averages reflect the proportions of those categories (e.g., 15% of the general sample completed the course).

Table 3 indicates very different completion rates between the general and secondary samples. In the general sample, only 15% finished the course, while in the secondary sample 87% finished the course, suggesting a great bias in the selection of the participants in the secondary sample. This may be explained because at the point of taking the second survey, about 50% of the course had passed, at which point participants were more likely to complete the course compared to those participants who only completed the entry survey. The differences between the samples are not only reflected in the rate of completion, but also in the participation of students in the forum, practices, evaluations, and exercises.

In general, correlations between most predictors were quite weak; however, there were moderate correlations between self-efficacy and motivation indexes ($r = -.46$, $p<.05$), and self-reaction and task strategies indexes ($r = -.54$, $p<.05$). This result was not surprising given the indexes measure the self-regulated learning dimensions and the cognitive and affective engagement of the student. The indexes were evaluated to verify multicollinearity with the Variance Inflation Factor (VIF); the results did not indicate multicollinearity problems.

Results

Description of the Samples

In the general sample, the completion rate was 15%. The sample comprised 64% men and 36% women. Of the men, 4% completed the course; 2% of women did likewise (see Figure 3). 62% of the participants reported having a university level education or comparable, the remainder of participants reported their education at preparatory level or lower. Half of the participants in this sample reported having had previous experience in MOOCs; thus this experience was the first one for 50%. The vast majority of participants (98%) reported having intentions or plans to complete the course in its entirety, with or without interest in obtaining the certificate. On average, participation in the discussion forum for this sample was .37 times.

The average age for those who completed the course was 31.24 years and 30.47 for those who did not. Regarding the trend in the completion of the course by age of the participants, the population enrolled in the courses is quite young; both, those who completed the course (N = 7,653; M = 31.24; SD = 11.47) and those who did not (N = 42,591; M = 30.47; SD = 9.97).

Since the participation in the pre-course survey and the survey during the course was completely voluntary, it is not surprising that the secondary sample is very different from the general sample; in the secondary sample, the completion rate was much higher (88%) than the result in the general sample. In the secondary sample, 67% are men and 33% are women. The trend in gender is similar to the general sample, 60% of men finished the course compared to 31% of women, however, the percentage of women who did not finish the course is much lower (2.2%). Seventy percent of the secondary sample have studies of bachelor education or above, the rest is at the preparatory level or below. Forty-three percent of the participants in this sample report having previous experience with other MOOCs. As in the general sample, the vast majority in this sample (97%) report having intentions or plans to complete the course in its entirety, with or without interest in obtaining the certificate. On average, participation in the discussion forum for this sample was 2.73 times, a result far greater than that of the general sample.

In the secondary sample, the results by age show the same trend as in the general sample regarding the participants who completed (N= 708; M=31.49; SD=11.70 and did not complete the course (N=99; M=33.31; SD=12.50).

Binary Logistic Regression Analysis

The variable “culmination of the course” was coded as binary (1 = yes, 0 = no) based on the weighting of the students' participation in course activities: (1) average of partial grades (quizzes), (2) participation in exercises, (3) participation in practices, (4) participation in challenges and (5) qualification in the final exam. A range of 0 to 1 was used in which the cut was defined as follows: "culminated course, yes"> = .6, “culminated course, no” <.6. Since the response variable is dichotomous, the logistic regression technique was chosen for the analysis of the variables in both samples, the general (n = 50,244) and the secondary (n = 808). It is important to emphasize that the variables used for the conformation of the dichotomous dependent variable were not used in the regressions to avoid multicollinearity. The results for the general sample are reported first.

As an initial step, participants with standardized residuals greater than 2 standard deviations (outliers) as well as those with missing data were eliminated, thus, in this first analysis, 44,881 participants were considered. A logistic model of four predictors was then fitted to test the hypothesis regarding the relationship between the probability of a participant completing a xMOOC and their gender, age, educational level and participation in the discussion forum. The logistic regression analysis was carried out by the Binary Logistic Regression procedure in SPSS version 23.

On the general evaluation of the model, the logistic model provided a better fit to the data by demonstrating an improvement over the null model (p<.001, χ^2 = 8937.995, 4) and a pseudo R^2 Nagelkerke =.544. The Goodness-of-fit test by Hosmer & Lemeshow was significant, however, studies report that the result of this test is not useful when dealing with large samples.

The results in this sample showed that the chances of a participant completing the course are positively related to forum participation and educational level (p <.001), and negatively related to gender (women) and age (p <.001, Table 4). In other words, the higher the participation in the forum and the higher the educational level of the participant, the more likely it is that the participant will complete the course. On the other hand, the odds of a woman completing the course are lower; this statement is confirmed by the negative coefficient associated with the gender predictor. In addition, for each increase in the participant's age, there is a .971 less chance of completing the course.

Table 4

Logistic Regression Analysis of 44,881 Participants to Complete a xMOOC in SPSS (Version 23)

Predictor	B (S.E.)	Wald	Exp (B)
Forum participation	1.071 (0.02)***	4454.684	2.917
Bachelor’s or above	0.269 (0.07)***	14.338	1.308
Female	-0.342 (0.07)***	25.889	0.71
Age	-0.03 (0.00)***	75.237	0.971
Constant	-3.455 (0.10)***	1162.527	0.032

Note. ***significance p<.001

In the secondary sample, we first searched for patterns in the missing data of the participants. Failing to find them, we performed the multiple imputation process with five iterations in SPSS. Thanks to the use of multiple imputation, the analysis was performed with the total sample without losing participants who lacked data (n = 808). The first step in the analysis consisted of an exploration for the identification of outliers; the sample was reduced to n = 774. With this last sample, a logistic model was adjusted with the Binary Logistic Regression procedure in SPSS version 23, with the intention of testing the hypothesis with respect to the relationship between the probability that a participant completes an xMOOC and their gender, age, educational level, participation in the forum, previous experience in xMOOCs, if they intend to finalize it or not, and the indexes of motivation, self-efficacy, strategies for tasks, satisfaction, and self-reaction.

The logistic model provided a better fit to the data by demonstrating an improvement over the null model ($p <.001$, $\chi^2 = 186.703$, 10, pseudo R^2 Nagelkerke = .490). The Hosmer-Lemeshow (HL) inferential goodness of fit test was not significant ($p> 0.05$, $\chi^2 = 5.836$, 8) suggesting that the model fit the data well, in other words, the null hypothesis of a good model adjusted to the data was sustainable.

The results of the secondary sample (Table 5) showed that the probabilities of a participant completing the course are positively and significantly related to participation in the forum, gender, motivation index and satisfaction index ($p <.05$). As in the general sample, the higher the participation in the forum, the more likely it is that the participant completes the course. In contrast or disagreement with the results of the general sample, the probabilities that a woman finished the course in this sample were greater than the probabilities for a man; this result should be interpreted with caution since trends for the gender variable in both samples is different.

Other variables positively and significantly related to the completion of a course were the motivation index and the satisfaction index. For each point of increase in the motivation and satisfaction indexes, the probabilities that a participant completes the course are higher. Results also showed that the odds of a participant completing the course are negatively related to age, the self-efficacy index, the task strategy index and previous experience in other xMOOCs ($p <.05$).The educational level and the self-reaction index did not have significant results in this regression ($p> .05$).

Table 5

Logistic Regression Analysis of 774 Participants to Complete an xMOOC in SPSS (v. 23)

Predictor	B	Wald	Exp (B)
Age	-0.042 (0.02)**	7.251	0.958
Bachelor's or above	0.573 (0.40)	1.975	1.774
Female	0.941 (0.40)**	5.81	2.564
Forum participation	0.427 (0.10)***	18.27	1.532
Previous experience in MOOC	-0.861 (0.34)**	6.212	0.423

Table 5 (Continued)

Logistic Regression Analysis of 774 Participants to Complete an xMOOC in SPSS (v. 23)

Motivation index	0.136 (0.07)**	3.873	1.146
Self-efficacy index	-0.441 (0.11)***	15.72	0.643
Task strategies index	-0.474 (0.10)***	21.99	0.623
Satisfaction index	0.733 (0.09)***	70.61	2.082
Self-reaction index	0.053 (0.09)	0.364	1.055
Constant	-0.618 (1.60)	0.169	0.539

Note. **significance p<.05; ***significance p<.001

It is important to explain that within the models, the cohort effect, defined as the effect that time, region, or life experiences may have on the development or perceptions of a particular group (Glen, 2005), was not taken into consideration for two reasons. First, the experiences lived by the group do not change significantly from one delivery to another; xMOOCs are self-contained courses that privilege the individual work of the participants. Second, the delivery of the courses was eight weeks long, and those that were given more frequently, were offered in a period of 16 months. In that sense, it was not considered a sufficiently long period of time for including the cohort variable in the models

Discussion

Before discussing the results obtained it is important to highlight some limitations of this study. One of the most important problems faced in this research was the bias identified in the selection of the secondary sample which includes a higher proportion of students who completed the course compared to the general sample. Unfortunately, and because the participation in the surveys was voluntary, it was not possible to solve this bias. In this way, the results of the secondary sample can only be generalized to the participants who answered them. A second limitation of the study is the applicability of the results to other types of MOOCs, such as a cMOOC with a more collaborative learning design compared with the content delivery design of the xMOOCs in this study.

In the general sample and in the secondary sample, the number of times the students participated in the discussion forum stands out as a predictive factor for the completion of an xMOOC when the rest of the variables remain constant. These results coincide with the existing literature (de Barba et al., 2016; Goldberg et al., 2015; Engle et al., 2015; Jiang, et al., 2014) and also support past research about a student's behavior in online environments, specifically emphasizing that peer interactions are an important component to support engagement (Kizilcec, Piech, & Schneider, 2013). These findings have important implications for how student interactions, in reference to collaboration and social integration, can be used to predict and encourage the completion of an xMOOC. It is necessary to rethink the way in which both—the collaboration and the social integration—are being promoted in xMOOCs. A plausible educative intervention would be to match the data of the platform with a social network at the student level; having these data merged could reveal richer results in three types of engagement: behavioral, cognitive and affective, allowing stakeholders and researchers to reference frequency and rigor of the participants' communicative discourse. Although this finding does not actually mean that by participating in the forum a student learns more, it does imply that higher achieving students engage more in social aspects of the course, including discussions. The learning intervention of this result could be focused on generating more opportunities for collaboration and social interaction.

The age variable maintained negative coefficients in both samples when the rest of the variables remained constant. In the models it was possible to show that for each year of increase in age, participants are less likely to finish the xMOOCs. This contrasts with the research of Greene et al. (2015) and Kizilcec and Halawa (2015) who showed that older students were more likely to persist in a MOOC. Again, the type of course (specific characteristics of the design and typology of the MOOC in question) could be a determining factor in the discord in results; this research only supports the behavior of xMOOCs participants.

A student with a high educational level (bachelor's degree, comparable or higher) was more likely to finish an xMOOC. This same pattern was shown in the investigations of Greene et al. (2015), Kizilcec and Halawa (2015) and Pursel et al. (2016), but not so for the Goldberg et al. (2015) and Heutte et al. studies (2014). It is important to mention that the interpretation of this result must be taken with caution, since the definition of "higher levels" can vary from study to study, making it difficult to compare results.

In the results of the general sample, a woman is less likely to finish an xMOOC, when the rest of the variables remain constant. This interpretation must be considered in depth since the five xMOOCs analyzed are related to electrical energy, a field which for decades men more than women were inclined to study. In fact, more men than women signed up for this course In the years to come, it would be worth investigating whether the gender gap persists.

The secondary sample suggests that most of the students who completed the course also completed the surveys (87%). Therefore, the results of the regression in this sample were somewhat different from the general sample in the educational level variable, which was not significant ($p > .05$) and in the gender variable, since women are more likely to finish the xMOOC ($= e^{2.564}$). The age in this sample however continued to have negative effects, whereas participation in the forum had positive effects in the completion of an xMOOCs.

In the secondary sample, having previous experience with other MOOCs had a negative effect, that is, those who reported that it was their first experience were more likely to finish an xMOOC. This result coincides with the research of Pursel et al. (2016) who found that the previous online learning experience had no impact on the completion of a MOOC.

Regarding the types of engagement, it is important to note that if there is a deep engagement towards the value of activities in the course and self-motivation (motivation index) before the course, there is a higher probability of finishing an xMOOC, and if during the course the satisfaction with the course itself and with the participant efforts is consistently high (satisfaction index), participants are more likely to finish the xMOOC. This indicates that the motivation and satisfaction that students bring and sustain before and during the course influence their engagement in activities. Although this is not new in educational literature (de Barba et al., 2016; Greene et al., 2015; Kizilcec & Halawa, 2015), this finding is new for xMOOCs.

On the other hand, when there are higher self-efficacy indexes and strategies for activities, the probabilities of finishing the course are lower. The negative coefficients in this regression imply that a participant who self-reports having a deep engagement in task strategies (elaboration of conceptual maps, use of tables, taking notes, use of diagrams, change of environment when it is required, among others), and that is over-confident in terms of the skills and competencies required to complete the course (for example, technological skills, search and analysis of information, study, use of social networks, among others), is less likely to finish an xMOOC. Research studies have shown that overconfidence is characteristic of humans because we tend to remember positive personality traits more easily than negative ones, or, we tend to overestimate abilities that we have to carry out an activity (Pajares, 1996). If also in xMOOC environments the confidence of a person in their performance in some activity is statistically and significantly higher than their performance in the activity, an appropriate thesis in the educational intervention for xMOOCs would be the induction of reinforcements or feedback during the course activities to return the balance to students who report overconfidence.

In sum, the results of this study provide information on some variables that show positive and negative relationships with the completion of an xMOOC. Although the results of the secondary sample may not be generalizable to other xMOOCs, the study of participant engagement demonstrated important behavior patterns that can support the design methods to keep students more involved in these types of learning environments. In a future study, the distal or post-course learning outcomes of both those who finished and those who did not finish the xMOOCs, as well as the characteristics of the learning environment that enhance such outcomes (see shaded parts of Figure 2) will be reviewed and contrasted with these results.

Acknowledgment

This research is the product of Project 266632 "Binational Laboratory for the Intelligent Management of Energy Sustainability and Technological Training", funded by the CONACYT-SENER Energy Fund. Sustainability (Agreement: S0019-2014-01)

References

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools, 45*(5), 369–386. doi: 10.1002/pits.20303

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* (March 2012), 13–25.

Christenson, S. L., Reschly, A. L., Appleton, J. J., Berman, S., Spanjers, D., & Varro, P. (2008). *Best practices in fostering student engagement.* In A. Thomas & J. Grimes (Eds.), Best practices in school psychology (5th ed., pp. 1099–1119). Bethesda, MD: National Association of School Psychologists.

Cleary, T. J., & Zimmerman, B. J. (2012). A cyclical self-regulatory account of student engagement: Theoretical foundations and applications. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 237-257). New York, NY, US: Springer Science + Business Media. doi:10.1007/978-1-4614-2018-7_11

Crossley, S., Dascalu, M., McNamara, D. S., Baker, R., & Trausan-Matu, S. (2017). *Predicting Success in Massive Open Online Courses (MOOCs) Using Cohesion Network Analysis.* In Proceedings of the12th International Conference on Computer Supported Collaborative Learning (CSCL'17) (pp. 103-110). Philadelphia, PA: International Society of the Learning Sciences.

de Barba, P. G., Kennedy, G. E., & Ainley, M. D. (2016). The role of students' motivation and participation in predicting performance in a MOOC. *Journal of Computer Assisted Learning, 32*(3), 218–231. doi: 10.1111/jcal.12130

Engle, D., Mankoff, C., & Carbrey, J. (2015). Coursera's introductory human physiology course: Factors that characterize successful completion of a MOOC. *International Review of Research in Open and Distributed Learning, 16*(2), 46–68. doi:10.19173/irrodl.v16i2.2010

Fredricks, J. A., & Mccolskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S.L. Christenson et al. (Ed.), *Handbook of research on student engagement* (pp. 763–782). Springer. doi:10.1007/978-1-4614-2018-7

French, B. F., Immekus, H. C., & Yen, H. J. (2013) Logistic regression. In T. Teo (Ed.), *Handbook of quantitative methods for educational research*. Rotterdam, Netherlands: Sense.

Glen, N. D. (2005). *Cohort analysis (quantitative applications in the social sciences)* (2nd ed.). Thousand Oaks, CA: Sage.

Goldberg, L. R., Bell, E., King, C., O'Mara, C., McInerney, F., Robinson, A., & Vickers, J. (2015). Relationship between participants' level of education and engagement in their completion of the Understanding Dementia Massive Open Online Course. *BMC Medical Education*, *15*(1), 1–7. doi:10.1186/s12909-015-0344-z

Greene, B. A. (2015). Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research. *Educational Psychologist, 50*(1), 14–30. doi: 10.1080/00461520.2014.989230

Greene, J. A., Oswald, C. A., & Pomerantz, J. (2015). Predictors of retention and achievement in a Massive Open Online Course. *American Educational Research Journal, 52*(5), 925–955. doi: 10.3102/0002831215584621

Guajardo-Leal, B. E., Navarro-Corona, C., & Valenzuela, J. R. (2019). Systematic Mapping Study of Academic Engagement in MOOC. *International Review of Research in Open and Distributed Learning. 20*(2), 114-139. Retrieved from: http://hdl.handle.net/11285/632316

Guajardo-Leal, B. E. & Valenzuela, J.R. (2017). *Transdisciplinary design of virtual learning environments: the case of an xMOOC on the study of electrical energy.* In Proceedings of the 4th International Conference Technological Ecosystems for Enhancing Multiculturality (TEEM'17), Cadiz, Spain: ACM Digital Library.

Halawa, S., Greene, D., & Mitchell, J. (2014). Dropout prediction in MOOCs using learner activity features. *eLearning Papers, 37*, 1–10.

Jiang, S., Williams, A. E., Schenke, K., Warschauer, M., & Dowd, D. O. (2014). *Predicting MOOC Performance with Week 1 Behavior.* Proceedings of the 7th International Conference on Educational Data Mining (EDM), 273–275. London, UK.

Joksimović, S., Dowell, N., Skrypnyk, O., Kovanović, V., Gašević, D., Dawson, S., & Graesser, A. C. (2015). *How do you connect? Analysis of social capital accumulation in connectivist MOOCs.* In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK'15) (pp. 64–68). New York, NY: ACM. doi:10.1145/2723576.2723604

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B. E., & Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning, 31*(6), 638–654. doi: 10.1111/jcal.12107

Joksimović, S., Poquet, A., Kovanovic, V., Dowell, N., Millis, C., Gasevic, D., & Brooks, C. (2018). How do we model learning at scale? A systematic review of research on MOOCs. *Review of Educational Research*, *88*(1), 43–86.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31–36.

Kizilcec, R. F., & Halawa, S. (2015). *Attrition and achievement gaps in online learning*. In Proceedings of the Second ACM Conference on Learning @ Scale (L@S'15) (pp. 57–66). New York, NY: ACM. doi: 10.1145/2724660.2724680

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers and Education, 104,* 18–33. doi: 10.1016/j.compedu.2016.10.001

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 (LAK'13) (pp. 170–179). Leuven, Belgium: ACM. doi: 10.1145/2460296.2460330

Miller, R. B., Greene, B. A., Montalvo, G. P., Ravindran, B., & Nichols, J. D. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology, 21*(4), 388–422. doi: 10.1006/ceps.1996.0028

Newmann, F., Wehlage, G., & Lamborn, S. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11–39). New York, NY: Teachers College, Columbia University. doi: 10.4236/ojapps.2014.45022

Pajares, F. (1996). Self-efficacy beliefs and mathematical problem-solving of gifted students. *Contemporary Educational Psychology*, *21*, 325-344.

Pekrun, R., & Linnenbrink, L. (2012). *Academic emotions and student engagement.* In S. L. Christenson, A. L. Reschly, & C. Whlie (Eds.), Handbook of Research on Student Engagement (pp. 259–282). New York, NY: Springer.

Pintrich, P. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40. doi: 10.1037/0022-0663.82.1.33

Pursel, B. K., Zhang, L., Jablokow, K. W., Choi, G. W., & Velegol, D. (2016). Understanding MOOC students: Motivations and behaviours indicative of MOOC completion. *Journal of Computer Assisted Learning, 32*(3), 202–217. doi: 10.1111/jcal.12131

Reschly, A. L., & Christenson, S. L. (2012). *Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct*. In S. L. Christenson, A. L. Reschly, & C. Whlie (Eds.), Handbook of Research on Student Engagement (pp. 3–20). New York, NY: Springer.

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. Practical Assessment. *Research & Evaluation*, *20*(5), 1–20.

Yukselturk, E., & Bulut, S. (2007). Predictors for student success in an online course. *Educational Technology & Society, 10,* 71–83.