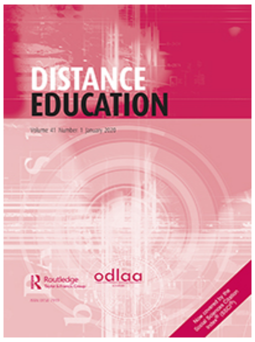

Distance Education

ISSN: 0158-7919 (Print) 1475-0198 (Online) Journal homepage: https://www.tandfonline.com/loi/cdie20

Going over the cliff: MOOC dropout behavior at chapter transition

Chen Chen, Gerhard Sonnert, Philip M. Sadler, Dimitar D. Sasselov, Colin Fredericks & David J. Malan

To cite this article: Chen Chen, Gerhard Sonnert, Philip M. Sadler, Dimitar D. Sasselov, Colin Fredericks & David J. Malan (2020) Going over the cliff: MOOC dropout behavior at chapter transition, Distance Education, 41:1, 6-25, DOI: 10.1080/01587919.2020.1724772

To link to this article: https://doi.org/10.1080/01587919.2020.1724772

Published online: 16 Feb 2020.

DISTANCE EDUCATION
2020, VOL. 41, NO. 1, 6–25
https://doi.org/10.1080/01587919.2020.1724772

Routledge
Taylor & Francis Group

ARTICLE

Going over the cliff: MOOC dropout behavior at chapter transition

Chen Chen[a], Gerhard Sonnert[a], Philip M. Sadler[a], Dimitar D. Sasselov[a], Colin Fredericks[b], and David J. Malan[c]

[a]Harvard-Smithsonian Center for Astrophysics, Harvard University, Cambridge, Massachusetts, USA; [b]HarvardX, Harvard University, Cambridge, Massachusetts, USA; [c]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA

ABSTRACT

Participants' engagement in massive online open courses (MOOCs) is highly irregular and self-directed. It is well known in the field of television media that substantial parts of the audience tend to drop out at major episodic, or seasonal, closures, which makes creating cliff-hangers a crucial strategy to retain viewers (Bakker, 1993; Cazani, 2016; Thompson, 2003). Could there be an analogous pattern in MOOCs—with an elevated probability of dropout at major chapter transitions? Applying disjoint survival analysis on a sample of 12,913 students in a popular astronomy MOOC that built participants' cultural capital (hobbyist pursuits), we found a significant increase in dropout rates at chapter closures. Moreover, the latter the chapter closure was positioned in the course sequence, the higher the dropout rate became. We found this pattern replicated in a sample of 20,134 students in a popular computer science MOOC that introduced participants to programming.

ARTICLE HISTORY

Received 26 April 2019 Final version received 30 January 2020

KEYWORDS

MOOC; learning analytics; dropout; cliff-hanger; learning sequence

Introduction

Within 10 years of development, massive online open courses (MOOCs) have become a revolutionary and democratizing force in higher education (Belanger & Thornton, 2013; Dillahunt et al., 2014; Dumitrica, 2017; Farrow, 2015; Haggard et al., 2013; Jacobs, 2013; Rice, 2014; Sanchez-Gordon & Luján-Mora, 2016). The originally economic concept of capital has been expanded and applied in the social sciences (Bourdieu, 1986; Putnam, 1995) and in education. In our context, it appears useful to distinguish human capital and cultural capital as desired educational outcomes. MOOC proponents claim that these courses not only reduce the cost of human capital training (Jones, 2015) but also transform the goal of higher education toward the cultivation of cultural capital and the satisfaction of lifelong learning (Baker et al., 2014; Hall, 2015). However, both academics and the public are concerned about the challenges that remain to be addressed in MOOCs (Ebben & Murphy, 2014), such as high dropout rates (Alraimi et al., 2015; Breslow et al., 2013; Coffrin et al., 2012; De Freitas et al., 2015; Hollands & Tirthali, 2014; Jordan, 2015) and inefficient learning

CONTACT Chen Chen chen.chen@cfa.harvard.edu Science Education Department, Harvard-Smithsonian Center for Astrophysics, MS-71, Cambridge, MA 02138, USA

management and support (Eynon, 2017; Fini, 2009; Guo & Renicke, 2014; Kizilcec & Halawa, 2015; Schophuizen et al., 2018). Despite the skepticism, many MOOC educators remain hopeful that the adoption of new media and technology can improve the user experience and persistence (Abidi et al., 2017; Bernacki et al., 2011; Kucirkova & Littleton, 2017; Laurillard, 2002).

It is well known in the field of television media that substantial parts of the audience tend to drop out at major episodic or seasonal closures, which makes creating cliff-hangers a crucial strategy to retain viewers (Bakker, 1993; Cazani, 2016; Thompson, 2003). Could there be an analogous pattern in MOOCs—with an elevated probability of dropout at major chapter transitions? Applying disjoint survival analysis on samples of two popular MOOCs (Cox & Oakes, 1984; Kleinbaum & Klein, 2005), we investigated this particular dropout pattern that would occur when learners finished the last unit of a chapter and were supposed to start the first unit of the next chapter.

Factors influencing MOOCs retention

Among the prominent factors that have been found to predict students' persistence in MOOCs, many are participant attributes, such as demographic characteristics (van de Oudeweetering & Agirdag, 2018; Zhu et al., 2018), user viewing history (He et al., 2015; Jiang et al., 2014; Kloft et al., 2014; Peng & Aggarwal, 2015), interaction (Gregori et al., 2018; Hone & Said, 2016; Jiang et al., 2014), self-reported motivation, and commitment or attitudes (Barak et al., 2016; Kizilcec & Halawa, 2015; Shao, 2018; Shapiro et al., 2017; Terras & Ramsay, 2015; Watted & Barak, 2018' Xiong et al., 2015). A few factors belong to the area of MOOC design (Guàrdia et al., 2013), such as the teachers' presence (Hone & Said, 2016; Joo et al., 2018) and accessibility (Hew, 2016), social network design (Corneli & Danoff, 2011; Li et al., 2018; Siemens, 2013), connectivist approaches (Bates, 2012; Siemens, 2012; Z. Wang et al., 2017), and assessment and feedback (Gerdes et al., 2017; Rivers & Koedinger, 2013, 2014; Vihavainen et al., 2012).

A key factor that has been frequently discussed in the MOOC literature is a learner's self-regulation (e.g., Martinez-Lopez et al., 2017; Pellas, 2014). Self-regulated learners are learners who can make plans, monitor progress, and adjust their engagement to achieve their learning goals (Carver & Scheier, 2011; Lee, 2018; McCardle & Hadwin, 2015; Reeve et al., 2008). The open structure in MOOCs affords learners a great degree of autonomy to self-regulate their course-taking behavior. The rationale behind this openness was grounded in prior research showing that learners learn the most efficiently when they achieve self-regulation (Broadbent & Poon, 2015; Pintrich, 2003; Tsay et al., 2011; C. Wang et al., 2013). Studies have shown that goal-setting (Maldonado-Mahauad et al., 2018) and time management (Lee, 2018; Lin et al., 2015; Papamitsiou & Economides, 2019)—skills that are considered easily trainable (Pintrich, 2000; J. C.-Y. Sun & Rueda, 2012)—are positively associated with students' engagement behaviors. However, research has also shown that participants' learning trajectories and patterns (Cohen et al., 2019; Rieber, 2017) are highly diverse, and that their self-directed learning pace (Cheng & Chau, 2013; Hood et al., 2015; Littlejohn et al., 2016; Milligan & Littlejohn, 2014) leads to highly irregular learning trajectories (Fini, 2009; Guo & Reinecke, 2014; Maldonado-Mahauad et al., 2018; Milligan et al., 2013)—so much so that scholars have cast doubt on the benefit of infinite freedom and called for restrictions to the open course structure in order to help students regain self-regulation (Kim et al., 2017; Zheng et al., 2015).

Learning trajectory

Numerous studies about MOOC retention (e.g., Greene et al., 2015; Allione & Stein, 2016; Wen et al., 2014; Yang et al., 2013) have adopted the survival analysis framework. The occurrence of an event (a dropout in the case of a MOOC) is the binary outcome variable of survival analysis, like in a logistic regression. However, different from logistic regression, which models a binary outcome as a function of time-invariant covariates, survival analysis models the outcome as a function of time itself (course milestones in the case of a MOOC) and of other covariates that may be time-invariant (e.g., prior knowledge) or time-variant (e.g., participation behavior). In MOOC retention studies, some learners drop out at some point during the course (termed *random censoring*), and some learners do not drop out by the end of the course (termed *right censoring*), that is, they "survive" the course. Including time in the model, survival analysis enables each participant to have a different duration in the course; therefore, the model can account for the censoring issues.

All these studies using survival analysis counted the completion of each MOOC unit, or the submitted response to the quiz of each unit, as surviving a milestone. Despite the existence of the above-mentioned irregular learning patterns, most of the studies assumed that the majority of the participants would follow the same sequence designed by the instructors. Another assumption held by many studies, particularly studies that adopted the survival analysis approach, is that the course content progresses incrementally unit by unit, without disconnects between the units. Therefore, dropout hazard was modeled as a linear function of time or milestones, rather than as a discrete function.

This assumption is arguably valid for most courses that aim to train on human capital (labor skills), in that each skill learned in prior units paves the way for more advanced skills to be acquired in the following units. Participants can hardly be equipped with sufficient skills or knowledge to fulfill specific labor market demands without completing a considerable proportion of the course. By contrast, participants who take courses that aim to build cultural capital (leisure or hobby), such as history or astronomy, might opt out at the junction of chapters (chapter in this article is defined as a set of units that congruently discusses one general topic that can be distinguished from other topics) for benchmark closure.

In this article, we call the increases in the risk of dropout at the junction of chapters the *cliff effect*. It has been well studied in the field of television that parts of the audience tend to drop out at major episodic, or seasonal, closures (Bakker, 1993; Cazani, 2016; Thompson, 2003), which makes cliff-hanger a necessary plot feature to retain viewers. Similarly, as the open education movement has pushed a coadaptation of education and entertainment to create something called edutainment (Moe, 2015), researchers have proposed using cliff-hanger strategies from entertainment media in informal learning (Fidalgo-Blanco et al., 2014) and in MOOCs (Lackner et al., 2015; Lackner et al., 2017). For those participants taking MOOCs as low-stake courses (cultural capital vs. human capital), it is possible that the more they have finished the course the more satisfied they are with their interim achievement and the less motivated they are to proceed to new chapters. Alternatively, it is also possible that the more effort and time participants have invested in a course, the more they may be motivated to proceed to new chapters in later stages of a course.

Self-determination theory

Self-determination theory (SDT) is among the most popular and empirically supported theories that explain MOOC learners' motivation and engagement (Durksen et al., 2016; Hartnett et al., 2014; Jeno et al., 2010; Ryan & Deci, 2000; Y. Sun et al., 2018). As portrayed by Ryan and Deci (2000), SDT posits that learners' intrinsic motivation is strengthened when the learning environment or instruction satisfies learners' needs for autonomy, competence, and relatedness. Y. Sun et al. (2018) showed empirical support for such a relationship. Furthermore, researchers have found that MOOC learners' intrinsic motivation positively influences their course engagement (Durksen et al., 2016; Xiong et al., 2015; Yang, 2014). In short, SDT predicts that high intrinsic motivation encourages a learner to engage and persist in a MOOC.

It is important to revisit the definition of intrinsic motivation in the SDT framework. Ryan and Deci (2000, p. 56) defined intrinsic motivation as the "doing of an activity for its inherent satisfactions rather than for some separable consequence". For example, Salmon et al. (2017) highlighted three intrinsic motivations among MOOC learners: to further existing knowledge, to acquire skills, and to apply knowledge or skills to practice. By contrast, extrinsic motivation was defined as "a construct that pertains whenever an activity is done in order to attain some separable outcome" (Ryan & Deci, 2000, p. 60), such as a certification. Based on this definition, students who drop out after completing a chapter should be considered to have low external motivation because they chose not to attain a certification as "separable outcome," but it can be argued that such students may have a relatively high intrinsic motivation because they finished a chapter without dropping out between units within the chapter. This behavior suggests that these learners were somewhat motivated to understand the complete concept domain in the chapter, but not interested in the rest of the course content, nor were they interested in the value of a certificate. In this sense, completing a whole chapter may have brought inherent satisfaction to the learners. Moreover, participants who drop out at chapter transitions in the later stages of a course could be considered even more intrinsically motivated because they have learned a substantial amount of the course content and could probably have easily obtained the certificate, but still were not motivated to attain that external reward. Could it be that SDT predicts that students with higher intrinsic motivation would be more engaged, but also more likely to drop out at chapter transitions particularly in the later stage of a MOOC, compared with students with lower intrinsic motivation? That may appear paradoxical because it contradicts the SDT prediction that highly intrinsically motivated learners should be more persistent. According to the essence of SDT, however, learners engage in a course to satisfy their inherent psychological need. If some learners' need is to learn enough of a specific concept domain, not full completion, and if they are satisfied with temporal closure, they can dropout at chapter transitions and still be considered motivated and successful MOOC learners, according to SDT and also according to many other MOOC advocates (Breslow et al., 2013; DeBoer et al., 2014; Evans & Baker, 2016; Kizilcec et al., 2013; Whitmer et al., 2014).

Research questions

In this study, we used data on students' characteristics, activities, and performance in the two MOOCs from HarvardX (on the EdX platform): "Super-Earths and Life" (SPU30x) and "Introduction to Computer Science" (CS50x). The courses were taught by a professor of astronomy and a professor of computer science, respectively, from Harvard University. The advantage of choosing an astronomy topic for this study was that this course is a typical MOOC that builds cultural capital (e.g., stimulating intellectual curiosity and amusement regarding the cosmos) rather than human capital for occupational needs (e.g., acquiring job-related skills), which we hypothesized to be more susceptible to the cliff effect (dropout risk at chapter transition). Moreover, this course adopted an interdisciplinary approach that discussed the search of life on exoplanets from the four perspectives of astronomy (exoplanets), chemistry (chemistry of life), biology (life on exoplanets), and engineering (the search for life). This course contained four chapters (each had 3–5 units), one for each of the four perspectives. Thus, there were three chapter-transitions, which provided opportunities to observe the hypothesized cliff effect.

The CS50x course contained three chapters: the first chapter (primarily using the C language) contained four units, the second chapter (primarily using the Python and SQL language) contained three units, and the third chapter (primarily using JavaScript) contained only one unit. Therefore, there are two chapter-transitions, which makes CS50x less ideal than SPU30x to investigate the change of the cliff effect over time. We used CS50x as a validation (model checking) of the model we developed based on SPU30x to examine if the cliff effect and its interaction with time still hold in a MOOC of a different topic. We did not start our analysis with a strong expectation that the cliff effect should manifest itself in the same form in both SPU30x and CS50x. Computer science skills are typically considered human capital in the labor market (as opposed to cultural capital obtained by learning about super-earths). We therefore expected that, in a computer science MOOC, the cliff effect should be minimal; and that, if there was a cliff effect, it should not increase in later chapters. If, however, a computer science MOOC demonstrated a similar cliff effect pattern to an astronomy MOOC, we would revise our expectation and argue that the cliff effect may be a more general dropout pattern applicable to different types of MOOCs.

Formally, our research questions were:

(a) Does a cliff effect exist? In other words, do MOOCs participants have higher risks of dropout at the chapter transitions (after finishing the last unit of a chapter)?
(b) Does such an effect increase or decrease at later chapters?
(c) Does an astronomy MOOC that builds cultural capital exhibit a different pattern of the cliff effect from a computer science MOOC that builds human capital?

We hypothesized that (a) A cliff effect exists; (b1) The effect increases over chapters (an interaction between the size of the cliff effect and the position of the chapter in the course sequence), which would result in participants' increased dropout after achieving intermediate benchmark closure; or (b2) The effect decreases over chapters, which would indicate that the more participants have invested in the MOOC, the higher their likelihood to proceed to a new chapter; (c1) The cliff effect is more pronounced in the astronomy MOOC than in the computer science MOOC, which would suggest that the cliff effect is specific to the topic and function of the MOOC; or, alternatively, (c2) Both MOOCs demonstrate similar patterns of the cliff effect, which would suggest it to be a more general effect in MOOCs.

Data and methods

A total of 12,913 participants enrolled in the MOOC SPU30x on HarvardX. Around 9% of the participants skipped at least one milestone in the sequence. The survival analysis is not applicable to irregular patterns. Therefore, we retained only the 11,721 regular participants. Similarly, for the MOOC CS50x, 20,134 participants finished the pre-survey, and 18,925 remained in the analytic sample after excluding irregular participants.

Pre-survey

SPU30x

In the SPU30x sample, there were 40% males and 60% females, with an overall average age of 29.5 years (*SD* = 11.2). Of the sample, 37% were living in the United States of America; 53% had a college or higher degree; 23% reported to be somewhat or very familiar with the topic; and 52% reported to be somewhat or strongly motivated to finish the course.

The pre-survey consisted of 12 items selected from the Astronomy and Space Science Concept Inventory Project (Sadler et al., 2010) to measure students' preconceptions about spatial science (the pre-test). On average, participants answered 7.95 items correctly (*SD* = 2.28).

SPU30x contained four chapters: Chapter 1 had four milestones, Chapter 2 had four milestones, Chapter 3 had five milestones, and Chapter 4 had three milestones. By the end of each milestone, participants were required to respond to a problem set (*pset*) as a marker of completion of the respective milestone. On average, participants finished 7.5 psets and spent 47 min on each milestone. A total of 3051 participants) (23%) completed all psets.

CS50x

The CS50x sample had 20,134 participants. There were 78% males and 12% females, with an overall average age of 28.8 years (*SD* = 9.88). Of the sample, 58% were living in the United States of America; 43% had a college degree as their highest educational level; 57% reported to be somewhat or very familiar with computer programming; and 67% reported to be somewhat or strongly motivated to finish the course.

Similar to SPU30x, the pre-survey in CS50x also included a pre-test. It contained 12 items testing pre-computational skills by posing logic, algorithmic, and pattern recognition questions. None of the items probed specific computer programming knowledge. These items were selected and adapted from the following:

- the University of Kent Computer Programming Aptitude Test, with the authors' kind permission (https://www.kent.ac.uk/ces/tests/computer-test.html)
- Tukiainen and Mönkkönen (2002)
- sample AP Computer Science A exam questions released by the College Board (see https://apcentral.collegeboard.org/pdf/ap-computer-science-a-course-and-exam-description.pdf for the current brochure)
- the American Computer Science League contests (https://www.acsl.org/samples.htm).

We selected 12 out of 31 questions, based on a pilot psychometric study with 911 Amazon Mechanical Turk participants. On average, CS50x participants answered nine items correctly (*SD* = 1.90).

CS50x contained three chapters: Chapter 1 had 4 milestones, Chapter 2 had 3 milestones, and Chapter 3 had only one milestone. Similar to SPU30x, CS50x required participants to respond to a problem set (*pset*) by the end of each milestone as a marker of completion of the milestone. On average, participants finished 1.5 psets and made 320 clicks in each milestone. A total of 996 participants) (4.8%) finished all 8 psets.

Analysis

The survival analysis model we specified was in many ways similar to a logistic regression model, except for allowing the outcome variable—hazard of dropout—to vary with milestones and including a time variable—milestone—and time-variant variables as predictors. The hazard was the number of dropouts at each milestone interval divided by the sample counted in that interval—similar to odds in logistic regression, except that at each milestone, both numerator (number of dropouts) and denominator (remaining sample) changed. The model was specified as:

$$logit\ h\left(m_{ij}\right) = \alpha_1 + \beta_1 M_{ij} + \beta_2 {M_{ij}}^2 + \beta_3 Covariate_{1ij} + \ldots + \beta_4 Cliff_{ij} + \beta_5 M_{ij} \times Cliff_{ij}$$

Numerous covariates were included in this model; most of them were time-invariant variables. They include students' age, gender, motivation, familiarity, pre-test score, foreign status, and self-described extrovert personality. Such variables were measured only in the initial questionnaire. There was one time-varying covariate, which was activity in the previous milestone (for milestone 1, activity in the previous milestone was defined as activities in the course introduction session before starting milestone 1). In SPU30x, the activity was measured by active time spent (standardized within milestone) by the participants; in CS50x, the activity was measured by clicks (standardized within milestone) made by the participants. We used activity measures as proxies for students' engagement, though the validity of such usage is arguable (Holmes et al., 2019). All of the covariates were included to control for prominent predictors of MOOC retention that had been identified by previous literature. They were, nevertheless, not the primary interest of our study.

The key predictor in our model was the variable *cliff*, which was a time-varying predictor (1 if a milestone was the first unit of a chapter; 0 if a milestone was not the first unit of a chapter, so that we could account for the participants who finished the last unit of the previous milestone, but did not proceed to the new chapter). For SPU30x, there were 16 milestones, and three of the milestone locations had *cliff* = 1 (milestones = 5, 9, 14). For CS50x, there was a total number of 8 milestones, and there were two milestone locations where *cliff* = 1 (milestones = 5, 8). If the parameter for *cliff* was positive, it would indicate that the likelihood of dropout increased at the chapter transition, compared with the predicted baseline at that milestone (as specified by M_{ij} and ${M_{ij}}^2$ explained below). We also specified an interaction effect between *cliff* and milestone (hereafter *cliff* × *M*). A significant interaction effect would indicate that the cliff effect changes over the duration of the course.

M_{ij} and M_{ij}^2 were linear and quadratic specifications of the baseline effect of milestone on logit hazard at milestone j for individual i. They represented the "natural" curve of logit hazard over the milestones if there were not any chapter transition. There were other possible specifications of the milestone effect (e.g., as dummy variables or as a linear effect only). Upon inspection of the logit hazard function, we noticed that the logit hazards did not exhibit a linear trend; there was indeed a decreasing trend, but a nonlinear one with a flat tail. The flat tail could be the result of an increasing cliff effect by the end of the course (interaction between cliff and linear effect of milestone). However, this could potentially lead to an over-interpretation of the flatness, as it might be attributable to the possibility that students lost stamina and became more likely to drop out by the later stages, even at regular (non-cliff) milestones, which can be modeled as simply a quadratic specification of milestone, hence the term M_{ij}^2 (it can also be understood as an interaction effect of milestone with itself, that is, that the milestone effect became modified by the later milestones). We decided that a quadratic specification would parsimoniously reflect the hazard function in our case and control for the natural increase of dropout during the later regular milestones. We also preferred this specification because it gave us a conservative estimate of the interaction effect between the cliff and milestone ($cliff \times M$), which was one of the key parameters of interest: if we adopted a linear specification (which we report as an add-on at the end of the Results section), the main effect of milestone would predict a linear downward trend without a tail. Therefore, the upward "force" that lifted the flat tail would be fully attributed to $cliff \times M$, which potentially overestimates this interaction effect. By including a quadratic term, part of the "lifting force" would be explained by the effect of milestone alone, rendering our estimation of $cliff \times M$ conservative.

For each sample, we built survival models separately. We first built models without the interaction effect, then added models with interaction effects included. The CS50x had fewer chapters, and the last chapter had only one milestone. Thus, it was not an ideal setting to test the cliff effect. For this reason, we focussed our interpretation on the parameters of the SPU30x models and used the CS50x models as a validation check for possible generalizability.

Results

Table 1 presents the parameters for the fitted models. It does not include the parameters for the controlled variables, however, although these variables have been controlled for.

Interpretation of the parameters is analogous to the interpretation of a logistic model: The coefficient (β) corresponded to the amount of change in logit hazard associated to one unit of change in the predictor. The logit hazard can be further converted to an odds ratio. For example, in M1.1, the main effect of chapter transition ($\beta_{cliff} = 0.645$) indicated that the logit hazard at a chapter transition was larger than the logit hazard for the same milestone if it was not a chapter transition by 0.645, controlling for other covariates. This could further translate to an odds ratio of 1.906 ($e^{.645} = 1.906$), which means that the odds of dropping out at a chapter junction were 1.906 times that of the odds of dropping out when there was no chapter junction. Similarly, in M2.1, the logit hazard of chapter transition was 0.654, which corresponded to an odds ratio of 1.923. In short, the main effects of chapter transition were nearly the same (roughly an odds ratio of 1.9) between

Table 1. Survival analysis predicting dropout from SUP30x and CS50.

	SPU30x				CS50x			
	M1.1		M1.2		M2.1		M2.2	
	β	*SE*	β	*SE*	β	*SE*	β	*SE*
(Intercept)	0.320	0.114**	0.497	0.117***	0.196	0.047***	0.250	0.047***
milestone	-0.370	0.013***	-0.451	0.017***	-0.219	0.010***	-0.247	0.011***
unitfirst	0.342	0.100***	-2.158	0.220***	0.703	0.061***	-0.938	0.200***
milestone x unitfirst			0.420	0.030***			0.297	0.034***
Controlling for:	gender, age, pre-test, familiar, motivation, time spent in the previous milestone				gender, age, pre-test, experience, motivation, clicks made in the previous milestone			

*Notes. ** p < 0.01, *** p < 0.001, after false discovery rate adjustment.*

the two MOOCs. Interpreting the *recent activity* terms in the same fashion, we concluded that, for both MOOCs, the more activity (in terms of time duration for SPU30x and number of clicking for CS50x) students engaged in at the previous milestone, the less likely were they to drop out.

Next, focussing on the interaction effect of *cliff* × *M* (M1.2) for SPU30x, we found that the interaction term was positive and statistically significant, which suggested that the logit hazard of dropout at cliffs increased over milestones. For example, in M1.2, at the first chapter transition (milestone = 5), the logit hazard increased by $\beta_{cliff} + 5 \times \beta_{cliff} = -1.064 + 5 \times 0.252 = 0.196$, which translated to an odds ratio of 1.217. This effect increased as the number of milestones increased to 9 (the second transition) and 14 (the third transition), where the logit hazards were 1.204 and 2.464 respectively, and the corresponding odd ratios were 3.333 and 11.751. We further calculated the estimated marginal probability of dropping out (comparing the probability of dropout at a given milestone when *cliff* = 1 against when *cliff* = 0) at milestones 5, 9 and 14, while controlling the other covariates at their *means*. We estimated that the probability of dropping out increased by 2, 2.5, and 3.6 percentage points. As illustrated in Figure 1, the dropout rate (decreasing in general with a flat tail) was bumped up by the chapter transition (counted at the first milestone of a new chapter), and this bump increased in magnitude as the course proceeded to latter stages.

Lastly, focussing on M2.2, we found the interaction effect of *cliff* × *M* to be statistically non-significant for CS50x. Using an analogous approach as above, we illustrated the cliff effect in CS50x in Figure 2. The probability of dropping increased by 14% at the first transition and 15% at the second transition. Visually, it appeared that the cliff effect at the second transition was larger than at the first transition. However, the interaction effect was not statistically significant, mainly because the overshoot at the second transition was partially explained by the quadratic term of the milestone effect. When we specified a linear model without the quadratic term, we detected both a significant main effect of transition ($\beta = -0.938$, i = 0.200, $p < 0.001$) and a significant interaction effect of *cliff* × *M* ($\beta = 0.297$, $SE = 0.034$, $p < 0.001$). For consistency, we also tested a linear model for SPU30x, which yielded a significant effect of milestone ($\beta = 0.420$, $SE = 0.030$, $p < 0.001$). The linear models excluded the possibility that the nonlinear upward-trending tail shown in Figure 1 and Figure 2 might be partially explained by the milestone itself. Rather, they attributed all nonlinearity to the interaction effect of *cliff* × *M*. We report the linear

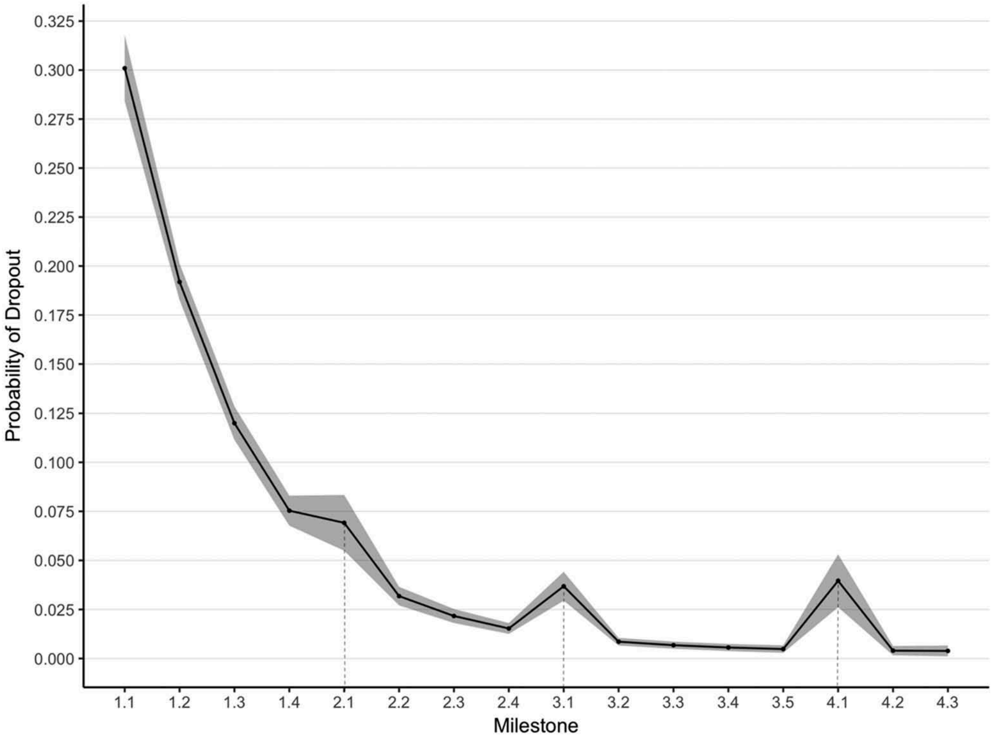

Figure 1. Plotting M1.2 fitted probability of dropout at each milestone, taking *cliffs* into consideration and controlling other covariates at their *means*.

models to demonstrate that the conclusion could vary depending on the model specification. However, we would like to focus on the quadratic models because they were less restricted and provided a more conservative estimation of the *cliff* $\times$ *M* term.

Discussion and conclusion

The most important finding of this study is the detection of a cliff effect, namely the overshooting of the dropout rate at chapter transitions. Cliff effects were defined as participants finishing the last unit of the previous chapter and not returning to the upcoming new chapter. The fact that the cliff effect increased across milestones, at least for SPU30x, suggests that the more participants had learned, the less motivated they were to learn a new topic. This result overturned our alternative hypothesis that the more participants had already invested and learned in the course, the more motivated they would be to finish the course. Whereas, in general, the latter appeared true and was captured in the typically falling dropout rates over time, the cliff effect might be explained by a fatigue factor in conjunction with a sense of accomplishment or closure by finishing a major milestone, which may have led to a reluctance to start a new chapter.

If we consider the total number of milestones completed to reflect commitment or motivation at the macro level, we can consider the time spent or clicks made at the

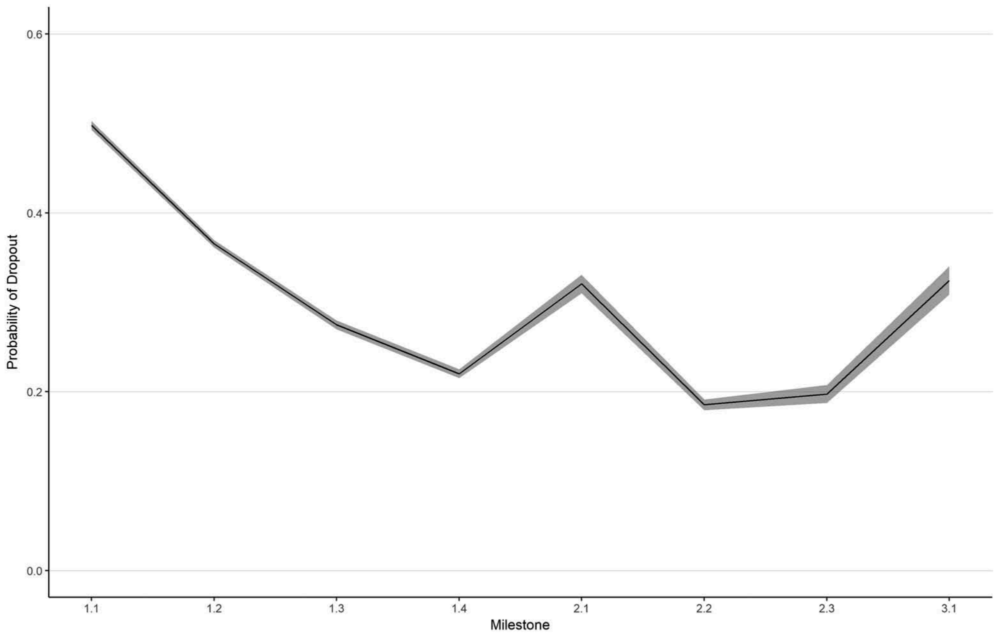

Figure 2. Plotting M2.2 fitted probability of dropout at each milestone, taking *cliff* into consideration and controlling other covariates at their *means*.

previous unit as the commitment or motivation at the micro level. We found an interesting contrast between the two levels. At the macro level, as discussed above, the more milestones participants had invested in learning the less likely they were to start a new chapter, whereas at the micro level, as suggested by the effect of *recent activity*, the more time participants spent, or the more clicks participants made, in the most recent unit, the more likely they were to remain in the upcoming unit. In other words, our alternative hypothesis of the positive impact of time investment on attrition was partially supported, except that it was not applicable at major chapter transitions. This pattern may partially resolve the paradox of SDT (Ryan & Deci, 2000) introduced earlier. Intrinsic motivation or inherent satisfaction pushes a student to complete a chapter in order to understand a coherent concept domain, but such a satisfaction may also provide the learner a psychological closure that reduces the likelihood of initiating a new concept domain.

When scholars consider motivations in an online course, they often talk about a learner's motivation to start a course and the motivation to complete a course. Seldom do they describe the decision of a learner who drops out in the middle of a course to be motivational. However, findings in this study inspired us to reflect upon the concept of learner motivation and contemplate a smaller 'grain size" of motivation. It is possible that a learner is motivated to finish a chapter to acquire the complete chapter of knowledge of interest, but drops out immediately after finishing the chapter. We can still consider such learners as intrinsically motivated in that they know what they need, retrieve what they need, and ignore the value added in completing the whole course. Prior studies categorized such learners as samplers (Coffrin et al., 2014; DeBoer et al., 2014). Our result show that samplers do not have to be irregular learners. Motivated samplers might also sample by chapters (clusters of units) rather than by single units.

We used the data from CS50x as a model check to either validate the result of the SPU30x analysis, or to discover that the pattern is highly course specific. We originally hypothesized that the cliff effect would be stronger for SPU30x, a course that built cultural capital, than for CS50x, a course that built human capital. Our reasoning was that cultural capital acquisition is subject to relatively unrestricted free choice, whereas human capital acquisition places a heavier weight on mastering a complete skillset. Nevertheless, our result from the CS50x sample largely replicated, and thus reinforced, the result from the SPU30x sample: there was a strong cliff effect at chapter transitions, and there was a positive effect of recent activity on attrition. The hypothesized effect of the type of capital was not detected. Therefore, we had stronger evidence to argue the cliff effect was a MOOC general phenomenon. Whereas, admittedly, a study of two MOOCs is still a rather shaky basis for generalization, the fact that a cliff effect of almost identical magnitude was found in two quite dissimilar MOOCs encourages further examination of the cliff effect across varying types of MOOCs.

We found only partial evidence of the interaction effect between cliff and milestones. The cliff effect increased by the latter stages of the course SPU30x. However, the effect did not increase, nor did it decrease, by the latter stages of the course CS50x. We argue that the reasons behind not detecting a significant interaction effect in CS50x were that we used a quadratic model that provided a conservative estimation of the interaction effect, as explained above, and relatedly, that the last chapter in CS50x had only one milestone, so that the overshoot of dropout rate at this milestone put a strong weight on the quadratic term. Had we had one additional milestone for the last chapter, we would be able to better estimate the quadratic term and parse out its effect from the interaction effect of $cliff \times M$. One clear takeaway from the analysis of CS50x, though, is that the cliff effect did not diminish over time. Even if we used the more conservative estimation, it stayed strong.

Findings in this study have strategic policy implications: If the goal of the MOOCs providers is to encourage participants to complete greater rather than lesser amounts of a course, these providers need to find strategies for preventing dropout at the moment of topic transitions. The general accepted principle to counteract the cliff effect is to implement cliff-hangers. Specifically, the guiding strategies might consider downplaying the distinction between chapters, overviewing the big picture at the beginning of the course so that participants understand that each chapter is only one piece of the puzzle, building suspension (e.g., raising new questions, discovering new confusions) by the end of a chapter, previewing of the upcoming chapter and explaining its connection and importance to what had been learned in previous chapters, embedding extra motivational work (e.g., extra doses of the abovementioned strategies) to the transitions at latter chapters, and explicitly asking learners about their chapters of interests and their sense of learning closure (which does not have to equate to full course completion) to have a fuller understanding of learners' motivation and a better anticipation of learners' completion. In fact, both SPU30x and CS50x have implemented some of the above mentioned strategies. This suggests that, without these implementations, the dropout at chapter transitions might have been even higher. Future studies and practices should experiment with enhanced cliff-hangers to chapter transitions in MOOCs and evaluate the effectiveness of such an intervention, and the effectiveness of different components of the intervention, in dropout prevention.

Garreta-Domingo et al. (2018) proposed the “teachers as designers” approach in MOOC development. According to their proposal, teachers should not only design the pedagogy for teaching the content knowledge but also take a learner-centered perspective through which they may capitalize on the students’ experience and shape the design of the MOOC structure, interface, and workflow to hold students’ attention. As noted by Terras and Ramsay (2015), the greater autonomy that MOOCs provide presents greater challenges because the burden of learning regulation shifts from the instructors to the learners. Nevertheless, instructors who have a design mindset and take a learner-centered perspective should partially share the burden of learning regulation. For example, instructors should take the cliff effect into consideration and make chapter transitions not only smooth in terms of the content, but also less segmented with regard to the interface.

Future replication studies should make amendments to avoid the limitations of this study. One major limitation is the small numbers of chapter transitions in the MOOCs, especially in CS50x. To make a more accurate estimation of the cliff effect and its interaction with milestones, we recommend analyzing MOOCs data that contain at least four chapters, and at least two milestones per chapter. Another limitation of the study was the small sample of MOOCs. In our case, there were only two MOOCs that had two different topics. Although we argued that the astronomy MOOC was tailored for cultural capital training and the computer science MOOC was tailored for human capital training, the division was not precise. A substantial proportion of computer science MOOC participants took the course for hobby and not professional development (67% of CS50x vs. 34% of SPUX30x participants reported that they were interested in obtaining certificates), which could explain the striking similarity in the cliff effects between the two MOOCs. We suggest a systematic survey of available MOOCs data to examine how widespread the cliff effect is across on different topics, pedagogies and platforms of MOOCs.

Whereas college education has traditionally dealt with a captive audience, where students face stiff penalties for dropping out during a course (e.g., the loss of tuition money and educational credits), the MOOC format has an extremely volatile audience. The makers of MOOCs, therefore, should consider adapting techniques from the entertainment and other leisure industries, intended to maximize the retention of their audience. In our study, for example, the discovery of a cliff effect calls for techniques such as the cliff-hangers to mitigate its effects on student dropout.

Acknowledgments

This work was supported by the National Science Foundation under the grant titled Outcome Predictions of Students in Massive Open Online Courses (OPSMOOC) (grant number DRL-1337166) and EAGER: Student Outcomes in a Computer Science MOOC (SOCSMOOC) (grant number DUE-1352696). Any opinions, findings, and conclusions in this article are the authors’ and do not necessarily reflect the views of the National Science Foundation. We thank Glenn B. Lopez for transmitting, and Alaalden Ibrahim and John Murray for processing, the MOOC data. We also thank those who gave us additional support and direction: Charles Alcock, Lori Breslow, Andrew Ho, Annie Valva, Rob Lue, and Wendy Berland.

Disclosure statement

No potential conflict of interest was declared by the authors.

Notes on contributors

Chen Chen is a postdoctoral fellow in the Science Education Department of the Harvard-Smithsonian Center for Astrophysics. He studies the MOOC retention problem, misconceptions in science learning, and media technology in education.

Gerhard Sonnert is a research associate in the Science Education Department of the Harvard-Smithsonian Center for Astrophysics, where he has worked on several large empirical studies in science, technology, mathematics, and engineering education, with a particular focus on gender aspects in STEM. He teaches a course on astrosociology in the Harvard Astronomy Department.

Philip M. Sadler is director of the Science Education Department of the Harvard-Smithsonian Center for Astrophysics. He is the F. W. Wright Senior Lecturer on Navigation in the Department of Astronomy. His research includes assessment of students' scientific misconceptions and how these change as a result of instruction or pedagogic techniques.

Dimitar D. Sasselov is professor of astronomy at Harvard University and Director of the Harvard Origins of Life Initiative. He is also a co-investigator on NASA's Kepler mission, searching for exoplanets the size of Earth. He teaches the MOOC "Super-Earths and Life" on HarvardX.

Colin Fredericks is senior project lead at HarvardX, where he helps to create MOOCS and advance the use of pedagogically driven technology. He was the course coordinator for the examined instance of the Super-Earths and Life MOOC.

David J. Malan is Gordon McKay Professor of the Practice of Computer Science in the School of Engineering and Applied Sciences and a Member of the Faculty of Education in the Harvard Graduate School of Education. He teaches "Computer Science 50," otherwise known as CS50, which is Harvard University's largest courses, one of Yale University's largest courses, and edX's largest MOOC.

ORCID

Chen Chen http://orcid.org/0000-0002-6065-8889

References

Abidi, S. H., Pasha, A., Moran, G., & Ali, S. (2017). A roadmap for offering MOOC from an LMIC institution. *Learning, Media and Technology*, *42*(4), 500–505. https://doi.org/10.1080/17439884.2016.1205601

Allione, G., & Stein, R. M. (2016). Mass attrition: An analysis of drop out from principles of microeconomics MOOC. *The Journal of Economic Education*, *47*(2), 174–186. https://doi.org/10.1080/00220485.2016.1146096

Alraimi, K. M., Zo, H., & Ciganek, A. P. (2015). Understanding the MOOCs continuance: The role of openness and reputation. *Computers & Education*, *80*, 28–38. https://doi.org/10.1016/j.compedu.2014.08.006

Baker, R., Evans, B., Greenberg, E., & Dee, T. (2014, February 10–12).Understanding persistence in moocs (massive open online courses): Descriptive & experimental evidence. In U. Cress & C. D. Kloos (Eds.), *Proceedings of the Second European MOOC Stakeholders Summit 2014* (pp. 5–10). Ecole Polytechnique Federale de Lausanne. http://hdl.voced.edu.au/10707/340125

Bakker, E. J. (1993). Activation and preservation: The interdependence of text and performance in an oral tradition. *Oral Tradition*, *8*(1), 5–20. http://admin.oraltradition.org/wp-content/uploads/files/articles/8i/8_1_complete.pdf

Barak, M., Watted, A., & Haick, H. (2016). Motivation to learn in massive open online courses: Examining aspects of language and social engagement. *Computers & Education*, *94*, 49–60. https://doi.org/10.1016/j.compedu.2015.11.010

Bates, T. (2012, August 5). What's right and what's wrong about Coursera-style MOOCs? *Online Learning and Distance Education Resources*. http://www.tonybates.ca/2012/08/05/whats-right-and-whats-wrong-about-coursera-style-moocs/

Belanger, Y., & Thornton, J. (2013). *Bioelectricity: A quantitative approach – Duke University's first MOOC*. Duke University. http://dukespace.lib.duke.edu/dspace/handle/10161/6216

Bernacki, M., Aguilar, A., & Byrnes, J. (2011). Self-regulated learning and technology-enhanced learning environments: An opportunity propensity analysis. In G. Dettori & D. Persico (Eds.), *Fostering self-regulated learning through ICT* (pp. 1–26). IGI Global. http://doi.org/10.4018/978-1-61692-901-5.ch001

Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood. https://www.gbv.de/dms/hebis-mainz/toc/009302689.pdf

Breslow, L., 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* (1), 13–25. http://www.rpajournal.com/studying-learning-in-the-worldwide-classroom-research-into-edxs-first-mooc/

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Carver, C. S., & Scheier, M. F. (2011). Self-regulation of action and affect. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (2nd ed., pp. 3–21). Guilford Press. https://psycnet.apa.org/record/2004-00163-000

Cazani, E. (2016). The cliffhanger phenomenon: The tension and the interruption. *REVISTA MEDIACAO*, *18*(23), 83–97. http://www.fumec.br/revistas/mediacao/article/view/4115/pdf

Cheng, G., & Chau, J. (2013). Exploring the relationship between students' self-regulated learning ability and their ePortfolio achievement. *The Internet and Higher Education*, *17*, 9–15. https://doi.org/10.1016/j.iheduc.2012.09.005

Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014, March 24–28).Visualizing patterns of student engagement and performance in MOOCs. In A. Pardo & S. D. Teasley (Eds.), *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 83–92). Association for Computing Machinery Press. https://doi.org/10.1145/2567574.2567586

Cohen, A., Shimony, U., Nachmias, R., & Soffer, T. (2019). Active learners' characterization in MOOC forums and their generated knowledge. *British Journal of Educational Technology*, *50*(1), 177–198. https://doi.org/10.1111/bjet.12670

Corneli, J., Danoff, C. J. (2011, June 30 – July 1). Paragogy. In S. Hellmann, P. Frischmuth, S. Auer, & D. Dietrich (Eds.), *Proceedings of the 6th Open Knowledge Conference* (pp. 13–23). CEUR-WS.org. http://ceur-ws.org/Vol-739/paper_5.pdf

Cox, D. R., & Oakes, D. A. (1984). *Analysis of survival data*. Chapman & Hall.

De Freitas, S. I., Morgan, J., & Gibson, D. (2015). Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision. *British Journal of Educational Technology*, *46*(3), 455–471. https://doi.org/10.1111/bjet.12268

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing "course" reconceptualizing educational variables for massive open online courses. *Educational Researcher*, *43*(2), 74–84. https://doi.org/10.3102/0013189X14523038

Dillahunt, T. R., Wang, B. Z., & Teasley, S. (2014). Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. *The International Review of Research in Open and Distributed Learning*, *15*(5). https://doi.org/10.19173/irrodl.v15i5.1841

Dumitrica, D. (2017). Fixing higher education through technology: Canadian media coverage of massive open online courses. *Learning, Media and Technology*, *42*(4), 454–467. https://doi.org/10.1080/17439884.2017.1278021

Durksen, T. L., Chu, M. W., Ahmad, Z. F., Radil, A. I., & Daniels, L. M. (2016). Motivation in a MOOC: a probabilistic analysis of online learners' basic psychological needs. *Social Psychology of Education*, *19*(2), 241–260. https://doi.org/10.1007/s11218-015-9331-9

Ebben, M., & Murphy, J. S. (2014). Unpacking MOOC scholarly discourse: a review of nascent MOOC scholarship. *Learning, Media and Technology*, *39*(3), 328–345. https://doi.org/10.1080/17439884.2013.878352

Evans, B. J., & Baker, R. B. (2016). MOOCs and persistence: Definitions and predictors. *New Directions for Institutional Research*, *2015*(167), 69–85. https://doi.org/10.1002/ir.20155

Eynon, R. (2017). Crowds, learning and knowledge construction: questions of power and responsibility for the academy. *Learning, Media and Technology*, *42*(3). https://doi.org/10.1080/17439884.2017.1366920

Farrow, R. (2017). Open education and critical pedagogy. *Learning, Media and Technology*, *42*(2), 130–146. https://doi.org/10.1080/17439884.2016.1113991

Fidalgo-Blanco, A., Sein-Echaluce, M. L., García-Peñalvo, F. J., & Escaño, J. E. (2014, October 1–3). Improving the MOOC learning outcomes throughout informal learning activities. In F. J. García-Peñalvo (Ed.), *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 611–617). Association for Computing Machinery. https://doi.org/10.1145/2669711.2669963

Fini, A. (2009). The technological dimension of a massive open online course: The case of the CCK08 course tools. *The International Review of Research in Open and Distance Learning*, *10*(5), 1–26. https://doi.org/10.19173/irrodl.v10i5.643

Garreta-Domingo, M., Sloep, P. B., & Hernández-Leo, D. (2018). Human-centred design to empower "teachers as designers". *British Journal of Educational Technology*, *49*(6), 1113–1130. https://doi.org/10.1111/bjet.12682

Gerdes, A., Heeren, B., Jeuring, J., & van Binsbergen, L. T. (2017). Ask-Elle: An adaptable programming tutor for Haskell giving automated feedback. *International Journal of Artificial Intelligence in Education*, *27*(1), 65–100. https://doi.org/10.1007/s40593-015-0080-x

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. https://doi.org/10.3102/0002831215584621

Gregori, E. B., Zhang, J., Galván-Fernández, C., & de Asís Fernández-Navarro, F. (2018). Learner support in MOOCs: Identifying variables linked to completion. *Computers & Education*, *122*, 153–168. https://doi.org/10.1016/j.compedu.2018.03.014

Guàrdia, L., Maina, M., & Sangrà, A. (2013). MOOC design principles: A pedagogical approach from the learner's perspective. *eLearning Papers, 33*. https://r-libre.teluq.ca/596/

Guo, P., & Reinecke, K. (2014, March 4–5).Demographic differences in how students navigate through MOOCs. In A. Fox, M. A. Hearst, & M. T. H. Chi (Eds.), *Proceedings of the First ACM conference on Learning @ Scale* (pp. 21–30). Association for Computing Machinery. https://doi.org/10.1145/2556325.2566247

Haggard, S., Brown, S., Mills, R., Tait, A., Warburton, S., Lawton, W., & Angulo, T. (2013). *The maturing of the MOOC: Literature review of massive open online courses and other forms of online distance learning*. Department for Business, Innovation and Skills. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/240193/13-1173-maturing-of-the-mooc.pdf

Hall, R. (2015). The implications of Autonomist Marxism for research and practice in education and technology. *Learning, Media and Technology*, *40*(1), 106–122. https://doi.org/10.1080/17439884.2014.911189

Hartnett, M., George, A. S., & Dron, J. (2014). Exploring motivation in an online context: A case study. *Contemporary Issues in Technology and Teacher Education*, *14*(1), 31–53. https://www.learntechlib.org/primary/p/114723/

He, J., Bailey, J., Rubinstein, B. I. P., & Zhang, R. (2015, January 25–29).Identifying at-risk students in massive open online courses. In D. Gunning & P. Z. Yeh (Eds.), *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (pp. 1749–1755). AAAI Press. https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9696/9460

Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. *British Journal of Educational Technology*, *47*(2), 320–341. https://doi.org/10.1111/bjet.12235

Hollands, F. M., & Tirthali, D. (2014). *MOOCs: Expectations and reality* (ED547237). ERIC. https://files.eric.ed.gov/fulltext/ED547237.pdf

Holmes, W., Nguyen, Q., Zhang, J., Mavrikis, M., & Rienties, B. (2019). Learning analytics for learning design in online distance learning. *Distance Education*, *40*(3), 309–329. https://doi.org/10.1080/01587919.2019.1637716

Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education*, *98*, 157–168. https://doi.org/10.1016/j.compedu.2016.03.016

Hood, N., Littlejohn, A., & Milligan, C. (2015). Context counts: How learners' contexts influence learning in a MOOC. *Computers & Education*, *91*, 83–91. https://doi.org/10.1016/j.compedu.2015.10.019

Jacobs, A. J. (2013, April 20). Two cheers for Web U. *New York Times*. https://www.nytimes.com/2013/04/21/opinion/sunday/grading-the-mooc-university.html?pagewanted=all&_r=0

Jeno, L. M., Grytnes, J. A., & Vandvik, V. (2017). The effect of a mobile-application tool on biology students' motivation and achievement in species identification: A Self-Determination Theory perspective. *Computers & Education*, *107*, 1–12. https://doi.org/10.1016/j.compedu.2016.12.011

Jiang, S., Williams, A.E., Schenke, K., Warschauer, M., O'Dowd, D. (2014, July 4–7).Predicting MOOC performance with Week 1 behavior. In J. Stamper, Z. Perdos, M. Marvrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 273–275). EDM. http://educationaldatamining.org/EDM2014/uploads/procs2014/short%20papers/273_EDM-2014-Short.pdf

Jones, C. (2015). Openness, technologies, business models and austerity. *Learning, Media and Technology*, *40*(3), 328–349. https://doi.org/10.1080/17439884.2015.1051307

Joo, Y. J., So, H. J., & Kim, N. H. (2018). Examination of relationships among students' self-determination, technology acceptance, satisfaction, and continuance intention to use K-MOOCs. *Computers & Education*, *122*, 260–272. https://doi.org/10.1016/j.compedu.2018.01.003

Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *The International Review of Research in Open and Distributed Learning*, *16*(3), 342–358. https://doi.org/10.19173/irrodl.v16i3.2112

Kim, T. D., Yang, M. Y., Bae, J., Min, B. A., Lee, I., & Kim, J. (2017). Escape from infinite freedom: Effects of constraining user freedom on the prevention of dropout in an online learning context. *Computers in Human Behavior*, *66*, 217–231. https://doi.org/10.1016/j.chb.2016.09.019

Kizilcec, R. F., & Halawa, S. (2015, March 14–18).Attrition and achievement gaps in online learning. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale* (pp. 57–66). Association for Computing Machinery. https://doi.org/10.1145/2724660.2724680

Kizilcec, R. F., Piech, C., & Schneider, E. (2013, April 8–12). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 170–179). ACM. https://doi.org/10.1145/2460296.2460330

Kleinbaum, D. G., & Klein, M. (2005). *Survival analysis* (2nd ed.). Springer.

Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. In C. Rose & G. Siemens (Eds.), *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs* (pp. 60–65). Association for Computational Linguistics. http://doi.org/10.3115/v1/W14-41

Kucirkova, N., & Littleton, K. (2017). Digital learning hubs: theoretical and practical ideas for innovating massive open online courses. *Learning, Media and Technology*, *42*(3), 324–330. https://doi.org/10.1080/17439884.2015.1054835

Lackner, E., Ebner, M., & Khalil, M. (2015). MOOCs as granular systems: Design patterns to foster participant activity. *eLearning Papers*, *42*(3), 28–37. https://graz.pure.elsevier.com/files/3217524/Design_Patterns_for_Open_Online_Teaching_and_Learning_In_Depth_42_3_1_.pdf

Lackner, E., Zimmermann, C., & Ebner, M. (2017). A case study on narrative structures in instructional MOOC designs. *Journal of Research in Innovative* Teaching & Learning, *10*(1), 48–62. https://doi.org/DOI 10.1108/JRIT-09-2016-0005

Laurillard, D. (2002). *Rethinking university teaching: A conversational framework for the effective use of learning technologies* (2nd ed.) RoutledgeFalmer. https://doi.org/10.4324/9781315012940

Lee, Y. (2018). Effect of uninterrupted time-on-task on students' success in massive open online courses (MOOCs). *Computers in Human Behavior*, *86*, 174–180. https://doi.org/10.1016/j.chb.2018.04.043

Li, B., Wang, X., & Tan, S. C. (2018). What makes MOOC users persist in completing MOOCs? A perspective from network externalities and human factors. *Computers in Human Behavior*, *85*, 385–395. https://doi.org/10.1016/j.chb.2018.04.028

Lin, Y. L., Lin, H. W., & Hung, T. T. (2015). Value hierarchy for massive open online courses. *Computers in Human Behavior*, *53*, 408–418. https://doi.org/10.1016/j.chb.2015.07.006

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education*, *29*, 40–48. https://doi.org/10.1016/j.iheduc.2015.12.003

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in massive open online courses. *Computers in Human Behavior*, *80*, 179–196. https://doi.org/10.1016/j.chb.2017.11.011

Martinez-Lopez, R., Yot, C., Tuovila, I., & Perera-Rodríguez, V. H. (2017). Online self-regulated learning questionnaire in a Russian MOOC. *Computers in Human Behavior*, *75*, 966–974. https://doi.org/10.1016/j.chb.2017.06.015

McCardle, L., & Hadwin, A. F. (2015). Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. *Metacognition and Learning*, *10*(1), 43–75. https://doi.org/10.1007/s11409-014-9132-0

Milligan, C., & Littlejohn, A. (2014). Supporting professional learning in a massive open online course. *The International Review of Research in Open and Distance Learning*, *15*(5), 197–213. https://doi.org/10.19173/irrodl.v15i5.1855

Milligan, C., Littlejohn, A., & Margaryan, A. (2013). Patterns of engagement in connectivist MOOCs. *Journal of Online Learning & Teaching*, *9*(2), 149–159. http://jolt.merlot.org/vol9no2/milligan_0613.htm

Moe, R. (2015). OER as online edutainment resources: a critical look at open content, branded content, and how both affect the OER movement. *Learning, Media and Technology*, *40*(3), 350–364. https://doi.org/10.1080/17439884.2015.1029942

Papamitsiou, Z., & Economides, A. A. (2019). Exploring autonomous learning capacity from a self-regulated learning perspective using learning analytics. *British Journal of Educational Technology*. *50*(6). 3138–3155. https://doi.org/10.1111/bjet.12747

Peng, D., & Aggarwal, G. (2015). Modeling MOOC dropouts. *entropy*, *10*(114), 1–5. http://cs229.stanford.edu/proj2015/235_report.pdf

Pellas, N. (2014). The influence of computer self-efficacy, metacognitive self-regulation and self-esteem on student engagement in online learning programs: Evidence from the virtual world of Second Life. *Computers in Human Behavior*, *35*, 157–170. https://doi.org/10.1016/j.chb.2014.02.048

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (1st ed., pp. 451–502). Academic Press.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*(4), 667–686. https://doi.org/10.1037/0022-0663.95.4.667

Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, *6*(1), 65–78. https://doi.org/10.1007/978-1-349-62965-7_12

Reeve, J., Ryan, R., Deci, E. L., & Jang, H. (2008). Understanding and promoting autonomous self-regulation: A self-determination theory perspective. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 223–244). Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9780203831076

Rice, J. (2014). MOOCversations: commonplaces as argument. In S. D. Krause & C. Lowe (Eds.), *Invasion of the MOOCs: Promises and peril of massive open online courses* (pp. 86–97). Parlor Press. http://www.parlorpress.com/pdf/invasion_of_the_moocs.pdf#page=101

Rieber, L. P. (2017). Participation patterns in a massive open online course (MOOC) about statistics. *British Journal of Educational Technology*, *48*(6), 1295–1304. https://doi.org/10.1111/bjet.12504

Rivers, K., & Koedinger, K. R. (2013, July 13). Automatic generation of programming feedback: A data-driven approach. In N. Le, K. E. Boyer, B. Chaudhry, B. D. Eugenio, S. I. Hsiao, & L. A. Sudol-DeLyser (Eds.), *Proceedings of AIEDCS 2013: The First Workshop on AI-supported Education for Computer Science* (pp. 50–59). AIED. http://ceur-ws.org/Vol-1009/aied2013ws_volume9.pdf

Rivers, K., & Koedinger, K. R. (2014, June 5–9).Automating hint generation with solution space path construction. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Proceedings of ITS 2014: The 12th International Conference on Intelligent Tutoring Systems* (pp. 329–339). Springer. http://doi.org/10.1007/978-3-319-07221-0

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., & Gould, R. R. (2010). The astronomy and space science concept inventory: Development and validation of assessment instruments aligned with the K–12 National Science Standards. *Astronomy Education Review, 8* (1), 010111-1–010111-26. https://doi.org/10.3847/AER2009024

Salmon, G., Pechenkina, E., Chase, A. M., & Ross, B. (2017). Designing Massive Open Online Courses to take account of participant motivations and expectations. *British Journal of Educational Technology*, *48*(6), 1284–1294. https://doi.org/10.1111/bjet.12497

Sanchez-Gordon, S., & Luján-Mora, S. (2016). Barreras y estrategias de utilización de los MOOC [Barriers to and strategies in using MOOCs]. In H. P. Gómez, G. B. Alba, & M. L. Carlos (Eds.), *La cultura de los MOOC para la innovación en educación superior desde contextos iberoamericanos* (pp. 141–160). Editorial Síntesis.

Schophuizen, M., Kreijns, K., Stoyanov, S., & Kalz, M. (2018). Eliciting the challenges and opportunities organizations face when delivering open online education: A group-concept mapping study. *The Internet and Higher Education*, *36*, 1–12. https://doi.org/10.1016/j.iheduc.2017.08.002

Shao, Z. (2018). Examining the impact mechanism of social psychological motivations on individuals' continuance intention of MOOCs: The moderating effect of gender. *Internet Research*, *28*(1), 232–250. https://doi.org/10.1108/IntR-11-2016-0335

Shapiro, H. B., Lee, C. H., Roth, N. E. W., Li, K., Çetinkaya-Rundel, M., & Canelas, D. A. (2017). Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers. *Computers & Education*, *110*, 35–50. https://doi.org/10.1016/j.compedu.2017.03.003

Siemens, G. (2012, July 25). MOOCs are really a platform. *eLearnspace*. http://www.elearnspace.org/blog/2012/07/25/moocs-are-really-a-platform/

Siemens, G. (2013, March 10). Group work advice for MOOC providers. *eLearnspace*. http://www.elearnspace.org/blog/2013/03/10/group-work-advice-for-mooc-providers/

Sun, J. C.-Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British Journal of Educational Technology*, *43*(2), 191–204. https://doi.org/10.1111/j.1467-8535.2010.01157.x

Sun, Y., Ni, L., Zhao, Y., Shen, X. L., & Wang, N. (2018). Understanding students' engagement in MOOCs: An integration of self-determination theory and theory of relationship quality. *British Journal of Educational Technology*, *50*(6), 3156–3174. https://doi.org/10.1111/bjet.12724

Terras, M. M., & Ramsay, J. (2015). Massive open online courses (MOOCs): Insights and challenges from a psychological perspective. *British Journal of Educational Technology*, *46*(3), 472–487. https://doi.org/10.1111/bjet.12274

Thompson, K. (2003). *Storytelling in film and television*. Harvard University Press.

Tsai, C.-W., Shen, P.-D., & Tsai, M.-C. (2011). Developing an appropriate design of blended learning with web-enabled self-regulated learning to enhance students' learning and thoughts regarding on- line learning. *Behaviour & Information Technology*, *30*(2), 261–271. https://doi.org/10.1080/0144929X.2010.51435

Tukiainen, M., & Mönkkönen, E. (2002, June 18–21).Programming aptitude testing as a prediction of learning to program. In J. Kuljis, L. Baldwin, & R. Scoble (Eds.), *Proceedings – Psychology of*

Programming Interest Group 14 (pp. 45–57). Psychology of Programming Interest Group. http://www.ppig.org/sites/ppig.org/files/2002-PPIG-14th-tukiainen.pdf

van de Oudeweetering, K., & Agirdag, O. (2018). Demographic data of MOOC learners: Can alternative survey deliveries improve current understandings? *Computers & Education*, *122*, 169–178. https://doi.org/10.1016/j.compedu.2018.03.017

Vihavainen, A., Luukkainen, M., & Kurhila, J. (2012, October 8–10).Multi-faceted support for MOOC in programming. In R. Connolly & W. D. Armitage (Eds.), *Proceedings of the 13th annual conference on Information technology education* (pp. 171–176). Association for Computing Machinery. https://doi.org/10.1145/2380552.2380603

Wang, C.-H., Shannon, D. M., & Ross, M. E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education*, *34*(3), 302–323. https://doi.org/10.1080/01587919.2013.835779

Wang, Z., Anderson, T., Chen, L., & Barbera, E. (2017). Interaction pattern analysis in cMOOCs based on the connectivist interaction and engagement framework. *British Journal of Educational Technology*, *48*(2), 683–699. https://doi.org/10.1111/bjet.12433

Watted, A., & Barak, M. (2018). Motivating factors of MOOC completers: Comparing between university-affiliated students and general participants. *The Internet and Higher Education*, *37*, 11–20. https://doi.org/10.1016/j.iheduc.2017.12.001

Wen, M., Yang, D., & Rosé, C. P. (2014, July 4-7). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper, Z. Perdos, M. Marvrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 130–137). EDM. http://educationaldatamining.org/EDM2014/uploads/procs2014/long%20papers/130_EDM- 2014-Full.pdf

Whitmer, J., Schiorring, E., & James, P. (2014, March 24–28).Patterns of persistence: What engages students in a remedial English writing MOOC? In A. Pardo & S. D. Teasley (Eds.), *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 279–280). Association for Computing Machinery. https://doi.org/10.1145/2567574.2567601

Xiong, Y., Li, H., Kornhaber, M. L., Suen, H. K., Pursel, B., & Goins, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach. *Global Education Review*, *2*(3), 23–33. https://ger.mercy.edu/index.php/ger/article/view/124

Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013, December 9–10).*Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses* [Paper presentation]. The 2013 NIPS Data-Driven Education Workshop, Lake Tahoe, NV, United States.

Yang, Q. (2014). Students motivation in asynchronous online discussions with MOOC mode. *American Journal of Educational Research*, *2*(5), 325–330. https://doi.org/10.12691/education-2-5-13

Zheng, S., Rosson, M. B., Shih, P. C., & Carroll, J. M. (2015, March). Understanding student motivation, behaviors and perceptions in MOOCs. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1882–1895). Association for Computing Machinery. https://doi.org/10.1145/2675133.2675217

Zhu, M., Sari, A., & Lee, M. M. (2018). A systematic review of research methods and topics of the empirical MOOC literature (2014–2016). *The Internet and Higher Education*, *37*, 31–39. https://doi.org/10.1016/j.iheduc.2018.01.002