

Journal Pre-proof

Beyond one-size-fits-all in MOOCs: Variation in learning design and persistence of learners in different cultural and socioeconomic contexts

Saman Rizvi, Bart Rienties, Jekaterina Rogaten, René F. Kizilcec



PII: S0747-5632(21)00296-X

DOI: <https://doi.org/10.1016/j.chb.2021.106973>

Reference: CHB 106973

To appear in: *Computers in Human Behavior*

Received Date: 3 February 2021

Revised Date: 29 July 2021

Accepted Date: 2 August 2021

Please cite this article as: Rizvi S., Rienties B., Rogaten J. & Kizilcec René.F., Beyond one-size-fits-all in MOOCs: Variation in learning design and persistence of learners in different cultural and socioeconomic contexts, *Computers in Human Behavior* (2021), doi: <https://doi.org/10.1016/j.chb.2021.106973>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Ltd.

Credit Author Statement

Saman Rizvi: Conceptualization, Methodology, Formal Analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing

Bart Rienties: Supervision, Writing - Review & Editing

Jekaterina Rogaten: Supervision, Writing - Review & Editing

René F. Kizilcec: Supervision, Writing - Review & Editing

Beyond One-Size-Fits-All in MOOCs: Variation in Learning Design and Persistence of Learners in Different Cultural and Socioeconomic Contexts

Saman Rizvi ^a, Bart Rienties ^a, Jekaterina Rogaten ^b, René F. Kizilcec ^c

^a Institute of Educational Technology, The Open University, Milton Keynes, UK

^b University of the Arts London, UK

^c Cornell University Ithaca, NY, USA

Beyond One-Size-Fits-All in MOOCs: Variation in Learning Design and Persistence of Learners in Different Cultural and Socioeconomic Contexts

FIRST AUTHOR'S NAME, INITIALS, AND LAST NAME

First author's affiliation, an Institution with a very long name

SECOND AUTHOR'S NAME, INITIALS, AND LAST NAME

Second author's affiliation, possibly the same institution

THIRD AUTHOR'S NAME, INITIALS, AND LAST NAME

Third author's affiliation, possibly the same institution

Abstract

How successful online learners are in achieving their goals varies along geo-cultural and socioeconomic dimensions, as well as with learning design features. Despite diverse enrollments, most online courses adopt a one-size-fits-all design that presents the same learning activities to all learners. We studied how learning design can be adapted to improve learner persistence rate. We leveraged data from ten FutureLearn MOOCs (n=49,582) to examine how variation in course activities, such as articles, videos, discussions and quizzes predicted learner persistence. We then assessed the heterogeneity in these associations by learners' geo-cultural and socioeconomic context. Our findings suggest that certain types of learning activities (e.g., discussion) facilitate progress for learners in one context (e.g., Anglo-Saxon), while inhibiting progress in another (e.g., South Asia). This research contributes new insights into the role of cultural variation in learning design preferences and can inform ongoing efforts to create online learning environments that are effective for learners from diverse backgrounds.

1 INTRODUCTION

Massive Open Online Courses (MOOCs) are recognised for their goal of scaling open access to high-quality educational resources. They have been characterised as "courses designed for a large number of participants that can be accessed by anyone anywhere as long as they have an internet connection, are open to everyone without entry qualifications, and offer a full/complete course experience online, for free." (Jansen & Schuwer, 2015). Despite the lofty aspirations of granting people universal access and opportunities for success, prior research has uncovered a range of individual and contextual factors, such as regional, cultural, and socioeconomic backgrounds, that affect individual learning behaviors and outcomes. In particular, recent studies in online learning environment uncovered

gaps in learner persistence and performance based on regional and cultural differences (Bozkurt & Aydın, 2018; Ogan et al., 2015) and socioeconomic status (Cai et al., 2017; Author D, 2017; Author D, 2015; Guo & Reinecke, 2014). With an increasing number of people across the globe now learning online, these findings highlight the need to design courses that are accessible and inclusive to all, irrespective of cultural differences.

Most online courses contain various types of learning activities that are delivered uniformly and in a predetermined sequence. In an online learning environment, pedagogical factors, such as the learning design, play a pivotal role in learner engagement, retention and persistence (Mangaroska & Giannakos, 2018; D. Davis et al., 2018; Author B, 2017; Author B, 2016; Pimmer et al., 2016; Hernández-Leo et al., 2014). Features of the learning design, such as content type and activity sequencing, can be crucial determinants of persistence, but there is limited empirical evidence on the role of learning activities in learners' course experiences. Moreover, we are limited in our scientific understanding of which learning activity types should be tailored to members of diverse socioeconomic and geo-cultural groups.

In this research, we contribute new insights into how course learning designs may be adapted for diverse online learners by exploring a heterogeneous learner population enrolled in ten large MOOCs. This study establishes new links between learner persistence and the types of learning activities they experience in a course. This study clearly demonstrates that an overall analysis of online learning data can mask geo-cultural and socioeconomic heterogeneity in the relationship between learning design and learner outcomes. Consequently, the results of analyses that do not take this heterogeneity into account primarily reflect the behavioral patterns of the largest group, which can stand in contrast to the patterns observed for smaller subgroups. If overall results are used to guide course design and iterative improvement, it can result in improved outcomes for the majority group while leaving behind members of underrepresented groups in the course. Therefore, this study critically, empirically examines the premise of one-size-fits-all learning design by simultaneously unpacking the complex interplay between online learning design and learners' cultural and socioeconomic diversity.

2 RELATED WORK

2.1 Relationship between learning designs and learner's persistence

A number of studies on formal online environments (Author B, 2017; Author B, 2017; Author B, 2016; Conde & Hernández-García, 2015; Hernández-Leo et al., 2014) and MOOC environments (Author A, 2020; Xing, 2019) show the critical role of learning design. While learning design can be defined in several ways, a better understanding of the nature of engagement with various learning activity types may lead to ways which we can use to improve persistence. Several other studies have also suggested that learning engagement can be better understood by examining the detailed interactions with various types of learning activities, such as reading material (Author A, 2020; Uchidiuno et al., 2018), video content (D. J. Davis, 2019; Guo et al., 2014), quizzes (Juhaňák et al., 2017; Li & Baker, 2018), and discussions (Zou et al., 2021; Allon et al., 2016; Sunar et al., 2016; Yang et al., 2013).

This study conceptualises learning design as the process of online course development where the designers create a series of learner-facing activities containing different types of learning materials (e.g., reading activities, learning material consisting audios or videos) (Author A, 2020; Sharples, 2015). Such sequenced activities are designed for learning, and therefore can be reused repeatedly, but with minimal possibility of modification, midway or between the course runs. A prepared design may have its own advantages. Still, recent literature suggests that a centralised learning design, containing a fixed number and types of activities may not necessarily be beneficial for

all learners, as it can be more useful for some learners while limiting others (Bearman et al., 2020; Margaryan et al., 2015). It is also not known how different proportions of these learning design factors enable or restrict learners' participation.

Cross et al. classified learning activities and established a taxonomy of learning design called OULDI (Open University Learning Design Initiative) (Cross et al., 2012). This taxonomy provides theoretical foundations for designing learning activities in FutureLearn MOOCs (Sharples, 2015). FutureLearn MOOC design primarily contains four types of learning activities; articles, videos, discussions, and quizzes (Sharples, 2015) and each type has its own advantages and limitations. This study specifically focused on three OULDI categories that represent the principal features in FutureLearn MOOCs. First, Assimilative activities refer to learning activities that help develop, process, and attain information in an online course (e.g., reading articles or watching videos). Previously in his seminal work on multimedia learning, Mayer (2005) asserted how effective learning occurs when supported by multimedia, for example, words and visuals (i.e., article) and instructional videos. Second, Communication-based activities allow learners to participate in course-related discussions. In FutureLearn MOOCs, discussion steps mimic a social media-style feed where learners can choose to start, like or comment on a discussion, or follow their peers and instructors (Sharples, 2015). FutureLearn is philosophically built on socio-constructivist principles and the idea that working together collaboratively drives learning. Therefore, discussions are embedded explicitly in all courses making the MOOC a *social learning space* (Manathunga et al., 2017). The third part of our work focused on Assessments (Quizzes) – another known driver of learning. Assessment activities and the associated feedback facilitate knowledge acquisition and learning successes (e.g., Hattie & Timperley, 2007). Increased skills and learning successes, in turn, can motivate learners to persist in the course (e.g., Marsh & Craven, 2006).

Therefore, in our first research question, we intend to understand if changing the number of various types of learning activities (articles, videos, discussions, and quizzes) is linked with improved persistence and lower withdrawal from a MOOC.

RQ1. To what degree does the number of (a) assimilative learning activities (e.g. articles, videos), (b) communication activities (e.g. discussions), and (c) assessment activities (e.g. quizzes) in a course predict learners' persistence?

2.2 Learning behaviour and geo-cultural or socioeconomic background

Culture, often defined as '*collective complexes of learned behaviours and perceptions of individuals in a society*' (Tylor, 1871), influences how an individual processes and interprets information. Therefore, education cannot remain disconnected from the cultural values of a society. The term 'culture' itself is multifaceted. The related term "geo-culture" (Wallerstein, 1991) is often employed in theoretical frameworks to explain how people act, behave and learn based on their geographical location or area of origin. Previous research has found the geo-cultural framework by House and colleagues (House et al., 2004) to be useful for understanding how learners from different geo-cultural regions on the globe engage with learning activities (discussed further in section 3.1). Recent work has also found that in various regions the originating culture strongly influences the respective design and content of the curriculum as well as formal classroom practices (Bozkurt & Aydin, 2018 ; Arnove et al., 2012). This research also suggested that learners' experiences and reactions to their educational practices may be grounded in and shaped by their culture and geographic region. Regional and cultural differences in online learner persistence have also been reported (Reich & Ruipérez-Valiente, 2019; Author A, 2019; Author B, 2015).

Initially, MOOCs, there was a strong expectation amongst some in the online learning and teaching community that these free, widely advertised online courses would potentially address the global disparity in education attainment (Rohs & Ganz, 2015; Pawlowski & Hoel, 2012). However, it turned out that most popular MOOC providers and the majority of active learners are currently located in developed countries, mainly in the Global North (Ruipérez-Valiente, Jenner, et al., 2020; Author D, 2017; Rohs & Ganz, 2015), with evidence suggesting that regional, cultural and socioeconomic factors influence engagement and persistence in online learning (Author A, 2019; Reich & Ruipérez-Valiente, 2019; Allione & Stein, 2016; Guo & Reinecke, 2014). Several studies reported disproportional participation and subsequent preference for particular types of learning activities among MOOC learners. Author D (Author D, 2015) found that learners from African, Asian and Latin American countries are less persistent accessing resources and exam materials. Liu and colleagues (Liu et al., 2016) also noticed a distinction between mainly English-speaking countries and Asian countries in their test-taking behaviour and discussion forum participation. Likewise, other researchers found similar evidence for significant disparities resulting from attributes such as ethnicity and race (Stich & Reeves, 2017; Wladis et al., 2015), average neighborhood household income or the national Human Development Index (Author D, 2017; Hansen & Reich, 2015), and levels of social integration that vary between different cultural groups (Author D, 2017; Liu et al., 2016). Pointing to this global divide, Author D (Author D, 2017) observed a "*wide MOOC retention and completion gap between less developed countries (LDC) and more developed countries (MDC)*".

The narrative was echoed in recent large-scale studies investigating the global participation in MOOCs where it was noted that most MOOC enrolments, and thereof certifications, have consistently emerged from the most affluent countries with much lower participation from less prosperous countries in the Global South (Ruipérez-Valiente, Halawa, et al., 2020; Ruipérez-Valiente, Jenner, et al., 2020; Reich & Ruipérez-Valiente, 2019). Other financial, psychological or social barriers were also repeatedly reported in region-specific MOOC studies (Castaño-Muñoz et al., 2017; Author D, 2017; Hansen & Reich, 2015). Ruipérez et al. (Ruipérez-Valiente, Jenner, et al., 2020) recently pointed out that differences in learning designs between regional and global MOOC providers influence demographic differences between course enrolments and completion rates. In MOOC learning environments, regional, financial or socioeconomic indicators have helped researchers understand learners' course engagement, and their decisions to enrol, persist or withdraw from a course (Reich & Ruipérez-Valiente, 2019; Hansen & Reich, 2015; Wladis et al., 2015).

Previous work suggests that there are cross-cultural and regional differences in participation in various types of learning activities. Comparing several regions or cultural traits, MOOC researchers noticed disproportional assessment or quiz attempts (Liu et al., 2016; Author D, 2015), unique video watching behaviour with or without accessing assessments (Liu et al., 2016; Uchidiuno et al., 2018) and social interactions with peers (Liu et al., 2016; Ogan et al., 2015). Relevant literature pointed out that various regions may have dissimilar preferences in assimilative activities, with some favouring reading-based materials and others with a fondness for video-based content (Uchidiuno et al., 2018; Liu et al., 2016; Reinecke & Bernstein, 2011). Indeed, settling for a learning design presents a methodological challenge if learners from numerous regions are given manifold choices in learning activities. However, such methodological challenges can easily be addressed with recommender systems and other machine learning methods that perform variable selection to surface essential and relevant feature combinations.

Despite these important advances to knowledge, there is still a paucity in research investigating an overall impact of various learning activity types and its relation to broader geo-cultural background or socioeconomic status. None of the studies discussed above was conducted using FutureLearn MOOC platform, which includes

explicitly many communication-based activities, which may help or potentially hamper learner engagement from specific geo-cultural regions. This leads to our next research questions on how the association between learning activity types and learners' persistence in MOOCs is moderated by learners' geo-cultural background (RQ2) and their regions' aggregated income group (RQ3). This study examined if the strength of the predictive association (from RQ1), differs between geo-cultural and socioeconomic subgroups. In particular, we intend to explore if certain types of learning activities were an enabler for one subgroup but constrained another subgroup, leading to the following key questions:

RQ2. To what extent does the association between course learning design and learner persistence differ between geo-cultural contexts?

RQ3. To what extent does the association between course learning design and learner persistence differ between socioeconomic contexts?

3 METHODOLOGY

3.1 Context and data pre-processing

Our data set comprises 49,582 learners from ten MOOCs developed by the Open University, UK and offered via the FutureLearn platform. The courses were structured into steps, activities, and weeks. Each week contains several steps comprising various activities. Such activities can be categorised using OULDI taxonomy (assimilative activities like articles and videos, communication activities like discussions and assessment activities like quizzes). Table 1 lists details of four learning design features in each course: Articles (A), Discussions (D), Videos (V), Quizzes (Q). For example, in the course aml1, 1564 learners were enrolled, whereby 1164 (or 74.4%) learners accessed more than 1% of activities. For those who accessed more than 1% of activities, median activity access was found to be 8% activities. The learning design contained 32 articles, 25 discussions, 27 videos and 24 quizzes.

In this study, we only consider open and freely available assessment Quizzes because participants are expected to pay a fee to access formal assessments (i.e., Tests). We include in our sample only learners with free and limited-time access to course content. Given our core focus on open education, we only included activities that were free to access for FutureLearn learners, because the certification fee can be prohibitively expensive for some learners. These sample exclusion criteria might have introduced selection bias. Still, the criteria were aligned with the research questions that examined the progress of learners through the open and freely accessible course material. Other sporadic miscellaneous learning activities were also omitted in this study, that includes activities such as exercises, peer reviews or composite steps consisting of assignment, assignment review and reflect.

➔ **Insert Table 1 about here**

On the FutureLearn platform, the full content of a course is made available to all learners at the beginning of the course. It remains accessible for a period called "extended enrolment", which is around two weeks. It is recommended that learners engage with the course for 3 to 6 hours per week, watch videos, read course materials, participate in the discussions, and complete quizzes. Since the full course contents were made available beforehand,

the number of learning activities remained constant throughout the course (i.e., time-invariant covariate). As illustrated in Table 1, the courses were diverse and largely dissimilar and were selected on the basis of variability in the learning design factors (types of learning activities).

The GLOBE framework for learners' geo-cultural categorisation (see for example (Mensah & Chen, 2013; House et al., 2004)) represents a comprehensive account of global regions and cultural constructs in those regions. Following the precedence set by many other studies of geographical and cultural variation in the context of MOOCs (e.g., Reich & Ruipérez-Valiente, 2019; Bozkurt & Aydın, 2018; Hennis et al., 2016), we classified learners' geo-cultural and socioeconomic backgrounds based upon their IP address. Using GLOBE, we first categorised learners' location into ten geo-cultural clusters; Sub-Saharan Africa (AF), Anglo-Saxon (AS), Confucian Asia (CA), Eastern Europe (EE), Germanic Europe (GE), Latin America (LA), Latin Europe (LE), Middle East (ME), Nordic Europe (NE), and South Asia (SA). The distribution over clusters by course is shown in Figure1 for more details. Anglo-Saxon (AS) subgroup made up a consistent majority of the learners in each course (between 15.8% to 57.2%), followed by learners from South Asia (SA) which varied between 10.2% to 31.8%. Whereas, the smallest subgroup was found to be Nordic European (NE) learners comprising only 0.5% to 1.6% learners.

➔ **Insert Figure 1 about here**

Next, using the world bank classification of world economies into different income categories (*World Bank Country and Lending Groups – World Bank Data Help Desk*, 2017), we aggregated learners' countries into four socioeconomic subgroups; high income (HI), upper-middle income (UMI), lower middle income (LMI), and low income (LI). Figure1 illustrates the distribution. In line with (Reich & Ruipérez-Valiente, 2019), the largest stratum (33.6% to 71.4%) originated from the flourishing economic subgroup of High-Income (HI) countries. The second-largest subgroup (17.2% to 47.3%) though belonged to relatively underprivileged countries with Lower Middle-Income (LMI) economies.

3.2 Measures

We used learners' persistence in a MOOC as the outcome measure, assessed based on how far in the course materials a learner progressed before they dropped out, defined as ceasing participation in a course. We calculated the percentage of activities accessed to operationalise persistence (% activities accessed). To predict the outcome, we defined a number of predictors or covariate measures. First, the four learning design predictors were the number of articles, videos, discussions, and quizzes in each course. These covariates contained ten data points, each representing one of the ten MOOCs. We first used cultural clusters (cc) and socioeconomic clusters (sec) to subset data. Next, we used these constructs as covariates within the regression equation to measure interactions. Following is the list of covariates we used in this study: (i) *Number of learning design (LD) features*: Articles (nA), Videos (nV), Discussions (nD), and Quizzes (nQ), (ii) *cc*: representing ten cultural subgroups, and (iii) *sec*: representing four socioeconomic subgroups.

3.3 Data Analysis Methods

We used survival analysis methods to measure the effect of the number of different types of learning activities on learner persistence (Labrador et al., 2019; Yang et al., 2013). Survival analysis, sometimes referred to as Time-to-event (TTE) analysis, is a useful statistical tool popular in disciplines such as medical research, engineering and

economics. For example, this method can be used to evaluate new drug and its dosage efficacy, assess actuarial losses, and reliability and life of devices under different environmental factors (like temperature or humidity). The factors associated within the ten MOOCs of the respective environment (number of articles [A], videos [V], quizzes [Q] and discussions [D]) were expected to influence learners' persistence. Unlike summative measure of performance (i.e., certification or course completion), persistence is temporal in nature where learners' dropout, (thereof hazard to the survival) occur as the course progresses.

Prior to the analysis, functional forms of all four learning design features were checked, and we found satisfactory evidence for linear assumptions. Thus, we assumed the relationship between learning design features and log hazard of early dropout (or hazard to the persistence) were linear in nature. During the survival analysis in this study, the measure of *event of interest* represented the respective event of learners leaving the course after accessing a certain fraction of activities. As indicated in Table 1, to perform the survival analysis, we defined the occurrence of an *event of interest* if a learner accessed at least 1% of course activities and then stopped interaction with the content and left the course (event = 1 when accessed activities >1%, otherwise 0). Table 1 also listed medians of per cent activity accessed by learners who had accessed at least 1% activities.

Firstly, the Kaplan-Meier (KM) curves were used to examine survival probabilities and median survivals. As is common in KM analysis, we assumed that the survival probabilities remained the same for the learners regardless of their recruitment time, i.e., whether they started early or late in the course. Secondly, the complementary log-rank tests were employed to test the differences between survivals. Both methods are non-parametric and make no assumptions about the distribution of the data. These two methods yield descriptive statistics that address questions such as what proportion of learners will continue learning in their respective MOOCs after a certain point in the course. Finally, we used semi-parametric Cox regression for a more nuanced, multivariate analysis (Fox, 2002).

Cox regression was used to quantify the impact of the different number of learning activities on the percentage of activities accessed by a learner (RQ1). Following, we checked if the degree of influence was different for the different geo-cultural subgroups (RQ2) and socioeconomic subgroups (RQ3). A large number of two-way interactions between the dimensions of interest (LD*cc, and LD*sec) raised a potential risk of finding spurious relationships. Since the interactions that significantly improve model fitting should be preserved in the model, we decided to use the regularisation technique of penalised regression. Table 4 and 5 in the results section highlight the most critical interactions retrieved from the regularised, ten-fold cross-validated outcome (Hastie & Qian, 2014).

➔ **Insert Table 2 about here**

The analysis for RQ1 provided overarching results but at the same time, masked heterogeneity in our diverse sample. Therefore, to answer RQ2 and RQ3, we first grouped data into ten subsets based upon cultural clusters (cc) and performed subgroup analysis on each subset using Cox regression. The method was repeated on four subgroups for socioeconomic clusters (sec). Developing a separate model for each subgroup enabled us to compare or contrast the findings with the overarching model and across the subgroups. It is important to highlight here that adopting a separate model for each subgroup assumed the subgroups to be entirely independent of each other. Next, we used two-way interaction terms within the regression equation to understand the joint effect of the interacting variables on the persistence hazard profile. This twice-over analysis approach helped us avoid overinterpretation of

potentially noisy subgroup results (Lagakos, 2006). The methods for answering each RQ are summarised in Table 2.

4 RESULTS

RQ1: Association between number of learning activities and persistence

We first confirmed a sufficient degree of variation in our outcome of interest, learners' persistence, across the ten courses ($\chi^2=882$, $df=9$, $p=<2e-16$). Next, we used cox regression to test for a quantifiable link between persistence and predetermined types of activities in the respective learning design (see Table 3). The assumption of proportionality of hazards (PH) was only mildly violated for variable nV ($0.33 > p > 0.02$ in covariate-adjusted Cox regression).

➔ Insert Table 3 about here

In Table 3, $\exp(\text{coef})$ represents the respective hazard ratio for each predictor. Overall, a higher number of learning activities in a course design was associated with an elevated risk of dropout, except for the number of discussion activities. We found that, while keeping all other activities constant, one unit¹ increase in number of discussion-based activities reduced the dropout risk by 3%. Similarly, increasing reading activities and quizzes by a unit, augmented the risk of dropout by 14% and 15% respectively. Intuitively speaking, Table 1 points to relatively low median activity access for the courses with a slightly higher number of articles and quizzes. The link between persistence and the change in the number of videos was minimal; about a 3% increase in risk of dropout from the course. In other words, a large number of articles and quizzes in a course was associated with an elevated hazard profile, while this association was reversed for discussions.

RQ2: Is the association (from RQ1) different in different geo-cultural subgroups.

Building on RQ1, we first confirmed that the survival experiences were dissimilar for different geo-cultural groups ($\chi^2=542$, $df=9$, $p=<2e-16$). As shown in Figure 1 (and Figure A1 in the Appendix), the Anglo-Saxon subgroup (AS) had the largest presence overall and exhibited the highest median persistence and survival probabilities. In contrast, the second and third largest subgroups, respectively, South Asia (SA) and the Middle East (ME), turned out to be least persistent throughout. Both subgroups had the smallest survival probabilities with 7% and 5% of learners, respectively, reaching the 50% activity access mark. Furthermore, only 3% and 2% respectively accessed all activities in the courses. (Figure A1 in the Appendix).

Addressing RQ2, Table 4 compares estimated coefficients from (i) subset analysis and (ii) two-way interaction tests, while we computed hazard ratios by exponentiating these estimates. During the subgroup analysis, the PH assumption was only mildly violated at three points among all 40 points of estimates. A mild violation in PH assumption is common when survival analysis is performed on big data, as the sensitive PH test was originally developed with much smaller datasets in mind. The second column for each activity type reports coefficients from

¹ We standardized the predictors, therefore, one unit represents 1SD for the respective predictor.

one full model where we interacted geo-cultural subgroups with learning design factors. The results from both types of examinations remained primarily the same.

➔ **Insert Table 4 about here**

Since the hazard ratios were computed by exponentiating the respective parameter estimates, for AF (see row 1, column 4 in Table 4), this result was $\exp(0.07) = 1.07$. In terms of interpretation, this points to a 7% increase in the expected hazard for every one-unit increase in number of articles given learner subgroup is AF. Overall, the outcomes suggested that keeping all other activities constant, a one-unit increase in articles meant an increased dropout risk of 7%, 28%, and 48% for African (AF), Anglo-Saxon (AS), and Latin American (LA) learners, respectively. However, this direction of association was the same for Germanic and Latin European (GE, LE) subgroups, with no significant cross-validated interactions. Surprisingly, the only meaningful interaction between cultural grouping and video-based activities we found was for Middle Eastern (ME) learners, where risk increased by 9% for one unit increase in video-based activities, holding other activities constant. Nevertheless, a relatively less critical interaction term, increasing the videos, lowered the hazard profile for South Asian learners (by almost 7%).

Discussion-based learning activities were found to be the most critical activity type, impacting various geo-cultural cluster differently. An increasing number of discussion steps were associated with a reduction in dropout risk ratio, except for the learners residing in South Asia and Africa. One unit increase in discussions reduced the risk of dropout for Latin American learners by almost 12%. In contrast, it elevated the risk to persistence by 21% for South Asian learners and nearly 9% for African learners. However, quiz-based assessments deterred more English-Speaking learners (Anglo-Saxon region) and Middle Eastern learners with an increment of 26% and 7% in risk profile.

➔ **Insert Figure 2 about here**

A subset analysis was most meaningful when we found that the potential change in the number of learning activities led to precisely the opposite results for the two largest geo-cultural subgroups in data; Anglo-Saxon (AS) and South Asia (SA). Except for discussion steps, the South Asian learners favoured a large number of learning activities overall. Furthermore, mirroring the overarching sample behaviour (Table 3), Anglo-Saxon learners preferred fewer activities overall; more discussion steps kept them engaged, though. Figure 2 illustrates this contrast in learning activity preferences.

RQ3: Is the association (from RQ1) different in different socioeconomic subgroups.

We repeated the same modelling steps for the four socioeconomic subgroups. Figure 1 (right) demonstrates the number of learners in each socioeconomic subgroup, and their estimated median activity access in percentages is

reported in Appendix Figure 2. High Income (HI) subgroup accessed approximately 3% to 4% more activities before leaving the course than their peers from Low-Income (LI) subgroup. Figure 2. in Appendix A, illustrates KM survival curves for the four socioeconomic subgroups. A log-rank test indicated significant differences between survival experiences ($\chi^2=320$, $df=3$, $p < 2e-16$). Each socioeconomic subgroup exhibited a distinct course engagement behaviour. The relatively deprived subgroup (LMI) showed the smallest survival probability, when only 7% of LMI learners made it to 50% activity access mark, as compared to 13% of learners from the High-Income (HI) subgroup. Displaying the consistent pattern, the fraction of these subgroups fell to 3% and 6%, respectively, towards the end of the course (See appendix A for more details).

➔ **Insert Table 5 about here**

As a result of the cross-validated analysis, we found no critical interaction in the smallest subgroup of Low-Income (LI) economies. Also, as can be seen from Table 5, a large number of learners (more than 36%) from this subgroup dropped out after accessing just one activity, as compared to 25% from the High-Income (HI) subgroup. The High-Income (HI) subgroup not only accounted for the majority of learners in our sample, but it also reflected the overall content engagement behaviour we noticed while answering RQ1 (Table 3). It is important to note here that considering the cultural context, the High-Income group of countries mainly belonged to the English-speaking Anglo-Saxon, Germanic, or Latin European cultural groups.

Increasing one unit in the number of articles increased the dropout risk for the High-Income subgroup by almost 25%. A change in the number of articles as well as videos was not significantly linked with any of the subgroups as we found no active interaction. As learners from advanced and wealthy countries preferred more discussion steps in the course, the same activity type increased learner's dropout risk for the second largest subgroup of Lower-Middle Income (LMI), by almost 13%. Whereas, learners from the economically affluent High Income (HI) and Upper-Middle Income (UMI) countries did not favour quizzes. Finally, the quiz-based learning activities did not interact with the other two less prosperous subgroups.

Figure 3 illustrates the contrasting behaviour of the two largest socioeconomic subgroups. In terms of geo-cultural grouping, Lower-Middle Income mainly consisted of countries from South Asian, African, and Middle Eastern regions. In other words, learners from wealthy, Anglo-Saxon countries or most European countries were repelled by a large number of articles or quizzes in a course. As for Lower-Middle Income (LMI) economies, the only significant link (from cross-validation) activity change we noticed was for discussions where a unit increase in discussion activities was found to be linked with a 13% increase in early dropout risk.

➔ **Insert Figure 3 about here**

5 DISCUSSION AND CONCLUSION

We set out to find the link between the persistence of 49,582 learners and four learning design factors in ten FutureLearn MOOCs developed in the UK. We also examined to what extent the learner persistence probabilities differed between socioeconomic subgroups or between geo-cultural subgroups. In this study, our aim was twofold. First, we examined whether a different number of assimilative activities (like articles and videos), communication activities (like discussions), and assessment activities (like quizzes) within a MOOC could be used to predict the

overall number of activities learners from across the globe access? Second, we compared the behavioural differences between ten geo-cultural clusters and four socioeconomic clusters. To address our first research question, we investigated the extent to which the number of various types of learning activities is linked with higher persistence in a MOOC. For our second and third research questions, we measured the variation in these associations in the contexts of geo-cultural and socioeconomic subgroups. For example, we found that with the only exception of learners residing in South Asia, learners overall favoured fewer activities in MOOCs. Earlier work suggested that communication-based learning activities are critical for success in a course (Zou et al., 2021; Author B, 2016). Still, we found that they may actually work against learners from disadvantaged, non-English speaking contexts originating from the Global South (LMI socioeconomic subgroup and geo-cultural subgroups such as SA and AF).

Concerning variations in persistence with respect to activity types, we found the following outcomes. Learners engaged heavily in courses with a large number of assimilative activities (articles and videos), specifically those with more videos. Since all the content was made available in the English language, one would expect that Anglo-Saxon learners will be slightly more interested in reading articles. However, a large number of reading-based material in a course was disfavored by most learners, surprisingly enough, even by learners residing in English speaking countries.

Against the general belief that learning videos should be the central learning feature in MOOC learning designs, we found an increasing number of videos to be positively linked with retention of South Asian learners only. The finding was consistent with previous research (Liu et al., 2016; Reinecke & Bernstein, 2011) that observed a preference for videos or visuals over text, particularly in collectivists countries such as those from South Asian region. But collectively, the growing number of video-based activities was slightly unfavourable to the rest of the regions, both collectivists and individualists.

In line with others (Liu et al., 2016; Ogan et al., 2015), we found that the FutureLearn platform's underlying pedagogy, '*education in a social learning space*', (Manathunga et al., 2017) worked better for the largest subgroup of western learners, who primarily belonged to English speaking or European countries with affluent economies. However, the same discussion-based activities were found to be constraining countries with limited resources and lower-middle-income level: South Asian or African regions. Open discussion of ideas and knowledge is a more common trait in individualistic societies with low power distance scores. This difference in cultural constructs might explain the increased risk profile of South Asian and African learners, associated with the number of discussion activities. Previous literature suggests that content in a MOOC can be influenced by the cultural values of the country where the course was produced (Bayeck & Choi, 2018). A FutureLearn MOOC developed in the UK, may probably be designed primarily with UK-based learners in mind (either explicitly or implicitly), while assuming that large number of discussion-based activities will facilitate learners in general.

Through MOOCs, learners preferred to be quizzed in moderation. A large number of quizzes led to large early dropouts, specifically for learners from European and Anglo-Saxon region – mainly upper or upper-middle-income countries. In contrast with a previous study (Author D, 2015), we found similarities in African and Latin American regions. However, we found a smaller negative association between the number of quizzes and risk to persistence in South Asian and Middle Eastern learners.

A change in the number of all four types of learning activities led to precisely the opposite results for the two largest geo-cultural subgroups in data: Anglo-Saxon (AS) and South Asian (SA) regions, as well as for the two largest socioeconomic subgroups of High Income (HI) and Lower-Middle Income (LMI) societies. This might potentially be

explained by the fact that most learners from advanced (European or English-speaking) countries may already have high income and education levels (Hansen & Reich, 2015), who would enroll in free online courses for "knowledge attainment" or merely learning for pleasure. Not interested in certification, these learners may be deterred by the increasing number of quizzes in an online course. In comparison, certification might be more important for learners from collectivist, high power distance regions such as South Asia, Africa, and the Middle East. We found either negative or a relatively small positive association between the number of quizzes and early dropout risk for these three geo-cultural clusters. Overall, our findings suggest that the result of changing types of learning activities on progress varies mainly with the context.

While measuring and reporting the subset analysis and term interactions, we considered that our study might not have covered all factors linked with learner persistence. We found no perfect recipe for the number of learning activities in any course. There is probably no ideal combination of coherent activities that work for all learners, and we found the search for inclusive design a complicated endeavour. In MOOC pedagogy and MOOC learning design, we can recommend elasticity and flexibility. Before we reach the goal of context-adaptive MOOC learning design, we suggest a balanced approach – a combination of all types of learning activities, not just video-driven, discussion-based, or reading MOOCs.

Overall, this study provides an evidence-based understanding of learner experience from different geo-cultural and socioeconomic backgrounds and relates those experiences with the proportion of course content type. This evidence has implications in developing culturally adaptive MOOC designs. Therefore, the key contribution of this research is that the findings inform about context-centred design. Indeed, by offering learners a choice to opt for such an adaptive design, we can ensure that the design is still learner-centred and not merely context-centred. This innovative research provides new insight into cultural variation in learning design preferences that inform ongoing efforts to create effective online learning environments for people from diverse backgrounds.

6 LIMITATIONS AND FUTURE WORK

All MOOCs used in this study were designed by the same learning design team at the OU, after consultation with the instructors and subject matter experts. However, the principal limitation in the approach we used is that we had no control over learning design factors as there was only one fixed, predetermined design for each MOOC. Another approach might be design-based, with a possibility of multiple learning design customisation during and between the course run. An experimental manipulation of design constructs in a course can provide causal evidence that extends beyond the correlational results we present in this study. Another limitation was that all MOOCs were developed by the same university and offered via a single platform. A multi-platform analysis may also yield an enhanced understanding. Here it is also important to highlight that learning design factors are a few of the many aspects linked with persistence.

The current analyses have treated geo-cultural clusters and socioeconomic clusters as independent covariates. However, more research is needed to understand the potential confounding impact of these two types of covariates. Further work should be undertaken to understand the three-way interaction between learning design factors, geo-cultural and socioeconomic contexts. Moreover, further research is needed to test other potentially influencing contextual and demographic factors such as learners' age or gender. Learners' trace data alone could not provide an in-depth, comprehensive understanding of learners' individual preferences. Therefore, to explore learners' choices and learning activities preferences, we recommend a follow-up analysis, preferably using qualitative and/or triangulated approaches.

REFERENCES

- Author A. (2020). [details removed for peer review].
- Author A. (2019). [details removed for peer review].
- Author B. (2017). [details removed for peer review].
- Author B. (2017). [details removed for peer review].
- Author B. (2016). [details removed for peer review].
- Author D. (2017). [details removed for peer review].
- Author D. (2017). [details removed for peer review].
- Author D. (2015). [details removed for peer review].
- 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.
- Allon, G., Van Mieghem, J., & Zhang, D. (2016). *Does Social Interaction Improve Learning Outcomes? Field Evidence from Massive Open Online Education*. The Field Experiments Website.
- Arnove, R. F., Torres, C. A., & Franz, S. (2012). *Comparative education: The dialectic of the global and the local*. Rowman & Littlefield Publishers.
- Bayeck, R. Y., & Choi, J. (2018). The Influence of National Culture on Educational Videos: The Case of MOOCs. *The International Review of Research in Open and Distributed Learning*, 19(1), Article 1.
- Bearman, M., Lambert, S., & O'Donnell, M. (2020). How a centralised approach to learning design influences students: A mixed methods study. *Higher Education Research & Development*, 1–14.
<https://doi.org/10.1080/07294360.2020.1792849>
- Bozkurt, A., & Aydin, İ. E. (2018). Cultural Diversity and Its Implications in Online Networked Learning Spaces. In *Supporting Multiculturalism in Open and Distance Learning Spaces* (pp. 56–81). IGI Global.
- Cai, Z., Fan, X., & Du, J. (2017). Gender and attitudes toward technology use: A meta-analysis. *Computers & Education*, 105, 1–13.
- Castañó-Muñoz, J., Kreijns, K., Kalz, M., & Punie, Y. (2017). Does digital competence and occupational setting influence MOOC participation? Evidence from a cross-course survey. *Journal of Computing in Higher Education*, 29(1), 28–46.

- Conde, M. Á., & Hernández-García, Á. (2015). Learning analytics for educational decision making. *Computers in Human Behavior*, 47, 1–3. <https://doi.org/10.1016/j.chb.2014.12.034>
- Cross, S., Galley, R., Brasher, A., & Weller, M. (2012). *OULDI-JISC Project Evaluation Report: The impact of new curriculum design tools and approaches on institutional process and design cultures*.
- Davis, D. J. (2019). *Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses*. <https://doi.org/10.4233/uuid:b8be8302-84a0-4b29-a6fe-761a3f872420>
- Davis, D., Seaton, D., Hauff, C., & Houben, G.-J. (2018). *Toward Large-Scale Learning Design*.
- Fox, J. (2002). Cox proportional-hazards regression for survival data. *An R and S-PLUS Companion to Applied Regression*, 2002.
- Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the First ACM Conference on Learning@ Scale Conference*, 41–50.
- Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. *Proceedings of the First ACM Conference on Learning@ Scale Conference*, 21–30.
- Hansen, J. D., & Reich, J. (2015). Socioeconomic status and MOOC enrollment: Enriching demographic information with external datasets. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, 59–63.
- Hastie, T., & Qian, J. (2014). Glmnet vignette. Retrieved June, 9(2016), 1–30.
- Hennis, T., Topolovec, S., Poquet, O., & de Vries, P. (2016). Who is the Learner: Profiling the Engineering MOOC Student. *SEFI 44th Annual Conference, Tampere, Finland*. URL: [Http://Www. Sefi. Be/Conference-2016/Papers/Open_and_Online_Engineering_Education_Engineering_Education_ Research/Hennis-Who-Is-the-Learner-108_a. Pdf](http://www.sefi.be/conference-2016/papers/Open_and_Online_Engineering_Education_Engineering_Education_Research/Hennis-Who-Is-the-Learner-108_a.Pdf).
- Hernández-Leo, D., Moreno, P., Chacón, J., & Blat, J. (2014). LdShake support for team-based learning design. *Computers in Human Behavior*, 37, 402–412.
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage publications.

- Jansen, D., & Schuwer, R. (2015). *Institutional MOOC strategies in Europe status report based on a mapping survey conducted in October—December 2014*. EADTU.
- Juhaňák, L., Zounek, J., & Rohlíková, L. (2017). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*.
- Labrador, M. M., Vargas, G. R. G., Alvarado, J., & Caicedo, M. (2019). Survival and Risk Analysis in MOOCs. *Turkish Online Journal of Distance Education*, 20(4), 149–159.
- Lagakos, S. W. (2006). The challenge of subgroup analyses-reporting without distorting. *New England Journal of Medicine*, 354(16), 1667.
- Li, Q., & Baker, R. (2018). The different relationships between engagement and outcomes across participant subgroups in Massive Open Online Courses. *Computers & Education*, 127, 41–65.
- Liu, Z., Brown, R., Lynch, C., Barnes, T., Baker, R. S., Bergner, Y., & McNamara, D. S. (2016). MOOC Learner Behaviors by Country and Culture; an Exploratory Analysis. *EDM*, 16, 127–134.
- Manathunga, K., Hernández-Leo, D., & Sharples, M. (2017). A Social Learning Space Grid for MOOCs: Exploring a FutureLearn Case. In C. Delgado Kloos, P. Jermann, M. Pérez-Sanagustín, D. T. Seaton, & S. White (Eds.), *Digital Education: Out to the World and Back to the Campus* (Vol. 10254, pp. 243–253). Springer International Publishing. https://doi.org/10.1007/978-3-319-59044-8_29
- Mangaroska, K., & Giannakos, M. N. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*.
- Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). *Computers & Education*, 80, 77–83.
- Mensah, Y., & Chen, H.-Y. (2013). *Global clustering of countries by culture—an extension of the GLOBE study*.
- Ogan, A., Walker, E., Baker, R., Rodrigo, M. M. T., Soriano, J. C., & Castro, M. J. (2015). Towards understanding how to assess help-seeking behavior across cultures. *International Journal of Artificial Intelligence in Education*, 25(2), 229–248.
- Pawlowski, J. M., & Hoel, T. (2012). Towards a global policy for open educational resources: The Paris OER Declaration and its implications. *White Paper, Version 0.2, Jyväskylä, Finland*.

- Pimmer, C., Mateescu, M., & Gröbhel, U. (2016). Mobile and ubiquitous learning in higher education settings. A systematic review of empirical studies. *Computers in Human Behavior*, 63, 490–501.
- Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science*, 363(6423), 130–131.
- Reinecke, K., & Bernstein, A. (2011). Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces. *ACM Transactions on Computer-Human Interaction*, 18(2), 1–29.
<https://doi.org/10.1145/1970378.1970382>
- Rohs, M., & Ganz, M. (2015). MOOCs and the claim of education for all: A disillusion by empirical data. *The International Review of Research in Open and Distributed Learning*, 16(6).
- Ruipérez-Valiente, J. A., Halawa, S., Slama, R., & Reich, J. (2020). Using multi-platform learning analytics to compare regional and global MOOC learning in the Arab world. *Computers & Education*, 146, 103776.
- Ruipérez-Valiente, J. A., Jenner, M., Staubitz, T., Li, X., Rohloff, T., Halawa, S., Turro, C., Cheng, Y., Zhang, J., Despujol, I., & Reich, J. (2020). Macro MOOC learning analytics: Exploring trends across global and regional providers. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 518–523. <https://doi.org/10.1145/3375462.3375482>
- Sharples, M. (2015). *FutureLearn Learning Design Guidelines*.
- Stich, A. E., & Reeves, T. D. (2017). Massive open online courses and underserved students in the United States. *The Internet and Higher Education*, 32, 58–71.
- Sunar, A. S., White, S., Abdullah, N. A., & Davis, H. C. (2016). How learners' interactions sustain engagement: A MOOC case study. *IEEE Transactions on Learning Technologies*, 10(4), 475–487.
- Tylor, E. B. (1871). *Primitive culture: Researches into the development of mythology, philosophy, religion, art, and custom* (Vol. 2). J. Murray.
- Uchidiuno, J., Koedinger, K., Hammer, J., Yarzebinski, E., & Ogan, A. (2018). How Do English Language Learners Interact with Different Content Types in MOOC Videos? *International Journal of Artificial Intelligence in Education*, 28(4), 508–527. <https://doi.org/10.1007/s40593-017-0156-x>
- Wallerstein, I. M. (1991). *Geopolitics and geoculture: Essays on the changing world-system*. Cambridge University Press.

- Wladis, C., Hachey, A. C., & Conway, K. (2015a). Which STEM majors enroll in online courses, and why should we care? The impact of ethnicity, gender, and non-traditional student characteristics. *Computers & Education*, 87, 285–308.
- Wladis, C., Hachey, A. C., & Conway, K. (2015b). Which STEM majors enroll in online courses, and why should we care? The impact of ethnicity, gender, and non-traditional student characteristics. *Computers & Education*, 87, 285–308.
- World Bank Country and Lending Groups – World Bank Data Help Desk. (2017, November 10).
<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>
- Xing, W. (2019). Exploring the influences of MOOC design features on student performance and persistence. *Distance Education*, 40(1), 98–113.
- Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. *Proceedings of the 2013 NIPS Data-Driven Education Workshop*, 11, 14.
- Zou, W., Hu, X., Pan, Z., Li, C., Cai, Y., & Liu, M. (2021). Exploring the relationship between social presence and learners' prestige in MOOC discussion forums using automated content analysis and social network analysis. *Computers in Human Behavior*, 115, 106582. <https://doi.org/10.1016/j.chb.2020.106582>

APPENDIX A

Figure A1: KM curves illustrating the differences between survival probabilities with respect to activities accessed for ten cc subgroups. The risk table shows absolute number of learners, and the percentage of learners at risk by that point.

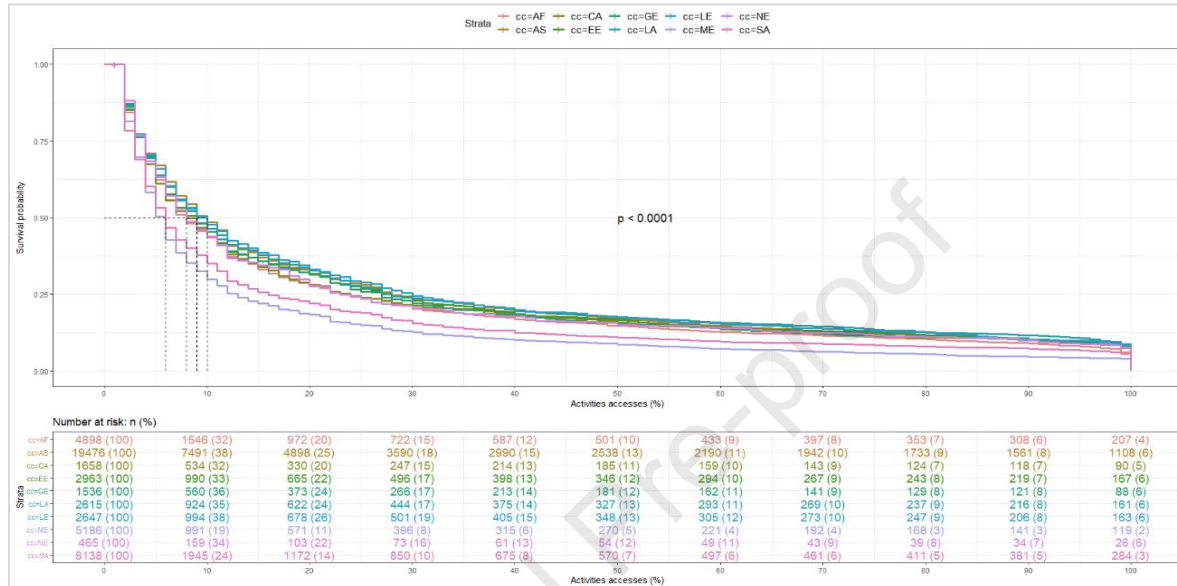


Figure A2: KM curves illustrating the differences between survival probabilities for four sec subgroups.

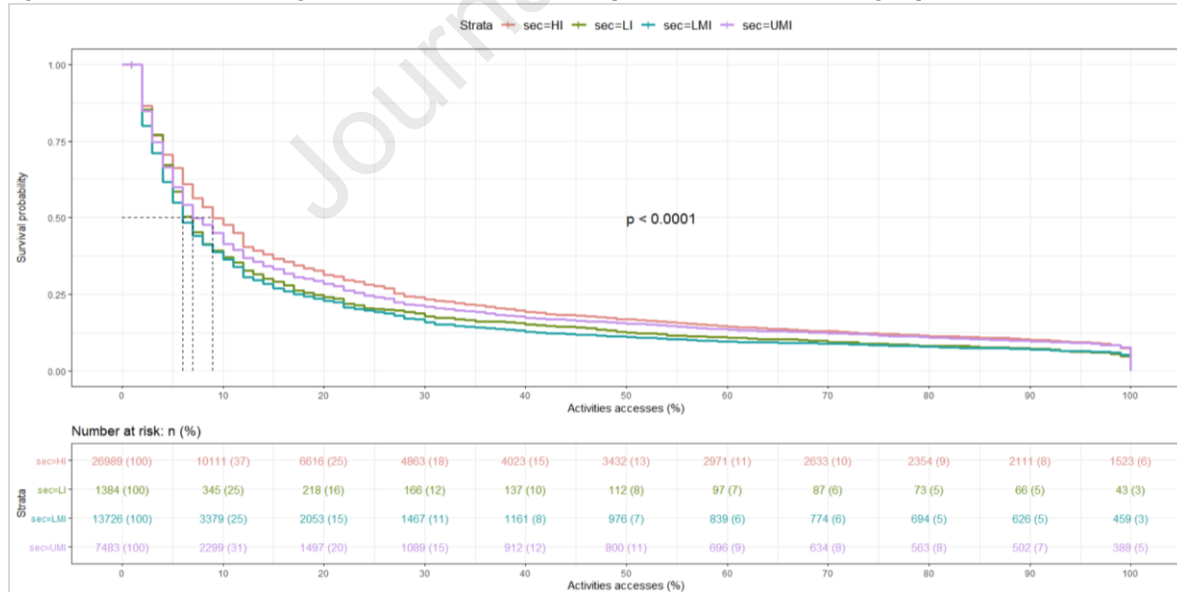


Table 1: Number of learners and learners who accessed at least 1% activities (%) in respective courses, median activities accessed (%), and the number of learning design (LD) features (Articles, Discussions, Videos and Quizzes).

Disciplines	Course	Learners (n=49582)	Learners with activities (n=34968) (%)	>1% Median activity (%) [LCL, UCL]	A	D	V	Q
Science, Engineering & Math and IT & Computer Science	aml1	1564	1164 (74.4)	8 [6, 8]	32	25	27	24
Business and Finance Fundamentals	bfec11	6232	3831 (61.5)	9 [7, 9]	40	13	16	5
Business & Management and Creative Arts & Media	bof7	1819	1264 (69.5)	11 [10, 13]	37	13	39	1
Nature & Environment and Science, Engineering & Math	ere6	1948	1367 (70.2)	24 [21, 24]	44	12	8	3
Business & Management	fbg7	1280	689 (53.8)	7 [7, 8]	52	7	24	5
Business & Management	fpb4	1506	956 (63.5)	12 [10, 15]	44	13	21	3
IT & Computer Science	ics18	7293	4498 (61.7)	9 [8, 9]	92	23	26	1
Language and Study Skills	leapfs1	9186	6461 (70.3)	7 [6, 7]	41	16	28	13
Business & Management	mllt11	4185	3208 (76.7)	8 [7, 8]	60	4	14	4
Creative Arts & Media and Literature	swf15	14569	11530 (79.1)	8 [7, 8]	78	16	14	6
Mean \pm SD					52.0 \pm 19.3	14.2 \pm 6.3	21.7 \pm 9.0	6.5 \pm 7.0

Table 2. Methods employed to address research questions

RQ	Methods	Predictors	Results
RQ1	One overarching survival model	nA, nV, nD, nQ	Table 3
RQ2	(i) Subset analysis: Subset data based on the ten geo-cultural clusters (cc) and fit one survival models for each subset. (ii) Interaction analysis: Fit one survival model to the full dataset and extract important interactions using cross-validation.	nA, nV, nD, nQ cc*nA, cc*nV, cc*nD, cc*nQ	Table 4
RQ3	(i) Subset analysis: Subset data based on the four socioeconomic clusters (sec) and fit one survival model for each subset. (ii) Interaction analysis: Fit one survival model to the full dataset and extract important interactions using cross-validation.	nA, nV, nD, nQ sec*nA, sec*nV, sec*nD, sec*nQ	Table 5

Table 3. Coefficient estimates and hazard ratios for the Cox regression model fitted to address RQ1.

Predictor	coef (se)	exp (coef)	PH test p-value
nA	0.13*** (0.009)	1.14	0.23
nV	0.03*** (0.007)	1.03	0.02
nD	-0.03*** (0.008)	0.97	0.33
nQ	0.14*** (0.007)	1.15	0.10

Significance codes: '***' $p < 0.0001$; '**' $p < 0.001$; '*' $p < 0.01$; '.' $p < 0.05$

Table 4. Comparison of coefficient estimates for the models fitted to address RQ2. The table reports coefficients from (i) subset analysis and (ii) two-way interaction tests. Hazard ratios are computed by exponentiating these estimates.

Cc	Learner with>1 % activities	Median activity (%)	Articles (A)		Videos (V)		Discussions (D)		Quizzes (Q)	
		[0.95LCL,0.95UCL]	cc subset analysis	Interaction test (cc*nA)	cc subset analysis	Interaction test (cc*nV)	cc subset analysis	Interaction test (cc*nD)	cc subset analysis	Interaction test (cc*nQ)
			coef (se)	Coef	coef (se)	Coef	coef (se)	coef	coef (se)	Coef
AF	3391	8 [7, 8]	0.07 ** (0.025)	0.07**	0.02 (0.021)	0.02	0.09 *** (0.021)	0.09 ***	0.10 *** (0.024)	0.09***
AS	14839	10 [9, 10]	0.25 *** (0.014)	0.24 ***	0.07 *** (0.011)	0.07 *	-0.13 *** (0.012)	-0.12 ***	0.23 *** (0.012)	0.23***
CA	1141	8 [7, 9]	0.14 * (0.054)	0.14	0.06 (0.038)	0.06	-0.13 ** (0.043)	-0.14 ***	0.11 ** (0.040)	0.12
EE	2064	9 [8, 10]	0.06 * (0.035)	0.06	-0.01 (0.026)	-0.01	0.02 (0.033)	0.02	0.04 (0.029)	0.04
GE	1159	9 [8, 10]	0.28 *** (0.052)	0.28 ***	0.03 (0.037)	0.03	-0.25 *** (0.045)	-0.24 ***	0.26 *** (0.040)	0.25 ***
LA	1851	9 [9, 10]	0.39 *** (0.040)	0.37 ***	0.09 *** (0.027)	0.09 *	-0.12 *** (0.036)	-0.12***	0.24 *** (0.031)	0.24 ***
LE	1972	10 [9, 10]	0.18 *** (0.039)	0.18*	0.08 ** (0.028)	0.08.	-0.11 ** (0.036)	-0.11 ***	0.13 *** (0.031)	0.12
ME	3039	6 [5, 6]	0.08 ** (0.032)	0.08	0.09 *** (0.025)	0.10 *	0.04 (0.026)	-0.05	0.07 * (0.027)	0.07
NE	345	8 [7, 10]	0.19 * (0.095)	0.20	0.04 (0.068)	0.04	-0.20 * (0.081)	-0.20 ***	0.19 * (0.075)	0.19
SA	5167	6 [6, 6]	-0.04 * (0.021)	-0.05***	-0.06 *** (0.018)	-0.07**	0.19 *** (0.019)	0.21***	-0.03 * (0.019)	-0.04 ***

Significance codes: '***' $p < 0.0001$; '**' $p < 0.001$; '*' $p < 0.01$; '.' $p < 0.05$ Entries with light grey font colors were found to be insignificant. Cells highlighted in blue indicate interaction terms that were selected in cross-validation (using `cv.glmnet` and default `lambda = 1se`, as recommended (Hastie & Qian, 2014))

Table 5. Comparison of coefficient estimates for the models fitted to address RQ3. The table reports coefficients from (i) subset analysis and (ii) two-way interaction tests. Hazard ratios are computed by exponentiating these estimates.

Sec	Learners with >1% activities	Median activity (%)	Articles (A)		Videos (V)		Discussions (D)		Quizzes (Q)	
		[0.95LC L,0.95U CL]	sec subset analysis	Interaction test (A * sec)	sec subset analysis	Interaction test (V * sec)	sec subset analysis	Interaction test (D * sec)	sec subset analysis	Interaction test (Q * sec)
			coef (se)	Coef	coef (se)	Coef	coef (se)	Coef	coef (se)	coef
HI	20267	9 [9, 10]	0.23 *** (0.01)	0.22 ***	0.06 *** (0.009)	0.06 ***	-0.13 *** (0.010)	-0.12 ***	0.20 *** (0.009)	0.20***
UMI	5111	7 [7, 8]	0.16 *** (0.02)	0.17*	0.04 * (0.017)	0.041	0.01 (0.019)	0.01***	0.15 *** (0.019)	0.15 *
LMI	8709	6 [6, 6]	0.01 (0.016)	0.01***	-0.01 (0.014)	0.01***	0.12 *** (0.014)	0.13***	0.02 (0.14)	0.02***
LI	881	7 [6, 7]	0.04 (0.05)	0.04 ***	0.01 (0.05)	0.01	0.09 * (0.04)	0.09***	0.04 (0.04)	0.04 **

Significance codes: '***' $p < 0.0001$; '**' $p < 0.001$; '*' $p < 0.01$; '.' $p < 0.05$

Entries with light grey font colors were found to be insignificant. Cells highlighted in blue indicate interaction terms that were selected in cross-validation (using cv.glmnet and default lambda = 1se, as recommended (Hastie & Qian, 2014))

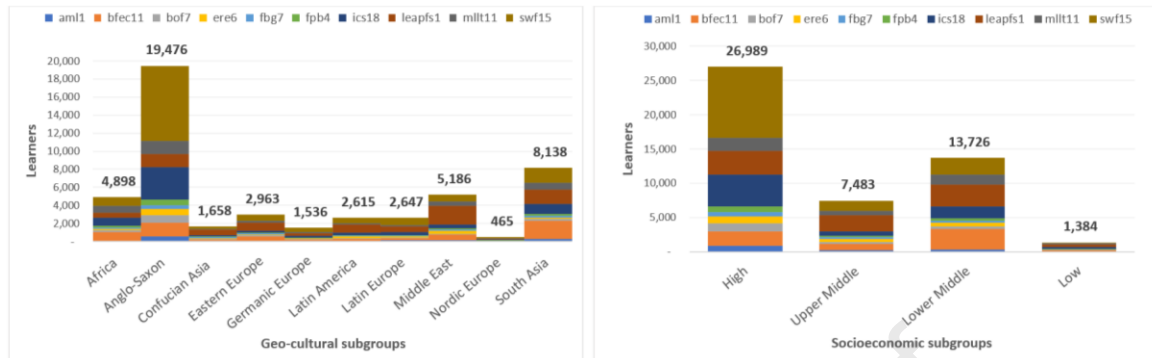


Figure 1: Distribution of learners in ten geo-cultural and four socioeconomic subgroups.

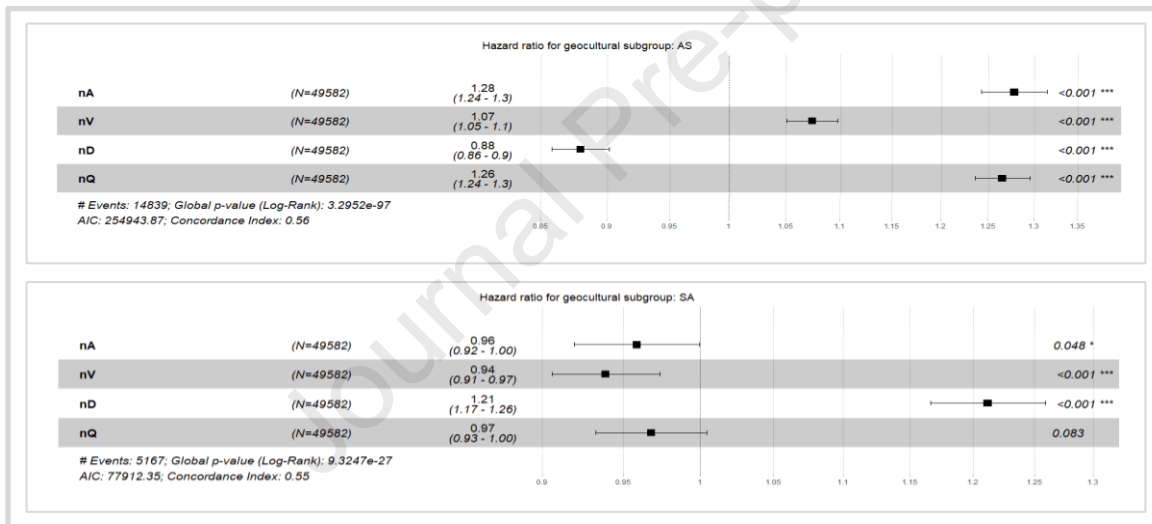


Figure 2. Contrasting behavior of learners in the two largest geo-cultural subgroups, Anglo-Saxon (top) and South Asia (bottom), showing the opposite results for the two largest geo-cultural subgroups in data.

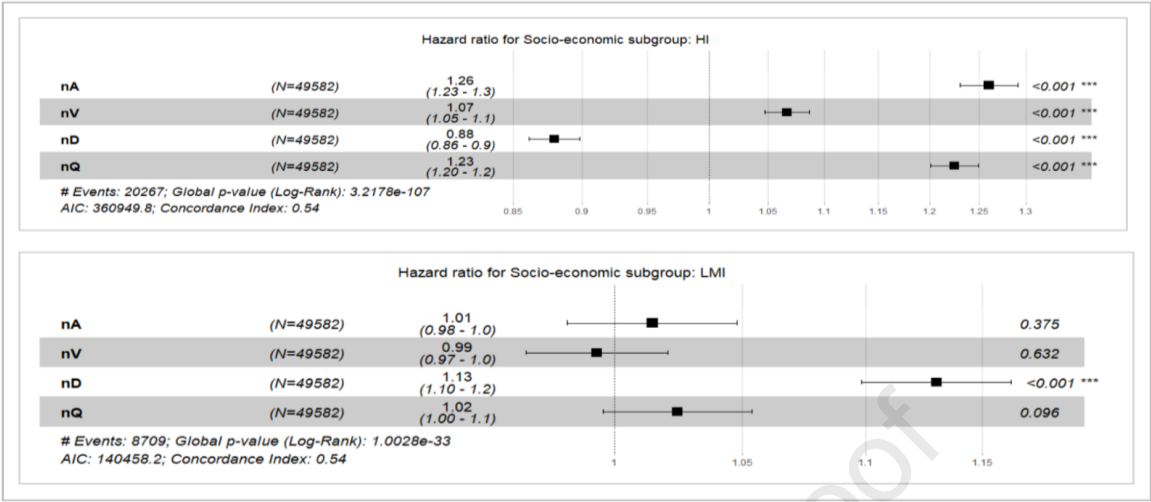


Figure3: The contrasting behavior of learners in the two largest socioeconomic subgroups: High Income (top) and Lower-Middle Income (bottom), showing the opposite results for the two largest socioeconomic subgroups in data

Highlights:

- There is no ideal combination of coherent learning activities that work for *ALL* learners.
- Certain learning activity types facilitate progress for learners in one context while inhibiting another.
- Preference for video over text was strong in collectivists, particularly those from the South Asian region.
- Making MOOC a *social learning space* supported western learners from affluent economies.
- Overall data analysis can mask geo-cultural and socioeconomic heterogeneity in the relationship between learning design and learner outcomes.