

November – 2020

Heterogeneity of Learners' Behavioral Patterns of Watching Videos and Completing Assessments in Massive Open Online Courses (MOOCs): A Latent Class Analysis

In Gu Kang
Boise State University

Abstract

Massive open online courses (MOOCs) have been touted as an effective way to make higher education accessible for free or for only a small fee, thus addressing the problem of unequal access and providing new opportunities to young people in middle and low income groups. However, many critiques of MOOCs have indicated that low completion rates are a major concern. Using a latent class analysis (LCA), a more advanced methodology to identify latent subgroups, this study examined the heterogeneity of learners' behavioral patterns in a MOOC, categorized them into distinctive subgroups, and ultimately determined the optimal number of latent subgroups in a MOOC. The five subgroups identified in this study were: *completing* (6.6%); *disengaging* (4.8%); *auditing* (4.6%); *sampling* (21.1%); and *enrolling* (62.8%). Results indicated this was the optimal number of subgroups. Given the characteristics of the three at-risk subgroups (disengaging, sampling, and enrolling), tailored instructional strategies and interventions to improve behavioral engagement are discussed.

Keywords: MOOC, learner behavioral engagement, tailored intervention, latent class analysis

Introduction

Increases in tuition and fees have reduced opportunities and accessibility to higher education especially for young people in middle- and low-income groups in the United States (College Board, 2016). To overcome this barrier, many universities and institutions offer massive open online courses (MOOCs), which are publicly available for free or for a small fee to anyone who wants to learn. Despite these efforts, critiques of MOOCs have pointed out that low completion rates have been a major concern (Kizilcec, Piech, & Schneider, 2013; Wang & Baker, 2015).

Some scholars claim that success of learners in MOOCs should be distinguished from success of students in traditional learning environments (Henderikx, Kreijns, & Kalz, 2017; Koller, Ng, Do, & Chen, 2013; Reich, 2014). Many educators tend to think of success of learners in MOOCs as official completion, meeting certain requirements, or earning credentials (Henderikx et al., 2017). However, many researchers assert that determining success in MOOCs needs to account for student intentions (Reich, 2014). It is known that approximately 5% of learners who enroll in MOOCs earn a credential indicating completion of the course (Jordan, 2014; Koller et al., 2013). However, if success is defined from the perspectives and intentions of learners, success rates range from 59% to 70% (Henderikx et al., 2017). Despite this argument, low completion rates have been of great concern to MOOC instructors and designers. To better understand student learning and success, instructors need to take a new approach, developing instructional strategies for MOOC learners by taking into account heterogeneity and the distributed nature of learners in MOOCs (Kizilcec et al., 2013; Koller et al., 2013).

Although research on MOOCs has rapidly grown, there is not yet an extensive body of literature on the advanced learning analytics of MOOCs. Given the very low completion rate, it is important to understand the patterns of learners' engagement in MOOCs in order to develop adaptive and specific learning mechanisms (Henderikx et al., 2017; Kizilcec et al., 2013). MOOCs are intended to include diverse populations in a single online learning environment. Learners often span wide ranges of age, location, educational background, and native language. Given the diverse profiles of MOOC learners, MOOC instructors and designers should select instructional strategies and intervention plans that match learner characteristics so that individual learners can be successful in MOOCs. This study examined the heterogeneity of learners' behavioral patterns, categorized learners into different behavior pattern groups, and ultimately developed tailored interventions for at-risk subgroups.

Literature Review

Learners' Behavioral Engagement in MOOCs

Most of the previous studies that investigated learners' engagement in academic activities were focused primarily on behavioral engagement (Jung & Lee, 2018). The most commonly used indicators of behavioral engagement were watching lecture videos, taking quizzes, completing tasks, and posting to forums (Jung & Lee, 2018; Li & Baker, 2016). It is important to identify patterns of learners' behavioral engagement since this allows MOOC instructors to understand how learners interact with content in MOOCs, detect at-risk

learner groups, and tailor interventions for improving engagement and learning outcomes (Bote-Lorenzo & Gómez-Sánchez, 2017; Ramesh, Goldwasser, Huang, Daume, & Getoor, 2014). Recently, Phan, McNeil, and Robin (2016) researched the relationship between learners' behavioral engagement (e.g., assignment submission and participation in discussions) and their performance in MOOCs and found that actively engaged learners showed better performance than those who were less engaged. Thus, among the multidimensional aspects of engagement, learners' behavioral engagement is a strong indicator of success in MOOCs.

Classification of Subgroups in MOOCs

Recent literature on MOOCs is increasingly moving beyond basic analytics to explore deeper level constructs such as intent, persistence, and behavior, as well as to make attempts at developing prediction models from findings. Kizilcec et al. (2013) investigated the pattern of learners' engagement, resulting in the creation of learner subpopulations identified as the trajectory of engagement using a *k*-means clustering algorithm. They found that while survival statistics counting *completing learners* are the most common measure of success in MOOCs, other types of learners such as *auditing*, *disengaging*, and *sampling* learners are subgroups found in MOOCs. Table 1 shows brief descriptions of the four subgroups identified by Kizilcec et al. (2013).

Table 1

The Trajectory of Engagement by Subgroups in MOOCs

Subgroups	Description
Completing learners	<ul style="list-style-type: none"> Completed the majority of the assessments At least attempted the assignments Were most similar to a student in a traditional class
Disengaging learners	<ul style="list-style-type: none"> Did assessments at the beginning of the course but then had a marked decrease in engagement Disengaged at different points in the course, but generally in the first third of class
Auditing learners	<ul style="list-style-type: none"> Did assessments infrequently, if at all Engaged by watching video lectures Followed course for the majority of its duration Did not obtain course credit
Sampling learners	<ul style="list-style-type: none"> Watched video lectures for only one or two assessment periods "Sampled" at the beginning of the course or briefly explored the material when the class was already fully under way

Ferguson and Clow (2015) examined the generalizability of the subgroup categories found in the research of Kizilcec et al. (2013). They replicated the research procedures and tested whether the same patterns of learner engagement were found in MOOCs where instructors used social constructivist pedagogy. Results

showed seven subgroups: *samplers*, *strong starters*, *returners*, *midway dropouts*, *nearly there*, *late completers*, and *keen completers*. Table 2 shows brief descriptions of the seven subgroups described by Ferguson and Clow (2015).

Table 2

The Subgroups Found in MOOCs

Subgroups	Description
Cluster I: Samplers	<ul style="list-style-type: none"> • Visited the course, but only briefly • Accounted for 37%-39% of learners and made up the largest cluster in all four MOOCs
Cluster II: Strong starters	<ul style="list-style-type: none"> • Submitted the first assignment, but then their engagement dropped off sharply, with very little activity after that • Made up 8%-14% of learners
Cluster III: Returners	<ul style="list-style-type: none"> • Completed the assessment in the first two weeks, returned to do so again in the second week, and then dropped out • Made up 6%-8% of learners in the three of the four MOOCs
Cluster IV: Midway dropouts	<ul style="list-style-type: none"> • Completed three or four assessments, but then dropped out about halfway through the course • Visited about half of the course (47%, 59%), and roughly half posted comments (38%, 49%), posting 6.3-6.5 comments on average
Cluster V: Nearly there	<ul style="list-style-type: none"> • Consistently completed assessments, but then dropped out just before the end of the course • Accounted for 5%-6% of learners in all four MOOCs
Cluster VI: Late completers	<ul style="list-style-type: none"> • Completed the final assessment, and submitted most of the other assessments, but were either late or missed some • Accounted for 6%-8% of learners in the three of the four MOOCs
Cluster VII: Keen completers	<ul style="list-style-type: none"> • Completed all the assessments, including the final one, and almost all of them on time (>80%). • Accounted for 7% to 13% of learners in the three of the four MOOCs.

In advancing the research, Ferguson and her colleagues wondered if these patterns of learner engagement could be applied by MOOC instructors or designers. In order to determine the applicability to other MOOCs, they examined five MOOCs with a variety of course duration (e.g., 3 weeks, 6 weeks, 7 weeks, and 8 weeks; Ferguson et al., 2015). Results showed that those same seven subgroups were found in seven- or eight-week courses. On the other hand, the seven subgroups did not show up in relatively short MOOC courses, such as three-week courses (Ferguson et al., 2015). Rather, there were variations and new emerging patterns of learner engagement such as *saggers*, *improvers*, *surgers*, and *weak starters*.

Taken together, there have been many attempts to identify subgroups in MOOCs based on learners' behavioral patterns. However, the categorizations across the previous studies were not matched accurately and still remain questionable. This study investigated profiles of subgroups in a MOOC by analyzing behavioral engagement patterns. To optimize the number of subgroups, it was necessary to use a more advanced and rigorous methodological approach to clustering subgroups in MOOCs.

Research Question

From the literature review, it was evident that there were inconsistencies in the categorizations and the number of subgroups in MOOCs, and that a more rigorous method to identify subgroups in MOOCs would benefit prediction models. To determine subgroups, two research questions were posed:

- How many subgroups would emerge from a latent class analysis (LCA)?
- What characteristics would each subgroup have in common?

Identifying the characteristics of subgroups provides a foundation to develop the features of adaptive learning in MOOCs. Answering the two research questions, we profile the homogeneity of behavioral patterns in each subgroup and develop tailored strategies for learning activities and student achievement.

Method

Course Description, Samples, and Demographics

A MOOC, *Job Success: Get Hired or Promoted in 3 Steps*, was selected for this study. This course was offered through Coursera, which is a platform to deliver MOOCs, and was taught only in English. The purpose of this course was to show job seekers how to stand out in a crowded applicant pool so that they would get hired, and to teach anyone who already had a job how to get recognized and promoted. This course was self-paced, but it was suggested that students spend three hours per week over the course of three weeks to complete the MOOC. This course encompassed three sections and included 10 video lectures and 4 quiz assessments. The quiz assessments were formative and could be taken multiple times.

To collect students' behavioral data from Coursera, we requested the clickstream data on the selected MOOC from September 1, 2016 to February 22, 2017 via Coursera's research exports (<https://github.com/coursera/courseraresearchexports>), where MOOC researchers request Coursera research data such as assessment submission data, course grade data, course progress data, demographic data, and discussion data. In total, 3,955 learners enrolled in the course and their behavior patterns were part of the data. In order to better understand learners' backgrounds, the demographic survey embedded in the course was part of this study, though participation in the survey was voluntary.

Among the 3,955 learners, 1,489 (37.6%) completed the embedded survey. This included 470 (31.6%) male and 724 (48.6%) female learners, with missing values of 295 (19.8%). Learners came from 121 countries (41 Asia, 35 Europe, 17 Africa, 15 North America, 10 South America, and 3 Oceania). All respondents to the survey were included in the clickstream data.

Using Kizilcec et al.'s (2013) four MOOC subgroups (see Table 1), learners were asked about their intentions. Most enrolled to complete the entire course, i.e., 765 (51.4%), while 185 (12.4%) intended to look around and review items of interest, 180 (12.1%) planned to follow most of the course lectures and videos without completing assignments, and 25 (1.7%) aimed to finish at least the first unit. To understand how many learners had powerful motivation to achieve their goals, the Grit Scale was included in the survey. Grit in this case is defined as "trait-level perseverance and passion for long-term goals" (Duckworth & Quinn, 2009, p. 166). The average score on the Grit Scale was 3.4, indicating learners were moderately passionate about achieving their long-term goals in a MOOC. Finally, most of these students had prior experience with MOOCs: 560 (37.6%) had completed 1 to 3 courses, 208 (14.0%) had participated in 4 to 6 courses, and 170 (11.4%) had taken 7 or more courses. Only 212 (14.2%) fell into the category of having had no experience with MOOCs.

Instruments

The clickstream data from Coursera are defined as learners' interactions, based on clicking information, which is automatically saved in the Coursera system. The clickstream data cover two domains: (a) video (interactions with lecture videos such as start, stop, pause, change subtitles, and heartbeats) and (b) access (accessing the course description page and course materials). As the raw clickstream data were very segmented and unclear in the Coursera platform, they had to be transformed into an analyzable format using Microsoft Access data mining functions. In this study, we defined patterns of learners' behavioral engagement with the two primary features of the course: video lectures and assessments. In the two representative studies of subgroups in MOOCs, i.e., Kizilcec et al. (2013) and Ferguson and Clow (2015), the same indicators of behavioral patterns were used. Subgroups classified by these indicators can be generalized into other MOOC contexts regardless of course content or instructional strategies.

Learners' interactions with 10 video lectures and 4 quizzes were used to determine the behavioral patterns of subgroups. For each of the 10 video lectures, a learner was coded as *0* for *not watching a video lecture* or *1* for *watching video*. To ensure watching a video lecture was completed, *1* was coded only to the learner who watched a video lecture to the end. If a learner started to watch a video lecture, but did not complete it, *0* was coded for that particular learner and video. In addition, for each of the 4 quizzes, a learner was coded as *0* for *not taking quizzes* and *1* for *taking quizzes*. For instance, if a learner watched all 10 video lectures and took all 4 quizzes, the coding for that learner was [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]. If a learner watched all 10 video lectures but took only one quiz, the coding for that learner was [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0]. If a learner enrolled in the course, but did nothing during the course, the coding for that learner was [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0].

Data Analysis

LCA was employed to cluster and determine the optimal number of latent subgroups. LCA is a statistical technique used to identify a set of mutually exclusive subgroups of individuals who have similar characteristics. In MOOC research, LCA has been recently adopted as a novel methodological approach to categorize relatively smaller and homogeneous subgroups from entire heterogeneous populations using learners' behavioral engagement patterns such as discussion viewing (Bergner, Kerr, & Pritchard, 2015), forum use (Poquet, Dowell, Brooks, & Dawson, 2018), and profiles of student motivation (Moore & Wang,

2020). The ultimate goal of these studies was to develop tailored instructional interventions for each subgroup based on profiles.

LCA is considered methodologically superior to any of the other algorithms such as *k*-means clustering in that (a) while *k*-means uses an ad hoc approach, LCA uses a probabilistic model that enables cases to be classified into clusters, and (b) while *k*-means provides no diagnostics for determining the number of clusters, LCA can use a variety of model selection indices such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) (Magidson & Vermunt, 2002). Based on this model, learners have membership in certain latent classes instead of the researcher finding clusters with arbitrarily chosen distance measures as is the case in *k*-means clustering.

To determine the optimal number of subgroups and maximize the model fit in LCA, model fit indices such as the AIC and BIC and interpretability were considered (Lanza, Collins, Lemmon, & Schafer, 2007). The optimal number of subgroups among MOOC learners was determined based on evidence that minimized AIC and BIC from among groups that measured a frequency of at least five percent.

Results

The results of the LCA showed five distinct subgroups as the optimal number in a MOOC. Students in each of the five subgroups showed similar behavioral patterns of learning engagement. Table 3 shows the results of fitting the latent class models (two to six classes) for indicators according to the model selection process.

Table 3

The Model Fit Information for Latent Class Models of MOOCs (N = 3,955)

Subgroups	2-latent class model		3-latent class model		4-latent class model		5-latent class model		6-latent class model	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Class 1	3,404	86	499	13	835	21	261	7	262	7
Class 2	551	14	888	22	346	9	190	5	172	4
Class 3			2,568	65	2,503	63	184	5	201	5
Class 4					272	7	2,485	3	2,432	61
Class 5							835	21	505	13
Class 6									383	10
Goodness of fit statistics										
AIC	31053.06		26016.39		24968.17		24211.24		23635.64	

BIC	31235.26	26292.83	25338.86	24676.16	24194.80
-----	----------	----------	----------	----------	----------

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion. A smaller number in BIC and AIC indicates a better model fit.

The final model specification and the optimal number of latent classes were determined by considering the model fit indices and the interpretability of the results (Lanza et al., 2007). First, model selection indices AIC and BIC were used to determine the optimal number of subgroups in a MOOC. According to Chen, Luo, Palardy, Glaman, and McEnturff (2017), AIC and BIC are preferable and the most effective model selection indices when the sample size is large and a model is misspecified, as compared to other model selection indices such as the Vuong-Lo-Mendell-Rubin likelihood ratio test (VLMR), and the adjusted Lo-Mendell-Rubin likelihood ratio test (ALMR). VLMR and ALMR have a tendency to over extract the number of classes when the sample size is large. Because of the large sample size ($N = 3,955$), AIC and BIC were the most effective model selection indices for this study. Analyzing the results of the model selection indices, the values of AIC and BIC decreased from the 2-latent classes model to the 6-latent classes model, indicating a better model fit as the number of classes increased. However, I stopped with the 6-latent classes model since, at that point, the proportion of class 2 was less than 5%, which is hard to interpret as a subgroup.

Second, to determine which of the 5-latent class or 6-latent class model had a better fit, the interpretability of the results was considered based on findings in previous literature profiling subgroups in MOOCs. Classes 1, 2, 3, and 4 in a 6-latent class model showed no or little change in the proportion compared to the 5-latent class model. Class 5 from the 5-latent class model was divided into only two subgroups in the 6-latent class model. These two subgroups were not found or interpretable based on previous literature. On the other hand, Class 5 in the 5-latent class model could be defined as the *sampler* or *sampling* group based on previous literature (e.g., Kizilcec et al., 2013; Ferguson & Clow, 2015). Consequently, the 5-latent class model was selected as the best fitting model given the fit indices AIC and BIC as well as the interpretability of the results.

Given the overall probability of endorsing the consequences, the five subgroups were categorized and named after discussions with the research team as follows: class 1 was the *completing* group; class 2 was the *disengaging* group; class 3 was the *auditing* group; class 4 was the *enrolling* group; and class 5 was the *sampling* group. The *enrolling* group (63%) exhibited the highest probabilities for being defined as a subgroup, followed by the *sampling* group (21%), the *completing* group (7%), the *disengaging* group (5%), and finally the *auditing* group (5%).

Overall, the behavioral patterns of each subgroup were almost the same as in the subgroups identified in Kizilcec et al. (2013). The *completing* group showed the highest level of completion of quiz assessments and watching video lectures and consisted of 6.6% of the entire enrollment. This group watched almost all 10 video lectures and completed all 4 quizzes. The *disengaging* group accounted for about 5% of the entire enrollment. In the *disengaging* group, almost half the students took the 4 quizzes and watched the first video lecture. However, the proportion of students watching video lectures decreased over time. The *auditing* group made up about 5% of the entire learner enrollment. In the *auditing* group, most learners watched almost all video lectures, but they completed only the first quiz or no quizzes during the course. This group aimed to get domain-specific knowledge from the course. Most of the learners in the *sampling*

group watched only the first 4 of the 10 video lectures. In addition, they took no quizzes during the course. This *sampling* group consisted of more than 20% of the entire enrollment. The *enrolling* group comprised the largest proportion of the entire enrollment. They consisted of almost 63% of the entire enrollment. This group enrolled in the course but did nothing or did not show up for the course. This group is a new subgroup found only in this study; it has never been categorized as a subgroup in previous studies. For instance, although Anderson (2013) briefly mentioned the similar *enrolling* group who enroll in MOOCs, but never login to or show up the course, those learners were not profiled or paid attention to as an important subgroup we should focus on in MOOCs. Table 4 shows the characteristics of each subgroup in this study.

Table 4

Characteristics of the Five Subgroups

Subgroup	Description
Class 1: Completing group	<ul style="list-style-type: none"> Learners watched almost all 10 video lectures and took almost all 4 quizzes. On average, 95% of learners watched 10 video lectures and 93% of learners took 4 quizzes.
Class 2: Disengaging group	<ul style="list-style-type: none"> About 30% of the learners watched video lectures. In the beginning, about half the learners watched the first video lecture and then the proportion decreased over time. On average, 54% of the learners took all 4 quizzes and tended to show a slight decrease in taking quizzes over time.
Class 3: Auditing group	<ul style="list-style-type: none"> About 86% of the learners watched most of the videos (8 out of 10), but they had a tendency to show a slight decrease in watching the last two video lectures. In addition, 61% of the learners took the first quiz and then the proportion who took the rest of the quizzes sharply decreased up until it reached only 3%.
Class 4: Enrolling group	<ul style="list-style-type: none"> Learners enrolled, but they did almost nothing in the course. Only 2% of these learners watched some of the video lectures and only 1% took the quizzes.
Class 5: Sampling group	<ul style="list-style-type: none"> Almost all learners watched the first two video lectures and then the proportion who watched decreased from the fourth video lecture. From the fifth video lecture, almost all learners did not watch. In addition, only 30% of learners took the first quiz and did not take the rest of the quizzes.

Discussion

Completing and auditing groups were considered successful subgroups since they achieved the goals they set going into the MOOCs. Thus, there remain three at-risk subgroups: disengaging, sampling, and enrolling. Tailored interventions are discussed in this section.

Different Tailored and Effective Interventions

Disengaging group. Learners in this group showed lower levels of engagement in the MOOC. For some reason, learners' intentions to complete the course changed negatively, which in turn resulted in lower levels of engagement over time. Kizilcec et al. (2013) pointed out that there were two reasons why learners in this group disengage over time: personal commitment and conflict with schedules at work. Personal commitment, combined with a lower level of motivation, resulted in a lower level of engagement. This often occurred when the learner failed to set clear goals. This was related to self-regulated learning skills, namely goal setting and strategic planning. Kizilcec, Pérez-Sanagustín, and Maldonado (2017) found that these two self-regulated learning strategies were positively related to goal attainment. Thus, an orientation session on self-regulated learning strategies would be helpful for increasing the confidence level of learners in this group.

Another primary cause of disengaging in MOOCs was conflict with schedules at work. According to Chen, Alcorn, Christensen, and Eriksson (2015), 60% of learners who completed MOOCs were full-time employees. They successfully managed their time and the effort needed for their MOOCs. However, some learners in this *disengaging* group failed to balance their work with their learning in the MOOC. This may have been caused by too much work, too many obligations, and a lack of organizational support. According to Waddoups (2016), only 17% of workers receive organizational and supervisory support for training, reaffirming that most organizations do not use and support MOOCs for professional development. To get support from supervisors and organizations, employees may pursue certificates through MOOCs since some employers are willing to pay for certificates and will support such types of professional development (Hamori, 2019). A further factor to consider is that some organizations will provide time off rather than tuition reimbursement to support employees' professional development through MOOCs, since MOOCs involve little or no cost (Hamori, 2019). However, to take time off for MOOCs, employees need to receive approval from supervisors and organizations before enrolling. MOOC instructors may consider developing a MOOC readiness assessment checklist that would help learners determine the status of supervisory and organizational support for MOOCs before they enroll.

Sampling group. Like the *disengaging* group, learners in this group showed a relatively lower level of engagement. They watched about 4 or fewer of the 10 video lectures and did not take any quizzes. One difference between the two groups is that while the *disengaging* group watched only the first few video lectures and were not engaged in the latter part of the course, the *sampling* group explored the video lectures both in the beginning as well as later in the session. Learners in this group aimed to explore the content and materials. In this sense, they were passive learners who were not willing to engage in diverse activities.

The reasons why the *sampling* group was passive can be found in the course's task design and/or level of facilitation during the MOOC (Cassidy, Breakwell, & Bailey, 2014). For instance, learners who prefer

learning in a group are more likely to engage in discussion forums, while others who only seek specific information are more likely to engage in individualized tasks and learning activities. Thus, MOOC instructors should develop a variety of tasks such as individual work (e.g., quizzes, tests, case studies, and knowledge checks) and group work (e.g., small group discussions and peer-reviewed assignments), taking into consideration the characteristics of individual learners. In addition, instructors should help learners get more involved in activities by posting announcements, participating in discussion forums, and encouraging completion of the course.

Enrolling group. Although learners in this group enroll mostly out of curiosity, and have no intention to complete their courses, they should be carefully dealt with since they comprise the largest portion of enrollees (e.g., 63% in this study). While little is known about the characteristics of this group, enrollment itself could be understood to mean that learners have an interest in or curiosity about the contents or learning in MOOCs. Participants in this group are potential subjects who may return to enroll in future MOOCs. According to Reich (2014), learners' intentions can change, and these "intention flips" are a good indicator of success in MOOCs. Thus, course designers and instructors should try to understand how they could help these learners change from having no or little intention to complete the MOOC into an intention to instead get involved.

There are several possible reasons why these learners do nothing after enrolling: (a) the course content is different from what was expected (irrelevance); (b) learners are unsure about their abilities to master the contents (less confidence); (c) learners have no experience of MOOCs (no experience); and/or (d) there is poor user interface design. Instructors could take a number of steps to address these issues.

Instructors could provide a video preview giving a quick glimpse of a course so that *enrolling* learners could determine whether their interests and intentions fit the course. Instructors could send a video preview link to the *enrolling* learners to remind them about participating in the course.

In addition, instructors could create a short description of the characteristics that lead to success in MOOCs, such as self-regulated learning. Those who have no experience of MOOCs are unlikely to have successful strategies for completing the courses they enroll in. Like the *disengaging* and *sampling* groups, this group may need to develop self-regulated learning skills and perform activities to set their goals and strategically plan at the beginning of the course.

Finally, if it is the user interface design that is causing problems, steps can be taken to improve the situation. For example, when learners first log in, a road map or short tutorial might be helpful to guide those who have no experience of MOOCs. A navigation pane would also be helpful for predicting the course structure. Table 5 shows a summary of tailored interventions for the three at-risk subgroups.

Table 5

Tailored Instructional Design and Intervention Strategies for the Three At-Risk Subgroups

At-risk subgroups	Instructional design and intervention strategies
Disengaging group	<ul style="list-style-type: none"> • Strategy 1. Create an orientation session for self-regulated learning strategies to teach goal setting and strategic planning. • Strategy 2. Create a MOOC readiness assessment checklist that helps learners review the status of supervisory and organizational support for MOOCs.
Sampling group	<ul style="list-style-type: none"> • Strategy 3. Develop a variety of tasks that consider individual learning preferences. • Strategy 4. Facilitate learners to participate in various learning activities by sending messages and reminders.
Enrolling group	<ul style="list-style-type: none"> • Strategy 5. Change no intention into a good intention by providing a video preview, creating a short description on learning in MOOCs, and improving the user interface design.

Implications

Methodological implications. In this study, LCA was adopted to profile individual learners' behavioral patterns of watching videos and completing assessments. LCA is a model-based approach to clustering subgroups from an entire population. An advantage of using a model-based approach over a data-driven cluster approach (e.g., *k*-means clustering) is to provide fit statistics that help researchers determine the most appropriate model for data and to compare models to arrive at the optimal number of subgroups for hypothesis testing. On the other hand, the results of the existing studies using *k*-means clustering showed an arbitrary or inconsistent number of subgroups in MOOCs. Furthermore, LCA is a more rigorous approach to clustering subgroups in MOOCs and helping researchers find model-based subgroups. LCA can be extended into any open and distributed learning environment (e.g., online learning, MOOCs, and blended courses) in order to cluster subgroups and develop tailored interventions. LCA as a novel approach has wider applicability to open and distributed learning environments.

Practical implications. The *disengaging* and *sampling* groups are at-risk learner groups since they start a course with good intentions and interest but they do not continue to engage in learning. For unknown reasons, their good intentions change into negative ones. To promote engagement in these two at-risk subgroups, this study suggests tailored interventions for the *disengaging* group (e.g., an orientation session for self-regulated learning strategies and a MOOC readiness assessment checklist) and for the *sampling* group (e.g., a variety of tasks responding to individual preferences and encouragement strategies

for active participation). These tailored interventions provide insights and information on effective instructional strategies for MOOC instructors and learners.

The characteristics of the *enrolling* group are unknown and have not been included in previous MOOC research. Although inactive, learners in this group could return to the course in the future. As this group comprises the largest proportion of the entire number of enrollees, they should be dealt with as one of the subgroups in MOOCs. Although the *enrolling* group has no clear intention and thus does not perform the behavior necessary to complete the course, their interest or curiosity caused them to enroll in the first place. Thus, the first step to develop strategies for this group is to find a way in which to help them discover their own intention to learn (e.g., through a video preview), and to translate these intentions into performing behaviors (e.g., the activity for goal setting and strategic planning).

Conclusion

This study identified the subgroups and determined the optimal number of subgroups in a MOOC using an LCA. First, focusing on the three at-risk subgroups, tailored interventions were developed. These tailored interventions will help MOOC instructors and designers get insight into instructional strategies for each of the at-risk subgroups. Second, LCA, a model-based approach to clustering subgroups in a MOOC, provided convincing evidence on the optimal number of subgroups so that MOOC instructors can develop tailored and effective interventions. In conclusion, this study is the first step towards more theoretically and empirically grounded research into learner engagement in MOOCs and contributes to developing the foundation of adaptive learning analytics.

Limitations and Suggestions for Future Research

Based on limitations, this study suggests some areas for future research. First, one of the limitations of this study was the use of only two types of behavioral engagement (watching video lectures and completing assessments) from the clickstream data to identify subgroups. However, the clickstream data included many other types of indicators such as average play speed of a video, stop, pause, and rewind, and whether the learner started the quiz and left without answering any questions. Using multiple indicators of learners' behavior in the clickstream data, the profiles of each subgroup could be elaborated. Thus, the granularity of the behaviors in the clickstream data could help refine the profiles of each subgroup.

Second, only one aspect of behavioral engagement was used to identify subgroups. In fact, there are other types of learner engagement in MOOCs such as cognitive engagement, affective/emotional engagement, and social engagement. Jung and Lee (2018) indicated that learner engagement in MOOCs includes the multidimensional approach to emotional, cognitive, and behavioral aspects. Thus, as the results of this study are limited to behavioral aspects of learner engagement, future research should delve into the multidimensional aspects of engagement in MOOCs by collecting data from multiple sources such as surveys, clickstream data, and interviews. For instance, follow-up interviews with learners in at-risk groups would help better understand why they are failing to complete courses. A variety of external factors could be addressed while collecting multiple data.

Finally, the data were collected from a single 3-week MOOC. Although this study provides empirical evidence of the sub-groups that emerged from an LCA, the findings may be hard to transfer or generalize into other MOOC courses and contexts. Thus, data could be collected from multiple MOOC courses, with diverse content, from different disciplines, of varying duration, and from various providers and platforms.

References

- Anderson, T. (2013, March). Promise and/or peril: MOOCs and open and distance education. *Commonwealth of Learning*, 3, 1-9. Retrieved from <http://www.ethicalforum.be/sites/default/files/MOOCsPromisePeril.pdf>
- Bergner, Y., Kerr, D., & Pritchard, D. E. (2015). Methodological challenges in the analysis of MOOC data for exploring the relationship between discussion forum views and learning outcomes. In J. G. Boticario, & O. C. Santos (Eds.), *Proceedings of the Eighth International Conference on Educational Data Mining* (pp. 234–241). Retrieved from <https://www.educationaldatamining.org/EDM2015/proceedings/full234-241.pdf>
- Bote-Lorenzo, M. L., & Gómez-Sánchez, E. (2017). Predicting the decrease of engagement indicators in a MOOC. In A. Wise, P. H. Winne, & G. Lynch (Chairs), *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (pp. 143–147). doi: [10.1145/3027385.3027387](https://doi.org/10.1145/3027385.3027387)
- Cassidy, D., Breakwell, N., & Bailey, J. (2014). Keeping them clicking: Promoting student engagement in MOOC design. *The All Ireland Journal of Teaching and Learning in Higher Education*, 6(2), 1–15. Retrieved from <http://www.icep.ie/wp-content/uploads/2013/12/CassidyBreakwellBailey.pdf>
- Chen, Q., Luo, W., Palardy, G. J., Glaman, R., & McEnturff, A. (2017). The efficacy of common fit indices for enumerating classes in growth mixture models when nested data structure is ignored: A Monte Carlo study. *SAGE Open*, 7(1), 1–19. doi: [10.1177/2158244017700459](https://doi.org/10.1177/2158244017700459)
- Chen, Z., Alcorn, B., Christensen, G., Eriksson, N., Koller, D., & Emanuel, E. J. (2015, September 22). Who's benefiting from MOOCs, and why. *Harvard Business Review*. Retrieved from <https://hbr.org/2015/09/whos-benefiting-from-moocs-and-why>
- College Board (2016). *Trends in college pricing 2016*. Retrieved from https://trends.collegeboard.org/sites/default/files/2016-trends-college-pricing-web_1.pdf
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment*, 91(2), 166-174. doi: [10.1080/00223890802634290](https://doi.org/10.1080/00223890802634290)
- Ferguson, R., & Clow, D. (2015). Examining engagement: Analysing learner subpopulations in massive open online courses (MOOCs). In J. Baron, G. Lynch, & N. Marziarz (Chairs), *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 51-68). doi: [10.1145/2723576.2723606](https://doi.org/10.1145/2723576.2723606)
- Ferguson, R., Clow, D., Beale, R., Cooper, A. J., Morris, N., Bayne, S., & Woodgate, A. (2015). Moving through MOOCs: Pedagogy, learning design and patterns of engagement. In G. Conole, T. Klobučar, C. Rensing, J. Konert, & E. Lavoué (Eds.), *lecture notes in computer science: Vol. 9307. Design for teaching and learning in a networked world* (pp. 70-84). Springer, Cham. doi: [10.1007/978-3-319-24258-3_6](https://doi.org/10.1007/978-3-319-24258-3_6)

- Hamori, M. (2019, May 21). MOOCs at work: What induces employer support for them? *The International Journal of Human Resource Management*. doi: [10.1080/09585192.2019.1616593](https://doi.org/10.1080/09585192.2019.1616593)
- Henderikx, M. A., Kreijns, K., & Kalz, M. (2017). Refining success and dropout in massive open online courses based on the intention-behavior gap. *Distance Education*, 38(3), 353-368. doi: [10.1080/01587919.2017.1369006](https://doi.org/10.1080/01587919.2017.1369006)
- Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. *The International Review of Research in Open and Distributed Learning*, 15(1), 133-160. doi: [10.19173/irrodl.v15i1.1651](https://doi.org/10.19173/irrodl.v15i1.1651)
- Jung, Y., & Lee, J. (2018). Learning engagement and persistence in massive open online courses (MOOCs). *Computers & Education*, 122, 9-22. doi: [10.1016/j.compedu.2018.02.013](https://doi.org/10.1016/j.compedu.2018.02.013)
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, 104, 18-33. doi: [10.1016/j.compedu.2016.10.001](https://doi.org/10.1016/j.compedu.2016.10.001)
- Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 170-179). doi: [10.1145/2460296.2460330](https://doi.org/10.1145/2460296.2460330)
- Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses. *Educause Review*, 48(3), 62-63. Retrieved from <https://er.educause.edu/articles/2013/6/retention-and-intention-in-massive-open-online-courses>
- Lanza, S. T., Collins, L. M., Lemmon, D. R., & Schafer, J. L. (2007). PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 671-694. doi: [10.1080/10705510701575602](https://doi.org/10.1080/10705510701575602)
- Li, Q., & Baker, R. (2016). Understanding engagement in MOOCs. In T. Barnes, M. Chi, & M. Feng (Eds.), *Proceedings of the Ninth International Conference on Educational Data Mining* (pp. 605-606). Retrieved from https://www.educationaldatamining.org/EDM2016/proceedings/edm2016_proceedings.pdf
- Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: A comparison with *k*-means. *Canadian Journal of Marketing Research*, 20(1), 37-44. Retrieved from <https://pdfs.semanticscholar.org/6add/265688cde63766bed6b920c4546e7c11ab99.pdf>
- Moore, R. L., & Wang, C. (2020, June 12). Influence of learner motivation dispositions on MOOC completion. *Journal of Computing in Higher Education*. doi: [10.1007/s12528-020-09258-8](https://doi.org/10.1007/s12528-020-09258-8)

- Phan, T., McNeil, S. G., & Robin, B. R. (2016). Students' patterns of engagement and course performance in a massive open online course. *Computers & Education*, 95, 36-44. doi: [10.1016/j.compedu.2015.11.015](https://doi.org/10.1016/j.compedu.2015.11.015)
- Poquet, O., Dowell, N., Brooks, C., & Dawson, S. (2018). Are MOOC forums changing? In M. Hatala (Ed.), *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (pp. 340-349). doi: [10.1145/3170358.3170416](https://doi.org/10.1145/3170358.3170416)
- Ramesh, A., Goldwasser, D., Huang, B., Daume III, H., & Getoor, L. (2014). Learning latent engagement patterns of students in online courses. In C. E. Brodley, & P. Stone (Eds.), *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* (pp. 1272-1278). Retrieved from <https://dl.acm.org/doi/10.5555/2893873.2894071>
- Reich, J. (2014, December 8). MOOC completion and retention in the context of student intent. *Educause Review Online*. Retrieved from <https://er.educause.edu/articles/2014/12/mooc-completion-and-retention-in-the-context-of-student-intent>
- Waddoups, C. J. (2016). Did employers in the United States back away from skills training during the early 2000s? *Industrial and Labor Relations Review*, 69(2), 405-434. doi: [10.1177/0019793915619904](https://doi.org/10.1177/0019793915619904)
- Wang, Y., & Baker, R. (2015). Content or platform: Why do students complete MOOCs? *Journal of Online Learning and Teaching*, 11(1), 17-30. Retrieved from https://jolt.merlot.org/vol11no1/Wang_0315.pdf

