Understanding the Roles of Personalization and Social Learning in a Language MOOC Through Learning Analytics

Napat Jitpaisarnwattana
Hayo Reinders
Pornapit Darasawang
King Mongkut's University of Technology Thonburi, Thailand

Abstract

In the last decade, there has been a great deal of interest in language MOOCs (LMOOCs) and their potential to offer learning opportunities for large audiences, including those in disadvantaged communities. However, experiences and research have shown MOOCs to suffer from several challenges. Chief among these have been low participation and completion rates, which are often attributed to limitations in how opportunities for personalisation and social interaction are implemented. For the current study, a dedicated LMOOC was designed and implemented, called the "Social and Personal Online Language Course (SPOLC)." This language learning environment incorporates a recommendation system and emphasizes personalisation and social interaction. The study identified the types of learning behaviour that were related to course completion and observed how 270 learners in the LMOOC used the various course features. The data were collected using learning analytical methods and analysed using binary logistic regression and feature extraction prediction model. The results demonstrated that working in groups and creating a learning plan were important factors associated with course completion, while interacting with other learners online was not. We conclude with several suggestions and implications for future LMOOC design, implementation, and research.

Keywords: Language MOOCs, personalisation, social learning, learning analytics

There has been a great deal of interest in Massive Open Online Courses for language learning (LMOOCs), as they hold considerable potential for addressing some of the existing practical challenges in online language learning, such as issues of accessibility and affordability (Hill, 2012). The open and free nature of most LMOOCs has contributed to addressing some of these practical challenges. However, a number of pedagogical issues have emerged from MOOC implementations and research studies. These include the teacher-centric nature of many courses, low attendance and completion rates, and limited interaction among MOOC learners. Of these, low completion rates have received widespread attention and have often been cited as the scale-efficacy tradeoff of the MOOC educational model (Onah, Sinclair, & Boyatt, 2014). In the context of LMOOCs, issues of participation, completion, and interaction are often attributed to a lack of personalisation and opportunities for social interaction for learners (Perifanou, 2015).

Personalisation involves giving learners choices in learning approaches, content, and pace in order to accommodate individual learning differences. Given the heterogeneous nature of LMOOCs, personalisation is crucial as learners from different backgrounds with different needs, goals, and preferences participate. Likewise, interaction with other learners has been seen as a key component for success in online L2 learning (Yang, 2011). LMOOC environments offer opportunities for learners to interact with other learners in the course given that there are by definition both large numbers of participants and multiple communication channels, including synchronous (e.g., chat facilities) and asynchronous (e.g., forums for communication) (Sokolik, 2014). However, studies of LMOOCs have shown interaction to be quite limited (Martin-Monje, Barcena & Read, 2013; Martin-Monje, Castrillo & Rodriguez, 2018; Rubio, 2015). There is thus a need for investigating how different design elements of LMOOCs may contribute to increased interaction.

One approach that has often been adopted is the use of an adaptive learning system that offers learners personalized feedback and content sequencing. This allows learners to be directed to the most appropriate learning materials based on their profiles (Godwin-Jones, 2014; Perifanou, 2015). Such intelligent systems have been implemented in many MOOCs. However, solely providing learners with adaptive or recommended content may not be enough. Rather, such a system needs to be placed in a learning environment that is also social and personalizable by the learner (Moreira Teixeira & Mota, 2014; Sokolik, 2014). There need to be ample opportunities for learners to interact with other learners through various types of collaborative work, peer assessment, discussion forums and other communication platforms. Furthermore, the personalized LMOOCs should afford learners enough freedom to tailor the way in which they want to participate in each course, thus allowing for personal learning (Downes, 2012) as well as engagement with a personal learning environment (Godwin-Jones, 2009, 2017) to manifest. The current study investigates the Social and Personal Online Language Course, or SPOLC, a MOOC-type language learning environment that deals primarily with essential English language skills for delivering presentations. This LMOOC incorporates a recommendation system and personalizable and social aspects into its design. The study aims to observe how learners in the SPOLC make use of the learning opportunities afforded by the course design and identify the types of learning behaviour that are related to course completion using learning analytical methods.

The next section of this paper discusses the concepts of personalisation and socialisation in LMOOC contexts and provides an overview of research and practices. After this, the steps taken in designing and implementing the SPOLC will be described; the results of the data

analysis will be reported and discussed in the later sections. Finally, implications for LMOOC implementation and practical applications will be raised considering the findings.

Review of Related Literature

Language MOOCs and Their Challenges

Barcena and Martin-Monje (2014) define LMOOCs as "dedicated web-based online courses for second languages with unrestricted access and potentially unlimited participation" (p.1). Despite early proliferation, their educational model has sometimes been criticized as "problematic" for language learning (Barcena & Martin-Monje, 2014, Barcena et al., 2015; Sokolik, 2014), with the majority of LMOOCs being based on xMOOC pedagogy and focusing on transmission of knowledge. This may not be suitable for the skill-based learning that language learning requires. The essential components of language acquisition, including ample L2 input, opportunities for L2 output and a scaffolded environment for L2 interaction, appear to be missing from most of the currently available LMOOCs. Further, as anyone can enroll in LMOOCs, their demography is extremely heterogeneous. Participants differ in their proficiency levels, interests, and learning styles, which pose significant challenges for developers. Currently, LMOOCs are not yet successful in personalizing learning experiences, which may be one of the reasons for their high drop-out rates (Loizzo et al., 2017). Another important challenge is the lack of interaction and socialisation in most LMOOCs (Rubio, 2015; Schulze & Scholz, 2018), as they mostly rely on discussion forums integrated into the course and often do not incorporate other communication tools. This can prevent learners from interacting with each other (Perifanou, 2015). Therefore, we propose that it is both theoretically important and empirically feasible for LMOOCs to start addressing these issues to maximize their potential.

Personalization and Social Interaction in LMOOCs

Personalisation refers to instruction that is tailored to learning needs, preferences and interests of different learners (Downes, 2016). Efforts to improve personalisation have received increased attention in recent years, helped by developments in educational technology. LMOOC environments hold considerable potential for increasing personalisation as a result of their online infrastructure and their adaptability to different pedagogical approaches. In addition, in online platforms learners can be encouraged and supported to create their own personal learning environment (PLE), or a learner-organized language learning environment in which learners can combine digital tools and resources to support different aspects of their learning process, from goal setting to materials selection to assessment (Author, 2014). According to Attwell (2007), PLEs afford learners with opportunities to be fully involved in the learning process by allowing them to be the co-creators of their knowledge. In CALL, the notion of PLEs has been widely adopted and examined in different contexts, including online and blended courses, mobile learning (Pegrum, 2014)) and social media (Devedzic, 2016).

The vast amount of data LMOOCs generate allows for the creation of learner profiles, which can be used to direct learners to learning resources that are suitable for their proficiency levels, learning goals and content preferences (Bull & Wasson, 2016). A concrete example of this is the use of a recommendation system, in which learners are presented with suggested learning materials or learning plans based on their profiles. A recommendation system has been utilized in various studies examining different language skills such as reading ability (Hsu, Hwang & Chang, 2013) and vocabulary (Nikiforovs & Bledaite, 2012). Since the PLE notion has often been adopted under the connectivist MOOC (cMOOC) model and the recommendation system has often been associated with a more structured xMOOC model, we argue that

personalisation in LMOOCs could benefit from addressing both forms of personalisation. In other words, LMOOC personalisation should provide personalized learning in the form of recommendations based on learner profiles, but at the same time allow learners to create and personalize their own learning pathways.

Interaction has been a mainstay in online language learning. Research into interaction in online courses has provided well-documented, positive results. Several meta-analyses demonstrate that learning is more effective when interaction and collaboration are facilitated and that interaction is positively correlated with learning outcomes (Bernard, et al. 2009; Ducate & Lomicka, 2008). Although researchers and practitioners are in general agreement that interaction is crucial and forms the basis for effective practices in online language learning environments (Bernard et al., 2009; Yang, 2011), interaction is a complex phenomenon and there are several key factors contributing to its successful integration in an online language course. Types of interaction are one of these key factors. Moore (1989) identified three components of critical interaction in educational contexts: learner–content interaction (L-C), learner–instructor interaction (L-I) and learner–learner interaction (L-L). In Moore's definition, L-C interaction encompasses reading texts, watching videos, searching for information, completing assignments and working on projects. For L-I interaction, learners interact with the course instructor either synchronously or asynchronously through emails or discussion forums. In L-L interaction, learners interact with other learners either individually or in groups and such interaction often takes place using through synchronous computer-mediated communication (CMC) tools (e.g., instant messaging) as well as asynchronous computer-mediated communication tools (e.g., emails and discussion forums).

These types of interaction provide a useful framework for LMOOC instructors and designers to understand what to consider when developing and delivering an LMOOC. Moore (1989) suggests that course designers maximize each type of interaction and provide suitable types of interaction in different subject areas. We argue that in LMOOC contexts where L-C interaction is almost a necessity and its 'massive' element makes L-I extremely difficult, L-L interaction has become a key design principle. The key design feature of current LMOOCs regarding interaction centres around encouraging participants to engage in forum discussion and providing peer feedback to other participants (Martin-Monje et al., 2018; Rubio, 2015). Despite its well-documented benefits for language learning (Blake, 2009; Harrison & Thomas, 2009; Wu et al., 2011), previous LMOOC designs have not yet been successful in facilitating L-L interaction and research studies on LMOOCs and interaction are unanimous in their observation that the level of L-L interaction is still quite low (Martin-Monje et al., 2013; Rubio, 2015; Martin-Monje et al., 2018). The types of interaction investigated in these studies included both exchanges in the discussion forums and peer feedback. Therefore, facilitating L-L interaction remains a challenge for LMOOC designers.

Personalisation and Social Interaction in LMOOCs: Research and Practice

LMOOCs offer learners opportunities to interact with a large number of peers from different countries. Despite studies of interaction in LMOOCs reporting a fairly high level of L-C and L-I interaction, the level of L-L interaction both in learning activities and discussion forums, is quite low (Martin-Monje et al., 2018; Rubio, 2015). In his study, Rubio (2015) compared learners' interaction in an LMOOC with the other two formats of delivery (blended and online) and found that, in the LMOOC format, the L-L interaction was quite low compared with L-C and L-I interaction. The study also reported a positive correlation between interaction levels and course outcomes. A similar finding emerged in a study looking at online interaction

(Martin-Monje et al., 2018) in that learners who were active in their participation and interaction were more likely to be successful in the LMOOC. Interestingly, however, participation in discussion forums and providing peer feedback were not factors associated with students' success.

In terms of course design, several personalisation initiatives have been implemented in the LMOOC context. One example of this is SpanishMOOC, which incorporates Instreamia, an adaptive learning system (Godwin-Jones, 2014). The system provided personalized feedback and content sequencing to the learners. Other intelligent systems have also been implemented. The Open Learning Initiative (OLI), which makes use of cognitive and example-tracking-tutors, offers self-study learning resources in several languages. The "open learners' profiles', in which learners' interactions with the system are collected and used to develop a more effective adaptive learning system were also used (Godwin-Jones, 2014). Although these efforts to offer personalized learning in LMOOCs were a good starting point, they have not yet been investigated empirically. On the basis of the above initiatives, we can conclude that despite initial efforts, it remains unclear to what extent personalisation can contribute to language learning in LMOOC environments and enhance course completion.

The available platforms have not yet succeeded in personalizing learning experiences and providing sufficient opportunities for social interaction and there is still considerable room in the LMOOC architecture for improvement. This study tackles this challenge by reporting on the development and outcomes of a Social and Personal Online Language Course (SPOLC), a MOOC-type language learning platform, that aims to provide a personalized learning experience within a social learning environment. This study is guided by three research questions:

1. To what extent can a specialised LMOOC environment encourage learners to personalize their learning?
2. To what extent can a specialised LMOOC environment encourage learners to interact with other learners?
3. What is the correlation between learning behaviours in an LMOOC and course completion?

Method

Design of the SPOLC

The SPOLC, an LMOOC-type course, was specifically designed for this study. It was developed on *Moodle* with additional plug-ins and a recommendation system. The design of the SPOLC is grounded in two primary theoretical foundations: personalisation and social learning. For personalisation, we align ourselves with Moreira-Teixeira & Mota (2014) and Sokolik (2014), who proposed that an optimal approach to designing an LMOOC is to provide an adaptive learning or a recommendation system in a personalizable learning environment. This idea allows for the combination of personalized learning with personal learning. The former refers to learning materials suggested to learners by a computer system, while the latter refers to learners' choices and decisions in planning their learning (Downes, 2012, 2016). For social learning, the SPOLC allows learners to work either individually or in a group on the final project. Several learning activities also encourage the use of peer feedback and peer assessment using provided rubrics.

The course delivered through the SPOLC is called *Presentation@work* and aims to help learners develop their English presentation skills in either a professional or educational context. The learning architecture of the SPOLC was based on a framework for operationalization and

implementation for learner autonomy proposed by Reinders (2010), in which self-directed learning is divided into seven stages: identifying needs, setting goals, planning learning, selecting resources, selecting learning strategies, practice, monitoring progress, and assessment and revision. The learning architecture of the SPOLC is visualized in Table 1 below.

Table 1
The Learning Architecture of the SPOLC

Learning stages in the framework	Learning stages in SPOLC	Learning activities	SPOLC design principles
	Registration	Create profiles	Personalized learning (Data collected for learner model)
	Stage 0: General	• Instruction on how to learn in the course and how to use features and tools in the platform	
Identifying needs + setting goals	Stage 1: Identify the type of presentation	• Identify the type of presentation they want to do	Personalized learning (Data collected for learner model)
Identifying needs	Stage 2: Self-evaluation and Identifying what you need	• Self-evaluation • Upload videos and get feedback from peers • Reflect on past experience with the topic and identify what needs improvements	Personalized learning pathway (PLP) generated and presented to learners Peer feedback and assessment
Setting goals + Planning learning + Selecting resources	Stage 3: Planning your learning	• Set their learning goals and create their learning plan • Discuss plan with their peers (for group) • Find additional learning resources outside of SPOLC	Personal learning and group learning
Selecting learning strategies and practice	Stage 4: Learning activities	• Learn and practice with a wide range of activities • Work on presentation • Get feedback from their peers	Personalized learning Personal learning Group learning Peer feedback and assessment
Monitoring progress and assessment and revision	Stage 5: Rehearsal	• Upload final presentation for feedback for them to improve upon	Group learning Peer feedback and assessment
Monitoring progress and assessment and revision	Stage 6: Final presentation	• Upload improved presentation • Uploaded videos are rated by other learners	Group learning Peer feedback and assessment

After registering and creating a profile, learners complete a series of learning activities. In stage 0, learners familiarize themselves with the platform, its structure, and features. Then they start thinking about the type of presentation that would be most beneficial for them, ranging from English academic presentations to annual company reports to a three-minute sales pitch. In stage 2, they self-evaluate different aspects of presentation skills, including delivery, engagement, and visual aids. They also upload their first video to get feedback from other learners (based on the rubrics provided). This is when the personalized learning pathway (PLP) based on their profiles and self-evaluation is generated by the system and provided to them.

The PLP provides each learner with a unique learning pathway, including recommended learning activities and the types of activities that would be most appropriate for their perceived ability. It is created by the system based on the data from the participants' profiles and their self-evaluation results. In stage 3, learners create an Individual Learning Plan (ILP), which includes deciding on their specific goals for the project, allocating a certain amount of time every week, and choosing whether to work alone or with others. They also consider what resources other than those available within the SPOLC they want to use, such as colleagues, English-speaking friends, favorite websites, etc. In other words, the system-generated PLP identifies the most suitable activities and sequence for completing these within the SPOLC, and the ILP, is learners' chosen program of study (or to put it metaphorically, the PLP is a recommended itinerary and the ILP the travel plan learners choose to follow, including how many stops to make and what to do in each place). For those opting to work in a group, they can hold meetings with other group members through their own personal communication channels at this stage. In stage 4, learners are given complete freedom to choose any activities that they want to learn. They can either opt to follow the personalized learning pathway or follow their own learning plan or they can follow neither. They can also work on the type of presentation that is most relevant to them. In stage 5, they upload their presentation to get feedback from other learners in the form of comments. The learners can use these comments to improve their presentation before resubmitting them in stage 6 when all the presentations are rated and ranked as part of the competition.

Participants

There was a total of 403 registered participants in this course. As this LMOOC was open to anyone, the background of the participants, gathered from learners' profiles, was highly diverse. There were 133 undergraduate students (33.01%), 98 graduate students (24.31%) and 172 working professionals (42.68%), including nurses, architects, engineers, medical staff, salespersons, teachers, and researchers. Although the majority of the participants were Thai, there were participants from the Philippines, Mexico and China as well. As for gender, 253 participants were female (62.78%) and 124 were male (30.77%), while 26 participants did not identify their gender (6.45%). However, only 270 participants started the course and we only focused on these participants in this study. The participants completed a self-evaluation questionnaire of their current knowledge of delivering a presentation in English, the focus of the course. The questionnaire asked the participants to evaluate their skills related to giving a presentation in English, including language, delivery, engagement, visual aids, and overall presentation. The evaluation classified the participants into four categories: 1) need overall improvement (39.9%) 2) need improvement in some areas (15.9%) 3) overall fairly good (41%) and 4) overall very good (3 %).

Data Collection and Analysis

The data were collected over a period of five weeks between October and November 2019 and involved the use of quantitative techniques. Learning-related data were logged using

the analytics system of the MOOC platform, in which data on activity completion, time spent in the course, following/not following the personalized learning pathway, devising/not devising their own individual learning plan, type of participation (group vs. individual), and their interaction in the forums and with other learners' videos were collected. The data set was processed using Microsoft Excel software and descriptive statistics on the use of personalisation features and interaction in the MOOC were generated using SPSS. Then two statistical approaches were applied: a binary logistic regression and a feature extraction prediction model.

A binary logistic regression model was developed and performed to evaluate the relationship between each learning factor and course completion. However, participating in an LMOOC is a complex non-linear process and there are several hidden learning patterns. Therefore, machine learning techniques were utilized to develop a prediction model that can identify the learning behaviours that affect course completion. As Al-Shabandar et al. (2017) note, machine learning is an effective analysis technique that can be applied to learning analytics because it can help to discover hidden patterns of students' learning behaviours and to analyze complex, non-linear relationships. In this study, the primary data set is made up of the *clickstream,* which means learners' behaviours relating to activity completion, posts in forums, interaction with peers' videos, access time, learning pathways, learning plans, and course completion. A brief description of the dataset attributes is given in Table 2.

Table 2
Data Collected

Features	Description
Course completion	The submission of the final presentation encoded as 1 (Yes) / 0 (No)
Follow PLP	Whether the participants followed the personalized learning pathway presented to them 1 (Yes) / 0 (No)
Create an ILP	Whether the participants created their own learning plan 1 (Yes) / 0 (No)
Access Time	A collective amount of time each participant spent in the MOOC
L-L Interaction	Whether the participants interacted with other learners in forums and video comments 1 (Yes) / 0 (No)
Number of messages	The number of messages each participant contributed
Activity completion	Whether the participants completed each learning activity (60 activities in total) / encoded as 1 (completed) / 0 (not completed)
Type of work	The type of work that the participants opted to do / 1 (individual) and 0 (group)

The model developed in this paper employed various linear and non-linear supervised machine learning models based on feature extraction techniques. These models include logistic regression (LR), Random Forest (RF), Recursive Feature Elimination (RFE), Chi-square test (Chi-2), Pearson's (r), and LightGBM. The machine learning prediction model can provide a computational prediction for the type of learner who is likely to complete the MOOC based on their learning behaviours. In other words, it provides a behavioral analysis in order to predict the participants' learning outcome (operationalized as completing the course).

Results

To what extent can a specialised LMOOC environment encourage learners to personalize their learning?

In investigating how the participants personalized their learning, the data were generated by the course's learning analytics tool, on which descriptive statistics were performed. Table 3 shows whether the participants followed the personalized learning pathway (PLP) provided to them at the beginning of the course.

Table 3
Participants' Use of the Personalized Learning Pathway (PLP)

Number of students	Followed PLP	Did not follow PLP
270	78 (28.9%)	192 (71.1%)

The majority of the participants (71.1%) chose not to follow the PLP provided to them, while only 28.9 % did so. Also, as described above, participants had a further choice—whether to complete their individual learning plan (ILP). The data on whether the participants created an LP is depicted in Table 4 below:

Table 4
Participants' Creation of an Individual Learning Plan (ILP)

Number of students	Created ILP	Did not create ILP
270	155 (57.4%)	115 (42.6%)

More than half of the participants created their ILP for the course, whereas slightly more than 40% opted not to. From the above, four different personalisation patterns are possible: follow PLP and create ILP, follow PLP but not create ILP, not follow PLP but create ILP, and neither follow PLP nor create ILP. The descriptive data on these four personalisation patterns are presented in table 5 below:

Table 5
Types of Personalisation

| Number of students	Follow PLP and create ILP	Follow PLP but not create ILP	Not follow PLP but create ILP	Neither follow PLP nor create ILP
270	68 (25.2%)	10 (3.7%)	87 (32.2%)	105 (38.9%)

As shown, the largest proportion (38.9%) of the participants did not follow the personalized learning plan provided to them, nor created their individual learning plan (as visible in the course analytics). A slightly smaller number of participants (32.2%) chose not to follow the PLP, but devised their ILP, while only 3.7 % of the participants followed the PLP without creating their ILP learning plan. Further, a quarter of the participants opted to use both features. These results demonstrated that although the participants were not so keen on following the provided PLP, creating an ILP was a fairly popular personalisation feature. This also suggests that when given choices, participants were more likely to "personalise" their own learning (ILP) rather than following the recommended pathways (PLP).

To what extent can a specialised LMOOC environment encourage learners to interact with other learners?

The course design allowed the participants two options for learning in the course: working individually or working as a group. The group learning option allowed participants to either form a group with their colleagues and join the course together or form a group with other learners online. It was found that a larger number of the participants opted to work as a group than to work individually at 61.1% (n = 165) and 38.9% (n = 105) respectively. Of those working as a group, the majority joined the course with their colleagues (94.54%), while only 5.46 % formed a group online. In addition, the course design provided the participants with several interaction opportunities including commenting on other learners' videos, participating in discussion forums and posting in a Facebook group. There was a total of 677 posts from the participants over the five-week period, or an average of 2.51 posts per person. The median number of posts was two and the mode was one, meaning that most of the participants posted only once. These posts were classified according to three different interaction channels. The majority of posts (93%) (n= 630) was in the form of comments on the videos of other learners, meaning an average of 0.46 comments per person per week, while only a very small number of posts were present in the discussion forums and the Facebook group at 1.8 (n = 12) and 5.2 % (n =35) respectively. The frequencies of the posts mean that the design of the current LMOOC could not encourage the majority of the participants to interact with other participants. Another important thing to take into consideration is how the interaction levels were spread across different phases of the course. The results are illustrated in Figure 1.

Figure 1

Pattern of Posts in Three Interaction Channels
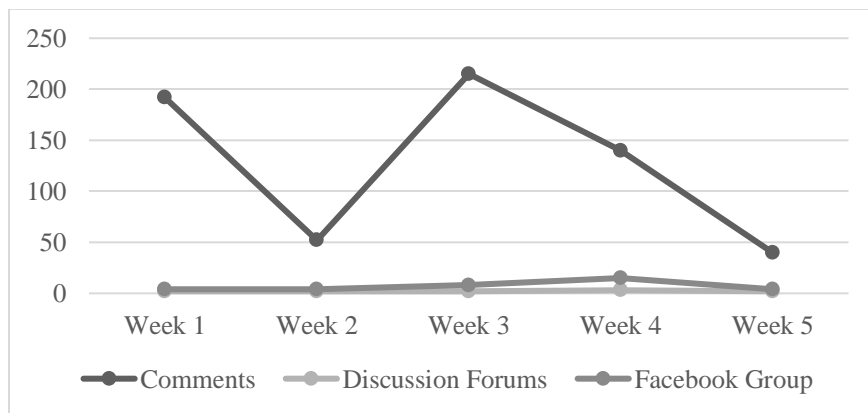
It is clear from the data that the pattern of the participants' comments coincides with the type of activities they engaged each week. Learning activities in weeks 1 and 3 encouraged the participants to give feedback on their peers' videos, whereas in week 2 most of the activities were individual. However, it is worth noting that there was a sharp decline in the number of posts in weeks 4 and 5 despite having similar learning activities as weeks 1 and 3. The number of posts in the Facebook group and the discussion forum were low across the weeks. The spread of the posts showed that the type of learning activities and the stages of the LMOOC might be factors affecting the participants' choices to interact with others in the course.

What is the correlation between learning behaviours in an LMOOC and course completion?

Of the 270 participants who started, 180 went on to complete the course (operationalized as submission of the final presentation), while 90 dropped out after starting the course—most (73.33%) in weeks 2 and 3. This gives the course a completion rate of 66.6%. This is, of course, a good completion rate compared with other LMOOCs and MOOCs in general. What is more interesting, however, is which factor(s) contributed to the participants completing the course. This section investigates this using two statistical techniques: a binary logistic regression and a computational machine learning prediction model.

Logistic regression analysis

The logistic regression model was computed to investigate the factors that are statistically associated with completing the course. The model was developed based on two sets of data: the characteristics of the participants (e.g., following a personalized learning pathway or working as a group) and participation in learning activities (e.g., completing learning activity 1.1). The analysis of the participants' characteristics is presented in Table 6 below:

Table 6

Logistic Regression Analysis of the Characteristics of the Participants

Characteristics	B	S.E.	Wald	df	Sig.	Exp(B)
Following the PLP	17.462	2724.069	0	1	0.995	38323829
Creating the ILP	4.777	1.41	11.485	1	0.001*	118.785
Time spent	0.025	0.013	3.815	1	0.051	1.026
Interaction	-4.741	4.124	1.322	1	0.250	0.009
Number of messages	2.228	1.891	1.387	1	0.239	9.278
Type of participation	-6.858	1.442	22.624	1	0.000*	0.001
Learning	-0.379	12.816	0.001	1	0.976	0.684
Forum						
Constant	-11.347	2724.07	0	1	0.997	0

Note: B = the coefficient for the constant, S.E. = the standard error around the coefficient for the constant, Wald = Wald chi-square test, Df = degree of freedom Sig. = Significant, Exp (B) = the exponentiation of the B coefficient

It can be seen from the analysis that creating an ILP and the type of participation are statistically significant to course completion (0.05). This means that the participants who created their own personal learning plan had a higher likelihood of completing the course. The negative coefficient in the type of participation means that the participants who opted to work as a group were more likely to complete the course than those who worked individually. However, other factors

including time spent in the LMOOC, following the PLP, interacting with other learners, the number of messages they posted, and participating in the learning forums did not statistically affect course completion.

In addition to the characteristics of the participants, participation in the learning activities is another important factor. Table 7 shows the results of the logistic regression analysis.

Table 7
Logistic Regression Analysis of the Participation in Each Learning Stage

Learning stages	B	S.E.	Wald	df	Sig.	Exp(B)
Learning stage 0	.047	0.116	0.165	1	0.685	1.048
Learning stage 1	1.194	0.965	1.533	1	0.216	3.302
Learning stage 2	0.596	0.163	13.431	1	0.000*	1.814
Learning stage 3	-0.159	0.178	0.792	1	0.374	0.853
Learning stage 4	0.047	0.033	2.009	1	0.156	1.048
Learning stage 5	-0.475	0.282	2.842	1	0.092	.622
Learning stage 6	0.581	0.156	13.782	1	0.000*	1.787
Constant	-3.763	.829	20.623	1	0.000*	0.023

Note: B = the coefficient for the constant, S.E. = the standard error around the coefficient for the constant, Wald = Wald chi-square test, Df = degree of freedom Sig. = Significant, Exp (B) = the exponentiation of the B coefficient

The analysis shows that participating in learning stage 2 (doing self-evaluation and uploading a presentation for feedback) is statistically related to the participants completing the course (Sig. < 0.05), meaning that participants who complete activities in learning stage 2 are more likely to complete the course (learning stage 6 is the submission of the final presentation). It should be noted that completing learning stage 5 (Rehearsal) also gives the participants a higher likelihood of completing the course, though less so than the first two variables (Sig. < 0.1). Nevertheless, completing activities in learning stages 0, 1, 3, and 4 does not affect course completion. In addition, a logistic regression analysis was performed with each learning activity in each learning stage (n = 54). The results of the analysis are shown in table 8 below:

Table 8
Logistic regression analysis of participation in each learning activity

Learning Activities	B	S.E.	Wald	df	Sig.	Exp(B)
Uploading the presentation for feedback (2.3.1)	4.030	1.394	8.352	1	0.004	56.528
Evaluating yourself (2.2)	3.593	1.777	3.964	1	0.046	34.416
Instruction for the final presentation (6.1)	2.513	1.427	3.101	1	0.078	12.344
How good is a good presentation (4.1.2)	3.436	2.092	2.697	1	0.101	31.017
Watch and rate your peers' videos (2.3.2)	-2.600	1.682	2.389	1	0.122	0.74
Constant	-3.763	.829	20.623	1	0.000	0.023

Note: B = the coefficient for the constant, S.E. = the standard error around the coefficient for the constant, Wald = Wald chi-square test, Df = degree of freedom Sig. = Significant, Exp (B) = the exponentiation of the B coefficient

The results demonstrate that uploading their presentation for feedback and self-evaluation are statistically significant to participants completing the course (Sig. < 0.05), meaning that participants who self-evaluated and uploaded their first presentation for feedback were more likely to complete the course than those who did not. However, participating in other learning activities did not statistically significantly affect course completion.

Feature of Importance Prediction Model

Participating in an LMOOC is a complex, non-linear process and there are patterns that may be hidden. To identify these, a machine learning prediction model, using several feature extraction techniques was developed to provide a more comprehensive analysis of the participants' behaviours. As Al-Shabandar et al. (2017) posit, a machine learning model can be an effective technique to discover hidden patterns of students' learning behaviours and to analyze complex, non-linear relationships in MOOC context. The building of such a prediction model could also show a more holistic picture of factors that may lead to learners completing the MOOC. The techniques applied in the model include Pearson correlation, Chi-square, recursive feature elimination (a feature selection technique), random forest (a type of decision tree algorithm), LightGBM (another type of decision tree algorithm) and logistic regression. The results of the analysis are presented in Table 9 below.

Table 9
The Feature Extraction Techniques and Results

Techniques	Features	Sig.*	Coefficient	Feature of Importance	Ranking
Pearson	*Type of work*		0.857		
	Creating an ILP		0.773		
	Learning activity 2.2		0.687		
	Learning activity 2.4		0.666		
	Learning activity 2.1		0.650		
Chi-Square	*Creating an ILP*	0.01			
	Interaction	0.01			
	Type of work	0.01			
	Learning activity 6.1.1	0.01			
	Learning activity 6.1.2	0.01			
RFE	*Following a PLP*				1
	Creating an ILP				1
	Interaction				1
	Type of work				1
	Learning activity 6.1.1				1
RF	*Creating an ILP*		0.159		
	Type of work		0.146		
	Time spent		0.121		
	Learning activity 2.2		0.063		
	Learning activity 2.3		0.033		

LightGBM	*Time spent*			244***	
	Creating an ILP			66	
	Type of work			58	
	Learning activity 0.1			58	
	Learning activity 0.4			57	
LogisticR	*Creating an ILP*		2.994		
	Type of work		2.245		
	Following a PLP		0.875		
	Learning activity 4.1.10		0.614		
	Interaction		0.527		

* Statistically significant at level 0.01
** only five most important features are presented
*** More numbers = more important

It is clear from the table that these feature extraction techniques yielded different results and each technique required different statistical interpretation of the importance of each of the features. For Pearson correlation, the analysis suggests that the type of work and creating a personal learning plan are the most important features affecting course completion, followed by the three learning activities. In addition, despite the type of work and creating an ILP being important, Chi-square analysis considers interaction in the course and participating in learning activities 6.1.1 and 6.1.2 as important features. Recursive feature elimination (RFE) is an algorithm that selects features of importance by recursively considering smaller and smaller features. In the process, the least important features are eliminated until the desired number of features is reached. The analysis demonstrates that the five most important features are following the PLP, creating a personal learning plan, type of work, interaction, and learning activity 6.1.1.

Random Forest (RF) is a type of decision tree algorithm that offers importance scores based on the reduction of criterion. The analysis shows that creating a learning plan, the type of work, and time spent in the LMOOC are the three most important features. Another algorithm included in creating the model is LightGBM, a type of decision tree algorithm. The model demonstrates that time spent in the LMOOC is the most important feature, followed by creating a personal learning plan, type of work, and learning activities 0.1 and 0.4 respectively. The final technique utilized was the logistic regression model, which showed similar results; creating a learning plan and the type of work the participants chose were the most important features. Subsequently, these six models were combined to create a prediction model for the types of learning behaviours that are likely to lead to completing the LMOOC. The model is illustrated in Table 10 below:

Table 10
The Prediction Model for Learning Behaviours in the Course

Features	**Pearson**	**Chi-Square**	**RFE**	**RF**	**LightGBM**	**LogisticR**
Type of work	Yes	Yes	Yes	Yes	Yes	Yes
Creating an ILP	Yes	Yes	Yes	Yes	Yes	Yes
Time spent	No	No	No	Yes	Yes	No

Interaction	No	Yes	Yes	No	No	No
Following a PLP	No	No	Yes	No	No	Yes
Learning Activity 2.2	Yes	No	No	No	Yes	No

As shown in Table 8, in this prediction model, only the type of work and creating an ILP are statistically associated with participants completing the course (i.e., they are considered important in all the models), while other features do not seem to be a probable predictor for course completion. It is interesting, perhaps, to discover that none of the learning activities are important features for course completion. From a learning analytics perspective, it is possible to say that, in this LMOOC, the participants who created their individual learning plan and who opted to work in a group are more likely to complete the course than those who did not.

Discussion

This study has attempted to determine how participants make use of the personalisation and interaction opportunities in an LMOOC and to identify the types of learning behaviours that are likely to lead to course completion. Regarding personalisation opportunities, participants were far less likely to follow a personalized learning pathway (PLP) (through a recommendation system) than to create their own individual learning plan (ILP). There are many factors that might influence this: individual preferences, expectations, or even the practicality of following the recommended plan. This, to a certain degree, resonates with Downes (2012, 2016), who argues for the importance of personal learning in the MOOC education model and reminds us that individual preferences might outweigh statistically oriented recommendations such as adaptive learning. Moreover, from an evaluative perspective, the fact that only about a quarter of the participants (28.9%) chose to follow the recommended learning pathway suggests that the pathway might not fit with what they needed in terms of the types of presentation they wanted to deliver, the number of activities they had to complete, and the amount of time they needed to invest in following the plan. Besides, over a third of the participants (38.9%) opted for neither option, a choice that was associated with diminished likelihood of completing the course.

In terms of social interaction opportunities, it is evident that the participants were active in commenting on their peers' videos, but not in the discussion forum and Facebook group. One possible explanation is that commenting on other participants' videos was seen as a part of the whole learning journey, while engaging in the forums and Facebook group was regarded as an extra activity, requiring additional effort. Furthermore, communicating in English might be a challenge for many participants, which may have prevented them from contributing more (something also noted in Sokolik (2014) and Martin-Monje et al., (2018). This might have been different if there had been minimum requirements for registration (e.g., B2 on CEFR level). Taking a more cultural perspective, since the majority of the participants are Thai, it might appear "unnatural" or "awkward" for them to communicate with other Thais in English beyond giving feedback, something we have observed in our own teaching in the country. In addition, despite a moderate number of posts per participant (2.51), the mode number of posts was still very low (N=1). This means that though some participants were active in posting comments, the majority of the participants were not.

As for the learning behaviours contributing to course completion, both the logistic regression analysis and the feature extraction prediction model yielded a similar result; the type of participation (working in group) and creating an ILP were the two factors that were statistically significantly associated with course completion. Regarding group learning, the collaborative experience that the participants had with their groups might have motivated them to keep learning in the LMOOC. Previous studies have shown that group learning could not only increase students' satisfaction, but also reduce drop-out rates (Sanz-Martínez et al., 2017; Bayeck, 2016). However, it is interesting to discover that participants' interaction in the course did not contribute significantly to course completion. This is contrary not only to our previous assumption when designing the course that L-L interaction should be a key feature of an LMOOC, but also with research in MOOCs in general that participation in forum discussions is a good indicator of course completion (Martin-Monje, 2017; Goldwasser et al., 2016). In the case of creating an individual learning plan, it is clear that providing the participants with the freedom to personalize their learning could encourage them to complete the course. The fact that participants can take different learning paths that lead to completion might give them a sense of "making learning your own'," keeping them in the LMOOC until completion. This analysis also empirically confirms Martin-Monje et al.'s (2018) contention that the LMOOC structure should be flexible and include numerous options to cater to a wide variety of participants. Since personalisation and social learning are imperative in LMOOC contexts, it is perhaps possible that there is an interplay between these two contributing factors and that the collaborative process within a personalizable learning environment is key to learning in such an environment. This relationship, however, needs to be investigated further in future studies.

Limitations and Conclusion

There are some limitations of this study that should be pointed out. First, although the LMOOC could be registered for by anyone in the world, the current demography is still largely localized, with most of the participants being Thai. Therefore, LMOOC designers should be cautious about adopting this design in other contexts. Also, as this LMOOC, to a certain extent, served as a laboratory to investigate a design concept, the number of LMOOC participants was smaller than in regular LMOOCs and as such the results may not be generalizable. Further studies might want to adopt the design principles of the current study and implement them with a larger group of participants and in different contexts.

In sum, this study examined the effects of personalisation and social learning on course completion in an LMOOC. Clearly, working in groups and creating an individual learning plan were important factors associated with course completion. Though the link is clear, it may be a stretch to claim that there is a causal relation between the two. What we can say, however, is that those participants who took up the personalisation and social learning opportunities were more likely to complete the course. The results relating to *personal* learning suggest that future LMOOC designers should consider making LMOOCs more flexible in terms of their course structure. Also, as the demography of LMOOC participants is becoming more diverse globally, it is advisable that future LMOOCs provide more options for participants to select different pathways for their learning.

Declarations

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

The authors received approval from an ethics review board for this study.

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Al-Shabandar, R., Hussain, A., Laws, A., Keight, R., Lunn, J., & Radi, N. (2017, May). Machine learning approaches to predict learning outcomes in Massive open online courses. *In 2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 713-720). IEEE. https://doi.org/10.1109/IJCNN.2017.7965922

Attwell, G. 2007. Personal learning environments—the future of eLearning? *eLearning Papers* 2(1): 1–8.

Barcena, E., & Martin-Monje, E. (2014). Language MOOCs: An Emerging Field. In E. Martin-Monje & E. Barcena (Eds.), *Language MOOCs: Providing learning, transcending boundaries* (pp. 1–10). Berlin: De Gruyter Open. https://doi.org/10.2478/9783110420067.1

Bayeck, R. Y. (2016). Exploratory study of MOOC learners' demographics and motivation: The case of students involved in groups. *Open Praxis*, *8*, 223–233. https://doi.org/10.5944/openpraxis.8.3.282

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A. & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, *79*(3), 1243–1289. https://doi.org/10.3102/0034654309333844

Blake, C. (2009). Potential of text-based internet chats for improving oral fluency in a second language. *The Modern Language Journal*, *93*(2), 227–240. https://doi.org/10.1111/j.1540-4781.2009.00858.x

Bull, S., & Wasson B. (2016). Competence visualisation: Making sense of data from 21st-century technologies in language learning. *ReCALL, 28*(2), 147–165. https://doi.org/10.1017/S0958344015000282

Devedzic, V. (2016). Not fade away? Commentary to paper education and the semantic web. *International Journal of Artificial Intelligence in Education*, *26*(1), 378–386. https://doi.org/10.1007/s40593-015-0051-2

Downes, S. (2016). Personal and Personalized Learning. *EMMA Newsletter*, Online, European Multiple MOOCs Aggregator. https://www.downes.ca/cgi-bin/page.cgi?post=65065

Ducate, L., & Lomicka, L. (2008). Adventures in the blogosphere: From blog readers to blog writers. *Computer Assisted Language Learning*, *21*(1), 9–28. https://doi.org/10.1080/09588220701865474

Godwin-Jones, R. (2009). Personal learning environments. *Language Learning & Technology*, *13*(2), 3–9.

Godwin-Jones, R. (2014). Global reach and local practice: The promise of MOOCS. *Language Learning & Technology, 18*(3), 5–15.

Godwin-Jones, R. (2017). Scaling up and zooming in: Big data and personalization in language learning. *Language Learning & Technology*, *21*(1), 4–15

Goldwasser, M., Mankoff, C., Manturuk, K., Schmid, L., & Whitfield, K. (2016). Who is a student? Completion in Coursera courses at Duke University. *Current Issues in Emerging eLearning, 3,* 125–137.

Harrison, R., & Thomas, M. (2009). Identity in online communities: Social networking sites and language learning. *International Journal of Emerging Technologies and Society, 7*(2), 109–124.

Hill, P. (2012). Four Barriers That MOOCs Must Overcome to Build a Sustainable Model. http://mfeldstein.com/four-barriersthat-moocs-must-overcome-to-become-sustainable-model/

Hsu, C.K., Hwang, G.-J., & Chang, C.-K. (2013). A personalized recommendation-based mobile learning approach to improving the reading performance of EFL students. *Computers & Education, 63*, 327–336. https://doi.org/10.1016/j.compedu.2012.12.004

Loizzo, J., Ertmer, P., Watson, W., & Watson, S. (2017). Adults are self-directed and determined to set and achieve personal learning goals in MOOCs: learners' perceptions of MOOC motivation, success, and completion. *Online Learning*, *21*(2) https://doi.org/10.24059/olj.v21i2.889

Martın-Monje, E., & Barcena, E. (Eds.). (2014). *Language MOOCs: Providing learning, transcending boundaries.* Berlin: De Gruyter Open. https://doi.org/10.2478/9783110420067

Martın-Monje, E. (2017). Mind the L in LMOOCs: The importance of language learning in massive courses. Paper presented at the DILLE Conference, Venice, Italy. https://www.researchgate.net/publication/313368938_Mind_the_L_in_LMOOCs_The_importance_of_language_learning_in_massive_courses

Martin-Monje, E., Castrillo, M. D., & Mañana-Rodigrez, J. (2018) Understanding online interaction in language MOOCs through learning analytics. *Computer Assisted Language Learning, 31*(3), 251–272. https://doi.org/10.1080/09588221.2017.1378237

Moore, M. (1989). Editorial: Three types of interaction. *The American Journal of Distance Education, 3*(2), 1–7. https://doi.org/10.1080/08923648909526659

Moreira Teixeira, A., & Mota, J. (2014). A proposal for the methodological design of collaborative language MOOCs. In E. Martin-Monje & E. Bárcena (Eds.), Language MOOCs: providing learning, transcending boundaries (pp. 33–47). Berlin: De Gruyter Open. https://doi.org/10.2478/9783110420067.3

Nikiforovs, P., & Bledaite, L. (2012), Personalized sequential language vocabulary learning recommender system. https://pdfs.semanticscholar.o/d901 /b5dcf47db028f1e56b60b931b72d8767cb6a.pdf

Onah, D.F.O., Sinclair, J., & Boyatt. (2014). Dropout rates of massive open online courses: behavioral patterns *Proceedings of the 6th International Conference on Education and New Learning Technologies (EDULEARN14)*: 1–10. https://doi.org/10.13140/RG.2.1.2402.0009

Pegrum, M. (2014). *Mobile learning: Languages, literacies, and cultures*. Palgrave Macmillan. https://doi.org/10.1057/9781137309815

Perifanou, M. (2015). Research report on the current state of language learning MOOCs worldwide: exploration, classification and evaluation. *LangMOOC project*. http://www.langmooc.com/?cat=7

Reinders, H. (2010). Towards a classroom pedagogy for learner autonomy: A framework of independent language learning skills. *Australian Journal of Teacher Education*, 35(5), 40-55.

Reinders, H. (2014). Personal learning environments for supporting out-of-class language learning. *ELT Forum*, 52(4), 14-19.

Rubio, F. (2015). The role of interaction in MOOCs and traditional technology-enhanced language courses. In E. Dixon & M. Thomas (Eds.), *Researching language learner interactions online: From social media to MOOCs. Computer Assisted Language* Instruction Consortium Series, *13* (pp. 63–68). CALICO.

Sallam, H., Martín-Monje, E., & Yan, L. (2020). Research trends in language MOOC studies: a systematic review of the published literature (2012-2018), *Computer Assisted Language Learning.* 33 1-28. https://www.tandfonline.com/doi/full/10.1080/09588221.2020.1744668

Sanz-Martínez, L., Martínez-Monés, A., BoteLorenzo, M., L., Muñoz-Cristóbal, J. A., y Dimitriadis, Y. (2017). Automatic group formation in a MOOC based on students' activity criteria. *Proceedings of the 12th European Conference on Technology Enhanced Learning, EC-TEL 2017.* 179-193. https://doi.org/10.1007/978-3-319-66610-5_14

Sokolik, M. (2014). What constitutes an effective language MOOC? In E. Martin-Monje & E. Barcena (Eds.), *Language MOOCs: Providing learning, transcending boundaries.* (pp. 6–32). Berlin: De Gruyter Open. https://doi.org/10.2478/9783110420067.2

Schulze, M., & Scholz, K. (2018). Learning trajectories and the role of online courses in a language program. *Computer Assisted Language Learning, 31*(3), 185–205. https://doi.org/10.1080/09588221.2017.1360362

Wu, W., Yen, L. L., & Marek, M. (2011). Using online EFL interaction to increase confidence, motivation, and ability. *Journal of Educational Technology & Society, 14*(3), 118-129.

Yang, Y. F. (2011). Engaging students in an online situated language learning environment. *Computer Assisted Language Learning*, *24*(2), 181-198. https://doi.org/10.1080/09588221.2010.538700