What are the best predictors of MOOC performance - demographics, assigned groups, or student motivation?

Abstract

Zhang, et al (in press) investigated the degree to which placing learners in groups and demographic variables affect performance in Massive Open Online Courses (MOOCs). The variables included the type of group (synchronous text, asynchronous text, synchronous video and audio, and control), age, gender, country of origin, English proficiency, education level, and intention to complete all, most, or some course assignments. Groups were created based on the participants’ grouping preferences and preferred time to study. Statistically significant differences were found in the preferred mode of communication based on gender, English proficiency, and education level. However the only variable (other than intention to complete) identified as having a significant effect on course completion was age, when learners were categorized as over or under 40 years old, and we found out that the group of over 40s performed better. There is no significant difference in achievement levels between the grouped and non-grouped participants. In this paper we develop a model to predict MOOC completion based on these variables, including motivations of taking this MOOC, previous online learning experience and etc.

Introduction

The Massive Open Online Course (MOOC) is becoming an intriguing online learning platform that attracts millions of enrollments. Unfortunately, it is also known for its high attrition rates. According to Liyanagunawardena and Williams (2013), most MOOCs have completion rate less than 10%. Many factors cause drop outs in MOOCs,
including participants with no intention to complete the course, participants’ lack of time and support (Hew & Cheung, 2014; Kellogg, Booth, & Oliver, 2014), participants’ change of job or location (Onah, Sinclair, & Boyatt, 2014; Gütl, Rizzardini, Chang & Morales, 2014).

Different MOOCs have taken various approaches to predict participants’ course completion and performance, and their findings have varied across these studies. Cisel (2014) indicated learner performance in MOOC is highly correlated with learners’ geographic location, employment status and time constraints. According to Cisel (2014), learners from high HDI (Human Development Index) countries and learners who are not working are more likely to complete the course. Guo and Reinecke (2014) stated that old students from low teacher-students ratio area are more likely to complete the MOOC. Morris, Hotchkiss and Swinnerton (2015) also identified four factors related to MOOC learners’ performance: age, prior online learning experience, educational level and employment status. They found that unemployed, older learners who received higher levels of education and with previous online learning experience tend to complete more of the course. On the other hand, Breslow et al. (2013) found out that age and gender did not affect students’ achievement or persistence in MOOC. According to Wang and Baker (2015), MOOC completers are interested in the course contents, while the non-completers are more interested in the MOOC experiences.

Although researchers have focused on understanding the set of potential factors that might impact participants’ MOOC completion and performance, there is not yet an accurate and well-defined model involving these characteristics that could be used by MOOC designers and instructors as a way to estimate or improve participants’ MOOC
completion rates. This scenario sets the stage to the focus of this paper: the development of a model to predict MOOC completion based on important variables as described above, and also including participants’ motivations of taking a MOOC and their previous online learning experience. Thus, we expect that this model can be used in different MOOCs and can be further refined with the interplay between research and practice.

**Theoretical framework/Literature review**

Several key factors have been identified in our study to predict learner performance and retention in MOOCs, which includes age, gender, education level, English proficiency, participation in group work, motivations of taking MOOCs, previous online learning experiences, and etc. Some of these factors are discussed below included in the MOOC completion model are generated from these variables. Social presence, teacher presence (Garrison et al., 1999; Garrison, 2007), and peer learning (Collison et al., 2000) are important factors contributing to successfully online learning experience. In a MOOC with thousands of enrolled students, there is not an opportunity for students to receive one-to-one instruction from the instructor (Pappano, 2012). Peer support and peer assessment are suggested by Yuan and Powell (2013) as effective ways to assist learning in a MOOC. Hiltz (1998) suggested assigning participants to work in groups to create an online community and enhance learning.

**Groups.** Many benefits of working in online groups have been documented in the literature. By working in a virtual team, learners have more flexibility of choosing the time and place to work online (Collis & Moonen, 2001). Guàrdia (2013) highlighted the importance of peer assistance and assessment in group work to facilitate collaborative learning and enhance the learners’ critical thinking skills. By getting involved in a community of practice (Lave & Wenger, 1991), participants feel more obligated and
committed to complete the course, meanwhile they also feel socially and emotionally engaged.

Walther & Boyd (2002) highlighted the importance of virtual teams from the following perspectives: “functions” (virtual teams can be formed in any public online media platform); “informational support” (learners receive answers, feedback and other resources from their peers); and “group support” (online activities replicate face-to-face activities with numerous participants).

**Presence.** Virtual teams also present “Social presence”, “cognitive presence” and “teaching presence”, which are crucial features of a community of inquiry (Garrison et al., 1999; Garrison, 2007). Social presence includes emotional support, cognitive presence helps the learners with deep understanding of the course topics through knowledge acquisition, negotiation and creation, and advanced learners offer “teaching presence” by helping less advanced learners in a team (Golbeck & El-Moslimany, 2013). Learners benefit from group work by getting advice, feedback and critique from others (Aalst, 2013).

Hmelo-Silver, Rosé and Levy (2014) implemented facilitation in group discussions in MOOCs and suggested combining Learning Analytics and Learning Sciences to identify students who need extra help and provide support for small groups in MOOCs. Discussion forums have been effectively used for peer support in MOOCs (Kellogg, Booth, & Oliver, 2014; Brown et al., 2015). Brown et al. (2015) also found out that a learner’s performance is highly correlated with their close friends’ in the network of MOOC discussion forum, and the frequency and quality of their talks are more like their friends’. By creating small groups and providing social support among peers in
MOOCs, we assess the differences in performance among those groups in combination with other factors (demographics, intention and motivations) in taking MOOCs as well as previous online learning experience.

**Methodologies**

In the Zhang, et al. (in press) data files, many of these variables are present, and it is possible to use this data to develop a MOOC completion prediction model. The purpose of this paper is to develop a model that will allow MOOC providers to better understand and/or predict retention and completion MOOCs. For that, this study proposes these main research questions:

1. What are the impacts of learner demographics, motivations and assigned groups on learners’ MOOC completion?

2. What indicators can best be used to predict learner’s course completion in MOOC?

**Data Sources:**

**Pre-course survey.** A call for participants of this grouping study was sent out to this Creativity, Innovation and Change (CIC) MOOC community, and 770 participants returned pre-course surveys (Zhang et al.). When combined with course performance data from Coursera, 655 valid student records were produced containing pre-course survey and course performance data. Pre-course survey data included demographic information such as gender, age, level of education, level of English proficiency, previous online learning experience, and employment status, as well as learners' motivations/reasons for taking the course.
**Coursera Data.** Original course completion data were collected through Coursera with three levels of completion: none, normal and distinction. In this paper, we summarize the completion into two levels: Complete (normal and distinction) and Non-Complete.

**Grouping Study Records.** Post course survey was distributed to grouped learners (Zhang et al) with 115 returned surveys, containing questions, such as what tools did you use to complete the course, how often did your group meet? To what extent did groups help you complete the course? The pre-course survey collects data includes the participants’ demographic information, intention of taking this course, grouping preference, and etc., meanwhile the post-course survey was conducted to gather learners’ group working experience in this course.

Based on the participants’ grouping preferences, we assigned them into groups following the order of their preferred language to communicate within a team, intention of completion, and mode of communication (synchronous text, asynchronous text, or synchronous video and audio) (Zhang et al.). Two general types of groups were formed containing grouped participants assigned into the treatment group and non-grouped participants (control group). Within the grouped participants, forty-two sub-groups were formed. At the very beginning of the course, grouping notification emails (see appendices for details) were sent out to the learners. In consideration of the large number of Chinese participants volunteering in this grouping study, we also translated the emails and instructions into Chinese as well. We suggested various online tools for different types of groups, for instance, we recommend MOOC discussion forum and email for asynchronous communication and Skype, QQ (Chinese instant messaging tool) for
synchronous communication. Meanwhile, we purchased ZOOM – a video conferencing tool for large group (mainly for control group) discussion.

**Data Analysis:** The pre-course and post-course survey data were exported from Qualtrics, students’ performance records were collected through Coursera, and those different data sets were retrieved and combined together using SQL database. The data analysis and its graphical representation were produced in ArcMAP and SPSS.

**Findings**

**Participant demographics and Grouping Preferences**

This CIC MOOC is the first English MOOC offered entirely in Chinese, for the sake of eliminating the language barriers and facilitating Chinese MOOC learners’ course comprehension. Table 1 shows that the Chinese learners account for the largest number of volunteers participated in our grouping study (25.2%) followed by learners from the United States (17.9%). The total number of Chinese enrollments in the entire CIC MOOC was 14442 (24%).

![Figure 1. Location and distribution of the participants](image-url)
As shown in figure 1, the majority of our participants were from North America, Europe and East Asia with the dots indicating the density of the distribution. The top five countries where the participants come from are China, United States, India, Mexico, and Canada. The course presented a larger percentage of female (61.3%) than male (38.7%) participated in our study, with four English levels: Poor (6.1%), Basic (23.7%), Fluent (41.7%) and Native (28.5%). There are six age levels of participants in our study, and 10.8% are from age 10-19, 35.8% are from age 20-29, 20.5% are from age 30-39, 14.7% are within age 40-49, and the rest are at age 60 and above. Sixty-five percent of our participants informed their intention of completing the entire course.
Participants’ first priority of grouping preferences were to work with people who spoke the same language as theirs. The second grouping preference was their intention of completing the course, e.g. how motivated they were to complete the course. The third preference was to set up a common time for team members to meet. However, accordingly to the post-course survey results, many participants complained about the difficulties of arranging online meetings due to the time zone differences and schedule conflicts. Other grouping preference factors include the education level, country they live in, age groups, and professions.

Table 3

<table>
<thead>
<tr>
<th>Variables in MOOC Completion Equation</th>
<th>B</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Level</td>
<td>.007</td>
<td>1.007</td>
</tr>
<tr>
<td>Education Level</td>
<td>-.084</td>
<td>.920</td>
</tr>
<tr>
<td>Age</td>
<td>.278**</td>
<td>1.320**</td>
</tr>
<tr>
<td>Personal Interest</td>
<td>.097</td>
<td>1.101</td>
</tr>
<tr>
<td>Connect with Others</td>
<td>.008</td>
<td>1.008</td>
</tr>
<tr>
<td>Institution</td>
<td>.173*</td>
<td>1.189*</td>
</tr>
<tr>
<td>Professors</td>
<td>-.290**</td>
<td>.748**</td>
</tr>
<tr>
<td>Earn Certificate</td>
<td>.110</td>
<td>1.117</td>
</tr>
<tr>
<td>Related to Program</td>
<td>.099</td>
<td>1.104</td>
</tr>
</tbody>
</table>
Friends Take | .104 | 1.109
Intent to Complete | .472*** | 1.604***
Gender | .041 | 1.042
Groups | .093 | 1.098
Previous Online Learning Experience | .133 | 1.142
Employment Status | -.033 | .967

p<.1. *** p<.01. **** p<.001

We include the demographics, assigned groups, motivation of taking the course, intention of completing the course, and other pertinent factors, such as previous online learning experience within this MOOC completion model. The demographic factors include education level, age, gender, employment status (full time/part time/not working), and English proficiency. For the motivation factors, the reasons for taking the course were described in the pre course survey as personal interest, interested in connecting with others, course offered by a certain institution or professor I like, earn a certificate, help with my current academic program, and friends’ taking.

Factors contained in the MOOC completion prediction model as shown in table 3, age and intention of completing the course are significant factors affecting the course completion. Learners at old age and intended to complete the course achieved higher performances in MOOC. Motivations of taking this course are not decisive factors compared with demographics and indicators of completing the course. However, learners who viewed certain institution that delivered the MOOC high completed more of this course. Surprisingly, there is a negative correlation between learners’ view of the importance of professors who teach this course with their course completion, which means learners who completed more of the course viewed the importance of professors low.
Table 4

Model accuracy estimation

<table>
<thead>
<tr>
<th>Predicted Completion</th>
<th>Real Completion</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Complete</td>
<td>Complete</td>
</tr>
<tr>
<td>Not Complete</td>
<td>68.6%</td>
<td>23.5%</td>
</tr>
<tr>
<td>Complete</td>
<td>3.2%</td>
<td>4.8%</td>
</tr>
<tr>
<td>Total</td>
<td>71.7%</td>
<td>28.3%</td>
</tr>
</tbody>
</table>

The numbers are rounded to one digit after the decimal point.

As seen in table 4, this model contains 23.5% of Type I error, which means the participants actually completed the MOOC while the model predicts non-completion of participants in this course. It also includes 3.2% of Type II error, by which the model predicts a learner’s completion of the course, while in facts he/she didn’t. However, this model produces 73.4% accuracy of predicting learners’ performance in this CIC MOOC.

Six hundred and thirty learners were included in this MOOC completion model with 25 invalid records with missing answers, such as they didn’t answer their ages or intention of completing the course.

Conclusions

We found out that working in a group didn’t affect completion, as the same finding from Zhang et al.’s (in press) paper, however, age (demographics factor), and intent to complete the course are highly related to students’ real course completion.

People at an older age with the intention of completing the course tend to advance more in this CIC MOOC. We also find out that learners who review the institution offering this MOOC high tend to finish the course. In contrast, learners who completed the course view the importance of MOOC instructors less.
Our findings are also closely aligned with Morris, Hotchkiss and Swinnerton (2015) and Guo and Reinecke’s (2014) conclusions that age is strongly correlated with MOOC completion. However, gender, education level, English proficiency, previous online learning experience, and employment status did not affect learners’ course completion in this CIC MOOC. Whereas, Engle, Mankoff and Carbrey (2015) concluded that educational level and English proficiency are influential factors for successful completion of a MOOC. Cisel (2014) identified employment status (full time versus part-time job) as a crucial factor of predicting course completion, and he found out that learners with part-time jobs achieved higher MOOC completion rate. Prior online learning experience is also established as a predictive factor of a learner’s MOOC performance in Morris, Hotchkiss and Swinnerton’s (2015) study.

Age and intention of completion are two significant factors to predict MOOC completion in our study. Data on these two factors were collected from the pre-course survey, and our findings imply that we may be able to use the pre-course/grouping survey data to predict learners’ course completion. To increase the course completion rate, the instructors and course designers may design and implement more vivid elements, such as multimedia products that include interactions between the course-learner and learner-learner to attract and retain younger participants.

**Discussions**

Grouping didn’t work in improving MOOC completion rate in this study, which maybe because of the less structured group formats and no scaffolds available. Besides, we didn’t monitor their group work, or keep track of group work progress. Breslow et al. (2013) indicated that students worked offline with people who had prior course
knowledge achieved higher in MOOC. The majority of our assigned groups worked online, and we found no difference of students’ achievement levels between grouped and non-grouped learners. For future study, we may include offline study groups as well.

Wen, Yang and Rosé (2015) found out that team leaders play an important role in keeping the cohesion of the team and improving team performance. For further studies, we would like to implement a more detailed teamwork instruction and assign team leaders or have the team elect the leaders to organize and lead group activities. Glance, Barrett & Hugh (2014) found out that the less activities implemented in MOOC design, the higher achievements learners could achieve. Instructors and course designers should also consider distributing the appropriate workloads for MOOC learners, and it may reduce the learners’ motivation of completing the course with too many activities added to the course.

This study also presents some limitations. First of all, those participants were called for working in groups, and the original research purpose was to evaluate the grouping effects on MOOC learners’ performances instead of assessing their course completion factors, and the predictive factors should be refined in our next study to improve the accuracy of the prediction model. Second, the majority of our participants in the grouping study are more motivated to learn compared with learners enrolled in the whole course, which is represented as a higher rate of the indication of completing the MOOC. Last but not least, this MOOC completion prediction model is created based upon the data from the CIC MOOC, due to the content and learner demographic variances in different MOOCs, this CIC MOOC prediction model may not be generalizable.
References


Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. In *Proceedings of the first ACM conference on Learning@scale conference* (pp. 21-30). ACM.


Appendix A: Email sent to Asynchronous groups

Dear CIC students,

You have been grouped based on your answers to our survey questions, including your preferred language to communicate in your group, your intention to complete course activities, and the way you would like to communicate with others (asynchronously - text chat offline).

Now, we encourage you to contact your group members (listed below). Here are some suggestions when you form groups. Take time to introduce yourself and get to know other team members. Since you won’t be together, online communication tools will play an important role. You may choose the tool/tools based on your needs.

Consider:

- using the hashtag #EAA1 in Coursera forums to find each others’ work and comment;
- create facebook group EAA1 and use it to post and discuss your work with the group members;

This will be a fun adventure! You will make friends from all over the world with various backgrounds and be exposed to fascinating ideas from different cultures. We hope you will learn and grow, build friendships and have fun!

Team member list:

Member A, contact information;
Member B, contact information;
Member C, contact information;
…
Let’s get this journey started! Wish you success in this course!

CIC MOOC Grouping Team
Appendix B: Email sent to Synchronous groups

Dear CIC students,

You have been assigned to the ESA7 group! This assignment was based on your CIC MOOC grouping survey preferences, including your preferred language to communicate in your group, your intention to complete course activities, and your preferred synchronous way (audio, text, video on line at the same time) to communicate with others. You provided the time slots when you are available to work in a group, and we found other people from different time zones that can work at the same time.

The best time to meet with your group members is Sundays from 7 am to 9 am your local time!

Now, we encourage you to contact your group members (listed below).

Here are some group building suggestions:

• Take time to introduce yourself and get to know other team members;
• Since you won’t be meeting face-to-face, online communication tools will play an important role in helping you stay connected with your group. Choose the tool/tools based on your needs;
• Use the hashtag #ESA7 in Coursera forums to find each others’ work and share feedback;
• Create Facebook group ESA7 and use it to post and discuss your work with your group members.
• Create a Zoom ESA7 space and meet on **Sunday from 7 am to 9 am your local time.**

This will be a great adventure! You will meet people from various backgrounds and from different corners of the world. You will be exposed to fascinating ideas from different cultures. We hope you will learn and grow, build friendships and have fun!

Wishing you success in this course!
Appendix C: Email sent to Chinese participants (Translated into Chinese)

Dear CIC students,

You have been grouped based on your answers to our survey questions, like your preferred language to communicate in your group, your intention to complete course activity, and the way you would like to communicate with others (synchronously - video and text online at the same time or asynchronously - text chat offline).

If you have chosen to communicate synchronously, we have tried our best to group you with students with same or similar schedules matched to your timezone. So now, we encourage you to contact your group members (all cc-ed here), select a leader and assign roles and tasks for team members. Each group member is expected to contribute to team projects to share diverse perspectives, pool knowledge and skills and hold each other accountable.

Here are some suggestions when you form groups, take time to introduce yourself and get to know other team members. Since the majority of groups in this CIC MOOC will work remotely, some online communication tools will place important roles in completing team projects. For Chinese users, we recommend QQ, WeChat, Skype or Blog, or any combination of those tools. You may choose the tool/tools accordingly based on your needs, which best facilitates team communication and knowledge building. Here are the QQ and WeChat links, and we recommend to start video-audio and text-based chats.

- Here is the tutorial to form QQ group:

- Tutorial to form WeChat group: (Using Mobile devices and Desktop computer)
  [http://www.360doc.com/content/14/0129/14/14747069_348734275.shtml](http://www.360doc.com/content/14/0129/14/14747069_348734275.shtml)

Skype instructions:
• If you have Skype names, exchange with group members. If you don't have Skype account, visit this link to download: http://www.skype.com/en/download-skype/skype-for-computer/

• Here is the link for downloading Google hangouts: https://www.google.com/+/learnmore/hangouts/

This will be a fun adventure! You will make friends from all over the world with various backgrounds and be exposed to fascinating ideas from different cultures. We hope you will learn and grow, build friendships and have fun!

Let’s get this journey started! Wish you success in this course!

CIC MOOC Grouping Team

亲爱的同学们：

良好的小组合作，能够促进学习、保持注意力和提升生活体验。在本次 CIC 课程中，我们将根据你在调查表中的回答意向（例如倾向的语言或母语、课程完成意图、倾向的小组规模等），将你划分到相应的学习小组之中。一旦当你收到关于组建学习小组的通知邮件时，每个小组都需要选出一个领导者，并为小组成员们赋予相应的角色和任务。每位小组成员都应为小组项目作出贡献，分享观点、积累知识技能，彼此负责。

当你们组建学习小组时，建议你们能够花时间介绍自我，与其它小组成员彼此认识。考虑到参加本次学习的大多数小组都将采用远程交流的方式，因此在线交流工具对于大家完成小组项目，具有重要的作用。对于中国学生来说，我们推荐你使用 QQ、Wechat、Skype 或博客，或者混和使用这些工具。你可以选择根据实际需求，来选择最有助于实现你们小组交流和知识构建的在线工具。通过小组项目，你能够掌握一系列有助于目前或未来获得成功的技能。
这将是一段有趣的学习之旅。你能够和来自世界各地不同背景的同学结识为友，并获得源于不同文化的奇妙创意。相信你将获取新知、建立友谊，体验快乐！

让我们开始学习之旅吧！期待与你合作，祝愿你学习成功！

CIC 研究团队
Appendix D: ZOOM instruction sent to Ad-Group

Download - less than a minute (30 sec) - if you are not taken to the room directly, there are two options:

- Click on the link from this email again;
- Click Join a meeting from the zoom website: https://zoom.us/ and enter the following IDs:
  - For Room 1 - 567 890 1234
  - For Room 2 - 678 901 2345
  - For Room 3 - 789 012 3456

- [https://www.youtube.com/playlist?list=PLKpRxBfeD1kEM_111d3N_Xl77fKDzSXe](https://www.youtube.com/playlist?list=PLKpRxBfeD1kEM_111d3N_Xl77fKDzSXe)

- 如果你所在的会议室中还没有其他人，你可以保持窗口打开状态并干点其他事，同时等待他人。当有其他人进入会议室时，会有响铃提示。如果你想学习关于 ZOOM 的更多功能，请点击:
  [https://www.youtube.com/playlist?list=PLKpRxBfeD1kEM_111d3N_Xl77fKDzSXe](https://www.youtube.com/playlist?list=PLKpRxBfeD1kEM_111d3N_Xl77fKDzSXe)。
  关于中文版 ZOOM 的功能介绍，你也可以参见：http://zoom.tom.com/.

- Thanks!!谢谢!

- [Click Here to Enter Meeting Room One](https://zoom.us/j/5678901234) (or open a browser and go to https://zoom.us/j/5678901234)

- 点击此处进入第一会议室 One (或直接在浏览器打开网址：https://zoom.us/j/5678901234)
- If that room is full, try
  - Enter Meeting Room Two (or open a browser and go
to https://zoom.us/j/6789012345)
  - 点击此处进入第二会议室 Two (或直接在浏览器打开网址:
    https://zoom.us/j/6789012345)

- If that room is full, try
  - Enter Meeting Room Three (or open a browser and go
to https://zoom.us/j/7890123456)
  - 点击此处进入第三会议室 Three (或直接在浏览器打开网址:
    https://zoom.us/j/7890123456)