

Who Is the English as a Second Language Speaker in this MOOC?

Gülüstan Doğan, Ayşe Saliha Sunar, İsmail Duru, and Su White

Abstract—Massive Open Online Courses have been widely used all over the world in recent years in entirely online learning context or as blended learning on campus. Most of these courses are offered in English. A high percentage of the users, however, are speaking English as a second language. Some of the authors of this paper who either used MOOCs for blended learning or a research subject are English as a second language speakers as well. They have observed whilst teaching students at university during blended teaching using MOOCs that the students struggle in courses offered in English. This has motivated us to explore this issue in MOOCs to contribute to the pedagogy of MOOCs. The main question that we consider is how can these platforms give a better experience to second language English speakers. There are many sub-problems of this big research question. In this paper we would like to briefly present our initial findings and give an overview of the research on this area.

Index Terms—Second language English speakers, MOOC, course engagement, FutureLearn, personalization.

I. INTRODUCTION

Online platforms which offer online courses with free registration to any learners who would like to participate, become one of the trend implementations in technology enhanced learning and are investigated from many different perspectives [1]–[5]. In order to improve teaching and learning in MOOCs, researchers have been investigating MOOCs from different perspectives such as stakeholders [6], design [1], MOOC pedagogy [2], learners' engagement [3], its effect on higher education [4], MOOC forums [7], [8] predicting performance [9]–[12] and possible improvements [13].

One of the main concerns in MOOCs is less personalized platform design and one static content that have been offered to such a diverse participants¹.

This personalization problem can be handled from many different angles such as age, ethnicity, sex, educational background and so on. For example, a learner who has experience in social science may struggle with a technical subject and may need basic introductions on the topic [14].

According to the statistics that class-central published in 2015, 75% of MOOCs were authored in English². This

research particularly consider language differences for international MOOC participants who come from diverse cultural backgrounds, and many of whom English will not be their first language.

The Ministry of Education in China predicted that over 10 million Chinese students would enrol in a MOOC by the end of 2016, which is almost 7 times higher than the rate in 2014³. Yet 90% of Chinese universities have not planned to launch a MOOC according to the same news. Although many MOOCs in China offered in Mandarin, we assume that many of these students will enroll in English language MOOCs. Many students choose to enroll in English MOOCs in order to improve their technical vocabulary and expertise.

Barak *et al.* state in their research investigating MOOC participants' motivation and engagement mediated by language [15] that learners coming from different cultural back-ground may differ in learning methods, communication styles, and patterns of behaviors.

Eriksson *et al.* [16] point out in their qualitative study investigating the reason of learners' dropouts that some learners had struggled with understanding the spoken language in the video and occasionally the instructors' accent. The MOOC participants also stated that subtitles in English were helpful. Contrarily, for example, Aboshady *et al.* [17] did not find English as a significant barrier in their investigation of Egyptian medical students' perspective towards a MOOC delivered in English (less than 20% of the students struggled). It may be due to medical students' advanced level of education.

One of our authors, Gulustan Dogan, is an assistant professor in Yildiz Technical University in Istanbul, Turkey. In 2016 Spring semester, she used blended learning approach in her teaching on campus, and used a MOOC as a teaching material for 8 of her first language Turkish speaking students. Every Wednesday evenings during a semester, she and her students organized classes via Google Hangout. Dr. Dogan and her students met on Google Hangouts instead of meeting physically. A different student was responsible for presenting the related week's material every week. As English is not their first language, the student who was responsible for presentation translated the videos to Turkish. That student also made a further search on the material that was being presented to present the class. The students stated that they had hard time understanding the professor's captured video. They complained that the professor was speaking English too fast and some phrases were difficult to understand. Since blended learning gives students and instructors flexibility to some extent, conscientious students often spent long hours

Manuscript received January 11, 2016; revised June 26, 2017.

Gülüstan Doğan and İsmail Duru are with Yildiz Technical University, İstanbul, Turkey (e-mail: gulustan@yildiz.edu.tr, iduru@yildiz.edu.tr).

Ayşe Saliha Sunar and Su White are with University of Southampton, UK (e-mail: ass1a12@soton.ac.uk, saw@ecs.soton.ac.uk).

¹ <https://computinged.wordpress.com/2013/01/04/moocs-are-a-fundamental-misperception-of-how-learning-works>

² <https://www.class-central.com/report/moocs-2015-stats>

³ <http://thepienews.com/news/chinese-mooc-learners-top-10-million-year-end>

translating and understanding the content and share their effort with the rest of the class an already translated MOOC content to overcome the language barrier. However, MOOCs in their first language, translated contents or peer study groups for further collaborations are not always available for all learners. Therefore, each international MOOC participant has their own personal experience while dealing with the language.

These examples suggest us that understanding behaviours of learners who speak English as a second language needs deeper investigation. In our research, we aim to investigate possible different engagement patterns of English as first language (FLL) and English as second language (ESL) speakers in MOOCs. We have access to datasets of FutureLearn MOOCs that are authored by the University of Southampton. This paper reports our method for data analysis regarding to detection of second language English speakers in a course. Additionally, we discuss challenges and opportunities of this research.

II. OBJECTIVE OF OUR RESEARCH

The main objectives of our research can be classified into three parts.

- 1) identifying second language English speaker (ESL) learners automatically from the FutureLearn dataset via a machine learning algorithm
- 2) identifying pattern of engagements of ESL learners
- 3) predicting the drop-out and/or certificate earn of ESL students.

III. METHODOLOGY

In order to conduct our research, a mix (quantitative and qualitative) research method has been taken. The following subsections demonstrate the data exploration process, preliminary findings, and tools that have been used for analysis and visualisation.

A. Datasets

We have used the forth run of the Understanding Language: Learning and Teaching MOOC that was run between 2016-04-04 and 2016-05-02. We have chosen this course as a start since it has attracted many international English language teachers around the world. Data files that we used in this study as follows;

- Enrolments: Includes demographics information, participation and enroll and unroll time and purchased statement certificate information.
- Step Activity: Includes step activity priorities related to course structure and activities' usage and completion time information.
- Comments: Includes comment priorities related to course structure, comment text, commenting user and comment relates social features.

B. Data Exploration: Clusters

In this study, we aimed to display differences between English as a first language participants (FLL) and English as a second language participants (ESL) in usage of

Understanding Language FutureLearn MOOC. Firstly, we used enrollment dataset to identify users' locations. In order to achieve this, we checked participants' country information which is provided in the Enrolments dataset. In the process of clustering, step users divided into three groups based on country information as using the Wikipedia link as reference⁴.

Fig. 1 and Fig. 2 show the overall country distribution in the course. The percentage of FLL and ESL students' is 13, students with no country information is 87%. It is also shown that ESL students are outnumbered than FLL students.

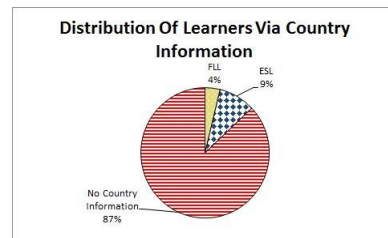


Fig. 1. Learners via country information.

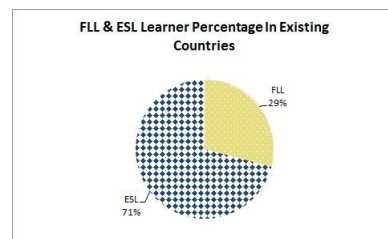


Fig. 2. FLL and ESL learner percentage in existing countries.

In the first group; there are users from countries listed in sub-titles: i) Countries in which English is a de facto official, but not primary language (e.g. Bangladesh), ii) Countries where English is a de jure official language (e.g. Canada), iii) Countries where English is a de facto official language (e.g. Australia). The only exception in this list is the country Mairutus. This country is not included in the Wikipedia list however it appeared as a common language along with other official languages. That is why we included it in the first language English speakers' group.

The second group consists of users who are not from the countries that are in the categories mentioned in the first group. In the third and last group, there are users who did not give the country information.

After the grouping process, the data files of the course, Comments, Enrollments and Step Activity, have been converted from csv format to MySQL format. Then, using the Mysql Workbench program, the data is divided into sub-categories based on their language.

According to the results of the analysis on three categories, it is seen that the method of separating the users according to the country information is insufficient in regarding to its accuracy. It was then decided to use the step asks learners to introduce themselves for better grouping and to examine the texts in the StepActivity data file which are located in step 1.5 in the first week. The reviews were first made manually and found that the records like listed below should be classified as different from country information.

⁴ https://en.wikipedia.org/wiki/List_of_territorial_entities_where_English_is_an_official_language

- A learner who lives in the UK and recorded as "GB" in the database stated that her first language is different from English. She also added that she studied several European languages and went on to do an English Language and Literature degree.
- A participant joining from South Australia expressed that his first language is not English. He said that he uses English everyday and it is easy to communicate with people speaking English as second language but he said that talking to native speakers can be rather complicated for him.
- Another participants who is located in the UK said that she was raised as bilingual (Ukrainian and Russian) and became a teacher of English.

Many examples like these show us that learners may have different first languages than the official language of the country where they live. In addition ESL students may have different English fluency level as it is mentioned. Therefore, we manually modified the categorises based on the participants' comments in the step 1.5.

IV. RESULTS

This section demonstrates graph to show the demographic and engagements of subgroups of learners by language. We used R programming language and the MySQL query language for visualisation.

A. Distribution of Learners Based on Purchased Statement

Fig. 3 and Fig. 4 display the percentage of FLL and ESL students based on purchased statement priority. The percentage of FLL and ESL are so low in all participants as it is very commonly seen in MOOCs.

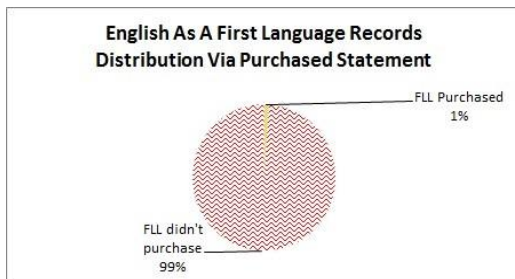


Fig. 3. FLL record distribution based on purchased statement.



Fig. 4. ESL record distribution based on purchased statement.

B. Distribution of Learners Based on Step Activity

In FutureLearn MOOC every activity of participants in a week is referred as a step activity.

In Fig. 5 and Fig. 6, A refers to learners up to 50% step activity completion, B refers to learners with step activity

completion between 50% to 80% and lastly C refers to learners completed higher than 80 percent of the steps.

It is seen in the figures that while step completion increases in FLL students, there is no big difference between two clusters.

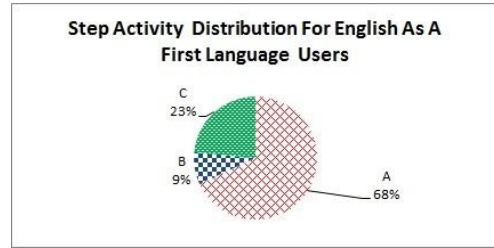


Fig. 5. FLL step activity distribution.

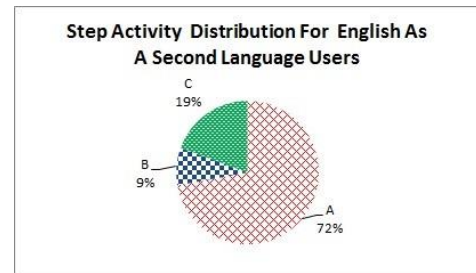


Fig. 6. ESL step activity distribution.

C. Distribution of Learners Based on Commenting Rate

Fig. 7 shows the general distribution based on comment rate. It shows that 16% of learners contributed to discussions at least with one post.

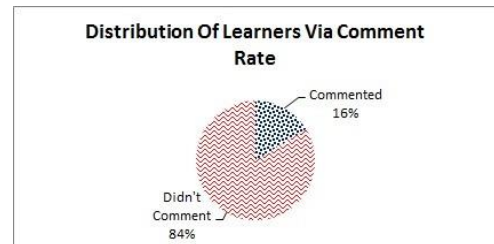


Fig. 7. Learners' distribution based on comment rate.

Fig. 8 and Fig. 9 show the distribution for FLL and ESL students based on comment rate. Both clusters have higher rate in contributing to discussion than the course average. However, there is no big difference between the two groups.

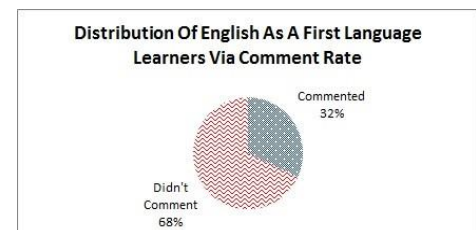


Fig. 8. FLL learners' distribution based on comment rate.

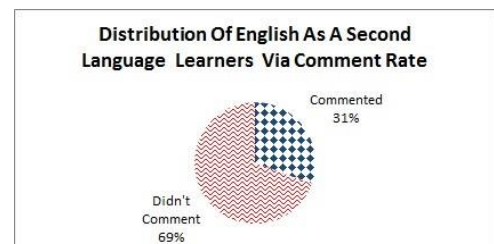


Fig. 9. ESL learners' distribution based on comment rate.

D. Distribution of Learners Based on Fully Participated at Ratio

In Fig. 10, the general distribution based on fully participation rate shown and it is just 8% of the all enrolled learners.

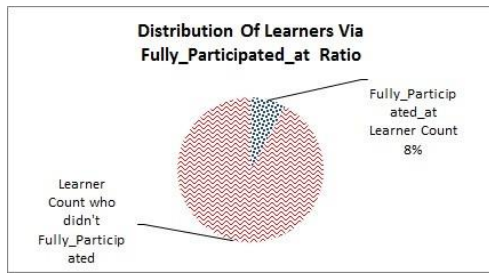


Fig. 10. Learners' distribution based on fully participated at ratio.

Fig. 11 and Fig. 12 show the distribution for FLL and ESL students based on fully participation rate. There is an important increase in two clusters and the percentage of FLL students is slightly higher than the ESL students.

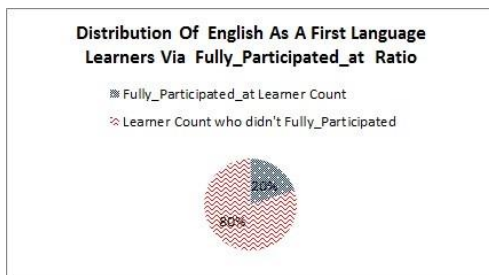


Fig. 11. FLL Learners' distribution based on fully participated at ratio.

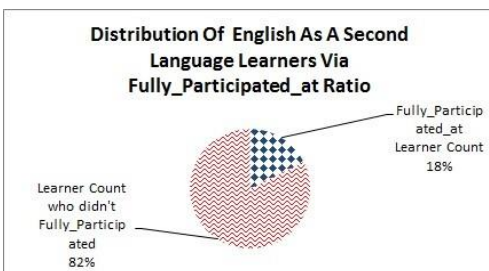


Fig. 12. ESL learners' distribution based on fully participated at ratio.

V. IDENTIFIED CHALLENGES AFTER THE PRELIMINARY STUDY

It is challenging to identify whether or not a participant is English as first language speaker. There are two ways to collect this information: i) Questionnaire, ii) Forum responses. There is a questionnaire at the beginning of the course. This questionnaire collects some demographic information from students. If we classify students according to where they live, this does not always tell us the right information. Such as a Turkish student might be living in England. In those cases, classifying students according to the country where they live can lead to erred classification. In order to detect these students we looked into the messages in the introduction section of the forum. In these messages students have written introductory sentences such as "I live in England but I am from Turkey". Some students who have not filled the questionnaire have written on the forum. Therefore the missing information can be detected. Another challenge is to detect the fluency of speakers. For example it might not

be right to put the students from Finland and Germany in the same cluster. Because statistics show that Finland and Germany has different English fluency level in English. Even for the students from the same country, English fluency can have different levels. Two Turkish students living in England can have different English proficiency. It might not be right to put these students in the same cluster. At the beginning of the study, we did a very high-level classification but the results suggest that a more detailed classification is required. We will do the more detailed classification as a continuation of this study.

VI. CONCLUSION

Potential contributions of this research is as follows.

- With the MOOCs education will be globalised. This globalisation will force us course providers to recognise international students and to present a personalised experience to ESL students.
- According to the learning patterns of ESL students, the platforms should be individualised.
- Based on the collected data learning strategies can be given to ESL students as well. For example the videos on the platform can come automatically with subtitles and at a slower pace once the platform detects that the user is an ESL student.

We believe that our results can be a new perspective for MOOC personalisation. As we do not hold the source code for popular MOOC platforms, we cannot make direct changes to MOOC platforms. However these identification methods can be used on their platforms for identification of ESL students. An ESL student who uses any MOOC platform such as FutureLearn would be more engaged if they are given a more personalised experience.

VII. FUTURE WORK

Intuitively we divided the set of students into two sets as first language English speakers and second language English speakers. We used a Wikipedia article⁵. However the characterisation of the two sets failed as these sets had a lot of common properties. Many ratios such as drop-out ratio, step activity ratio were very similar in these two groups. As we progress, we plan to divide the students into smaller subsets such as speakers from countries where English is their first language, speakers from countries where English is not first language but official language. According to this classification India and England will not be in the same set. But in our current classification they are in the same set. As we are still in data preprocessing phase, these kind of changes are normal. Once we learn the dataset by heart, we will be able to classify more successfully.

One other method to better identify ESL learners is to make use of the comments the users have entered. For instance in the forum students have introduced themselves and written about their fluency in English.

⁵https://en.wikipedia.org/wiki/List_of_territorial_entities_where_English_is_an_official_language

ACKNOWLEDGMENTS

The dataset used in this paper is provided by the University of Southampton for the ethically approved collaborative study, ID number 23593. This work has been done under the project numbered 2016-04-01-DOP05 in Yildiz Technical University (YTU). The authors would like to thank Prof. Dr. Banu Diri from YTU for her help in our studies.

REFERENCES

[1] F. Brouns and O. Firssova, "The role of learning design and learning analytics in MOOCs," presented at the 9th EDEN Research Workshop, Oldenburg, Germany, October 2016.

[2] L. Guàrdia, M. Maina, and A. Sangrà "MOOC design principles: A pedagogical approach from the learner's perspective," *ELearning Papers*, no. 33, 2013.

[3] A. Sunar, S. White, N. Abdullah, and H. Davis, "How learners' interactions sustain engagement: A MOOC case study," *IEEE Transactions on Learning Technologies*, December 2016.

[4] S. White and S. White, "Learning designers in the 'third space': The socio-technical construction of moocs and their relationship to educator and learning designer roles in he," *Journal of Interactive Media in Education*, 2016, no. 1, November 2016.

[5] I. Duru, G. Dogan, and B. Diri, "An overview of studies about students performance analysis and learning analytics in moocs," in *Proc. IEEE International Conference on Big Data*, pp. 1719–1723, IEEE, 2017.

[6] S. White, H. Davis, K. Dickens, M. León, and M. M. Sánchez-Vera, "MOOCs: What motivates the producers and participants?" in *Proc. International Conference on Computer Supported Education*, pp. 99–114, Springer, October 2014.

[7] C. Tucker, B. Pursel, and A. Divinsky, "Mining student-generated textual data in moocs and quantifying their effects on student performance and learning outcomes," presented at 2014 ASEE Annual Conference, Indianapolis, Indiana, Indianapolis, Indiana, 2014.

[8] Y. Feng, D. Chen, Z. Zhao, H. Chen, and P. Xi, "The impact of students and tas' participation on students' academic performance in mooc," in *Proc. the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2015, pp. 1149–1154, ACM, 2015.

[9] G. Kennedy, C. Coffrin, P. Barba, and L. Corrin, "Predicting success: How learners' prior knowledge, skills and activities predict mooc performance," in *Proc. the Fifth International Conference on Learning Analytics and Knowledge*, pp. 136–140, ACM, 2015.

[10] M. Klíšener and A. Fortenbacher, "Predicting students' success based on forum activities in moocs," in *Proc. 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 2, pp. 925–928, IEEE, 2015.

[11] C. Ye and G. Biswas, "Early prediction of student dropout and performance in moocs using higher granularity temporal information," *Journal of Learning Analytics*, vol. 1, no. 3, pp. 169–172, 2014.

[12] B. Xu and D. Yang, "Motivation classification and grade prediction for moocs learners," *Computational Intelligence and Neuroscience*, vol. 4, 2016.

[13] C. Coffrin, L. Corrin, P. Barba, and G. Kennedy, "Visualizing patterns of student engagement and performance in moocs," in *Proc. the Fourth International Conference on Learning Analytics and Knowledge*, pp. 83–92, ACM, 2014.

[14] M. K. Khrib, M. Jemn, and O. Nasraoui, "Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval," in *Proc. Eighth IEEE International Conference on Advanced Learning Technologies*, pp. 241–245, IEEE, 2008.

[15] M. Barak, A. Watted, and H. Haick, "Motivation to learn in massive open online courses: Examining aspects of language and social engagement," *Computers & Education*, vol. 94, pp. 49–60, March 2016.

[16] T. Eriksson, T. Adawi, and C. Ståhr, "Time is the bottleneck": A qualitative study exploring why learners drop out of MOOCs," *Journal of Computing in Higher Education*, pp. 1–14, November 2016.

[17] O. A. Aboshady, A. E. Radwan *et al.*, "Perception and use of massive open online courses among medical students in a developing country: multicentre cross-sectional study," *BMJ Open*, vol. 5, no. 1, p. e006804, December 2014.



Ayşe Saliha Sunar received her BSc in mathematics at Gazi University in Ankara, Turkey. She then took a one year professional teaching programme to teach mathematics to secondary school students. In 2011, she pursued an MSc at the Graduate School in Social and Information Science Department at Nagoya University, Japan. Her MSc thesis focussed on learning languages with the aid of intelligent tutoring systems. Currently, she is a PhD candidate in the Web and Internet Science Research Group (WAIS) at the University of Southampton, UK where she contributes to the research of the MOOC Observatory. Her research interests include big data, learning analytics, MOOCs, learning at scale, technology enhanced learning, predictive models and personalisation of learning.



Ismail Duru Ismail Duru is currently a research assistant and PhD student at Yildiz Technical University, Istanbul, Turkey. He received her BSc in Computer Engineering at Istanbul University in Istanbul, Turkey. In 2014, he completed his MSc at Computer Engineering Department at Yildiz Technical University, Turkey. His MSc thesis name was "Application of Agile Software Development Method in Mobile Application Development". He has been managing two research and development (RD) projects named as "Learning Analytics Technologies on Online Education Platforms", "Language Teaching Technologies on Mobile Applications and Social Networks" which are funded by Turkish Government. In his PhD thesis he study on predicting English as a second users' behavior and performance analysis. His research interests include Big Data, Educational Data mining, Performance Prediction and Behavior Analysis in MOOCs.



Gulustan Dogan Gulustan Dogan is currently working at Yildiz Technical University, Istanbul, Turkey as an Assistant Professor. She worked at NetApp and Intel as a software engineer in Silicon Valley. She received her PhD degree in Computer Science from City University of New York. She received her B.Sc degree in Computer Engineering from Middle East Technical University, Turkey. She is one of the founding members of Turkish Women in Computing (TWIC), a Systems community affiliated with Anita Borg Institute. Her research interests are big data, predictive analytics, distributed computing and networks.



Su White Su White is an associate professor who researches in the Web and Internet Science Group within Electronics and Computer Science (ECS) at the University of Southampton. She supervises a broad group of PhD students interested in MOOCs and Learning, and is the co-director of the MOOC Observatory co-located in WAIS and the University's Institute for Learning Innovation and Development (ILiAD). The MOOC Observatory maintains a large, and growing open MOOC citation list in Mendeley and curates Grey Literature and Web sources for this fast moving area via a ScoopIt collection. The group has a growing publication list based on their own research, and there is a large body of work in progress incorporating stakeholder analyses, blended MOOCs, learning analytics, educational modelling, digital literacies and personal learning. The group has hosted a number of visiting academics and welcome collaboration of this kind.