Getting to know English Language Learners in MOOCs: Their Motivations, Behaviors and Outcomes

Selen Türkay
Harvard University
Cambridge, MA, USA
selen_turkay@harvard.edu

Hadas Eidelman
Harvard University
Cambridge, MA, USA
hadas_eidelman@harvard.edu

Yigal Rosen
Harvard University
Cambridge, MA, USA
yigal_rosen@harvard.edu

Daniel Seaton
Harvard University
Cambridge, MA, USA
daniel_seaton@harvard.edu

Glenn Lopez
Harvard University
Cambridge, MA, USA
glenn_lopez@harvard.edu

Jacob Whitehill
Worcester Polytechnic Institute
Worcester, MA, USA
jrwhitehill@wpi.edu

ABSTRACT

Massive Open Online Courses (MOOCs) promise to engage a global audience and emphasize the democratic achievement of free, university-level education. While such open access enables participation, it is unclear how learners who are not fluent in English (ELLs) engage with MOOC content. After all, the language of MOOCs is English. In order to improve accessibility for ELLs in digital learning environments, we must first have a clear understanding of the educational landscape: who are the non-native English speakers enrolled in MOOCs? Where are they located geographically? What are their current online learning behaviors, motivations and outcomes? In this paper we start answering some of these questions by analyzing data from 100 HarvardX courses, using self-report and log data. Preliminary analysis show evidence that ELLs are motivated by more utilitarian goals compared to non-ELLs.

Author Keywords

MOOCs; English language learners; motivations; behaviors; outcomes

ACM Classification Keywords

K.3.1. Computers and Education: Computer Use in Education

INTRODUCTION

MOOCs have the potential to promote equity in educational opportunity for traditionally underserved learners worldwide. Often, the most underserved learners are those in less economically developed nations, where the native

language is not English. However, because English is the primary language for MOOCs, for learners with beginning and intermediate English language ability, authentic materials are often beyond their language proficiency and may become incomprehensible without help [12]. A recent survey by Guokr [4] with Chinese learners (n=13,526) presented dire language barriers in online learning. 82% of the respondents said that they wouldn't be able to follow the course unless Chinese close caption is provided. Consequently, more than half of the participants have not taken an online course (n=7809) and 31% of them pinpointed the language barrier as the main reason. For those who have taken online courses, 11% reported that they failed to complete the course because they could not follow the content in English.

In this paper, we aim to learn more about ELLs, their motivations, behaviors, and outcomes in HarvardX MOOCs by analyzing data from 100 courses which is a subset of 150 courses that includes only those learners who filled out pre-course surveys self-reporting their English proficiency and motivations for enrolling in the MOOC (11.35% of all use cases, representing 521221 unique users). Finally, we use recent data from one course, CS50, to produce analysis of transcript usage. CS50 has twelve languages available to learners to use as transcripts.

Opportunities and Challenges for English Language Learners in Online Learning and MOOCs

The education of non-native English speakers in MOOCs is a critical issue both within the United States and abroad. On a domestic level, the U.S. Census Bureau [10] indicates 59.5 million people aged 5 and over spoke a language other than English at home. Only 58% of this population reported speaking English "very well" whereas 7% did not speak English at all. ELLs' educational outcomes, including high school graduation rates (57% versus 79%), tend to be lower compared to native English speakers [5, 11]. Outside of the U.S. context, one of the promising features of MOOCs is the ability to deliver high-quality educational content to learners in disadvantaged portions of the globe. As it was

noted earlier, ironically, the geographic areas that could most benefit from MOOC content are often those where English is not the main language of communication. Therefore, both within and outside of the American education system, the issue of access to MOOC curricula for students who are non-native English speakers is critical from the perspective of equalizing opportunity for underserved learners.

There are several reasons why MOOCs represent learning environments offering a unique opportunity to educate learners from vulnerable populations. Compared to traditional classroom-based learning environments, online learning and MOOCs in particular offer the potential to (a) cross geographic and educational boundaries to deliver high-quality content; (b) give students agency to learn when and where they want; and (c) provide educational content in multiple modalities (e.g., audio, video, text) to suit students' differing needs or preferences (e.g., closed captioning, adjustable playback speed) [5]. The asynchronous nature of MOOCs provides ELLs more privacy and preparation time by lowering barriers [12]. For ELLs, these advantages can facilitate subject matter comprehension while simultaneously promoting learning of the language of instruction itself [7,8].

However, without deliberate design for accessibility, online education risks broadening the gap between advantaged and disadvantaged learners [5]. The digital divide in online education has two fronts: content access driven by economic status and students' capability to benefit from that content [2]. The latter prevents many ELLs from accessing courses developed in English-speaking countries [3]. Considering the large number of students with low English proficiency in the US and around the world, only 20% of the world population can fully benefit from English-only educational content [1], highlighting the need for implementing various accommodations to increase access for these students to digital learning environments. Some most common accommodations are video captions and subtitles. Lecture transcripts are beneficial by-products of these accommodations. ELLs can download the transcripts to aid their language comprehension.

A previous interview study showed evidence that ELLs may have different motivations than non-ELLs [9]. In this paper, we will investigate ELLs' motivation to take MOOCs, using self-reported survey questions, and how learners' motivations may differ based on their level of engagement in the course. We will also compare ELLs' motivations to non-ELLs' motivations to find potential differences between these learner groups.

DATA

HarvardX enrollment data and events data generated by participants across all HarvardX courses were analyzed using the HarvardX/MITx edx2bigquery canonical dataset Person-Course-Survey. This comprehensive dataset includes all the common standard Qualtrics pre-course survey questions (including English proficiency), and merges together the edX user data and learning metrics including course enrollment and registration information, demographics, resource access, course progression and time-on-task for all HarvardX courses.

By the time we wrote this paper, HarvardX pre-course survey included four questions on learners' English fluency in reading, writing, speaking, and listening (e.g., How fluent are you in English, the language of this course - reading?), and users rated their English fluency on a 5-point Likert scale (0=Weak; 1=Basic; 2=Intermediate; 3=Proficient; 4=Fluent). Here, we define ELLs as learners who answer in the fluency questions *not fluent*. If a learner answers one of these language skills as not fluent, we categorize this learner as an ELL. If s/he answers all four items as "fluent", we categorize these learners as non-ELLs.

RESULTS

Geographic Locations

The top five countries with most ELLs in HarvardX MOOCs are India, USA, Brazil, China, and Spain. The fact that USA is among these top five countries with ELL MOOC users strengthens our emphasis on national dimension of ELL user population. From USA, California, New York and Massachusetts have the most number of ELL registrants in HarvardX courses.

Certification Rates

For this portion of our analysis, in order to avoid wildly imprecise estimates of certification rates that might occur in particularly small courses, we restricted our sample of MOOCs to only those that had 100 or more ELLs enrolled and 100 or more students who received certification (*n*=64). In these 64 courses, using the operationalization for ELLs that we presented above, we found that the proportion of ELLs by course ranged from 6.0-29.6%. Similarly, restricting the data to only those UN geographic regions that had 100 or more ELLs and 100 or more learners who received certification (*n*=17), we found that the proportion of ELLs ranged from 3.9% in Northern America to 56.5% in Northern Africa. At the course level, certification rates for ELLs ranged from 0.33-35.6% for ELLs and 0.32-56.3% for non-ELLs. We fit logistic regression models estimating students' probability of certification by ELL status, controlling for age, gender, geography, developing nation status, and online behaviors. Preliminary modeling in the full sample suggests the odds of ELLs certifying are roughly 0.6 times the odds for non-ELLs ($p<0.001$), accounting for the nesting of students in courses and in different configurations of covariates. We conclude with sensitivity analyses using our two alternate versions of the language-proficiency measure.

Transcript Use

Our dataset included a transcript variable: *ntranscript*, which describes the number of video transcript events from tracking logs. These transcript events include download transcript, and toggle on and off closed captions. We found

that ELL users, on average, create significantly more transcript events than non-ELL users ($t(279720) = -21.82$, *cohen's d* = 0.07, M_{ELL}=14.36, $M_{non\text{-}ELL}$=11.15).

Comparison of ELL and non-ELL users' behavior of transcript download show that ELLs downloaded transcripts more than non-ELLs did ($t(239560)$=-15.9, *p*<.0001, *cohen's d*=0.05). (Reader should keep in mind that we are able to track transcript downloads only since May 2016.) Another transcript related video player feature is the running transcripts on the right side of the video. Learners can choose to view these or close it by clicking on the "CC" button. This event is registered as *show_transcript* in our data. Considerable numbers of ELLs used this function (*n*=59518). However we are unable to conclude whether their goals were to show or close running transcripts.

The HarvardX course with the highest number of ELL registrants is CS50: Introduction to Computer Science course, which is also the largest course offered at HarvardX. CS50 also offers 12 different transcript languages to assist learners. When investigating further, approximately 38% (*n*=8790) of CS50 learners who filled out the pre-survey are ELLs. Examining log event data related to transcripts, we find that 32% (*n*=2851) of these learners use transcripts other than English as their modal transcript language, and 2771 learners have downloaded the video transcripts to aid their comprehension.

Video Events and Forum Participation
We analyzed video and forum interaction events for ELL and non-ELL users based on the cumulative data on each type of interaction (i.e., video play, pause, forum threads, forum comments). On average, we found that ELLs take significantly more video play ($t(247930) = -25.11$, *Cohen's d* = 0.03), video pause ($t(304220) = -10.31$, *Cohen's d* = 0.08), and seek actions ($t(354200) = -6.75$, *Cohen's d* = 0.02) compared to non-ELL users (see Figure 1). Forum participation metrics revealed that compared to ELLs, non-ELL users start more forum threads ($t(52269) = 3.06$, *Cohen's d* = 0.02, p=.002), make more comments($t(61398) = 11.05$, *Cohen's d* =0.08, p < 2.2e-16), and votes more on other learners' posts (t(67943)=13.52, *Cohen's d* =0.10, p < 2.2e-16) (see Figure 2).

Comparison of ELL and non-ELL Users' Motivations to Take MOOCs
In pre-course surveys, learners answer questions on their motivations to take the course that they signed up for. They are asked to rate their motivation to take a MOOC on ten dimensions (How important were the following reasons in choosing to register for this course?) These dimensions are: Engaging in lifelong learning; learn from the best professors and universities; advancing learners' formal education; participate in an online community; curiosity about online learning; career advancement; access learning opportunities not otherwise available to them; learn about course content; earn a certificate; better serve their

community. See Table 1 for ratings by learners where Cohen's d is larger than 0.2 (a small effect size).
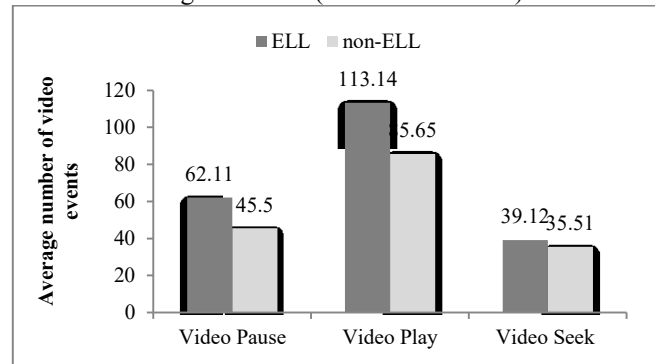

Figure 1. Average number of video events per person by language proficiency.
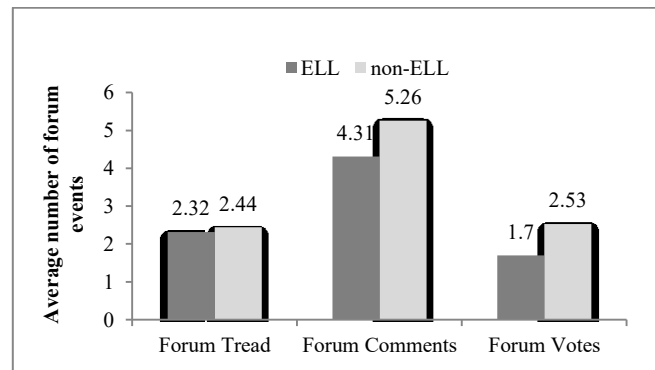

Figure 2. Average number of forum events per person by language proficiency.

	t	$M_{(non\text{-}ELL)}$	$M_{(ELL)}$	Cohen's d
Lifelong learning	69.3	3.37	3.17	0.23
Best professors	-64.98	3.19	3.39	0.21
Education advan.	-92.55	2.55	2.93	0.30
Participate	-89.65	1.57	1.94	0.29
Career advance.	-84.76	2.48	2.85	0.27
Certification	-82.15	1.57	1.94	0.27
Community	-69.83	2.14	2.43	0.23

Table 1. Two sample t-test statistics and effect sizes, comparing ELL and non-ELL users' motivations.

Results show evidence that ELL and non-ELL users may be motivated by different goals to take MOOCs. When we further compare ELL and non-ELL registrants' motivations by course progress (i.e., certified, explored, viewed), the differences across motivations become even more salient. In Figure 3, we can see that ELL and non-ELL users tend to view courses if their motivation is to engage in life long learning. For all other levels of engagement, the

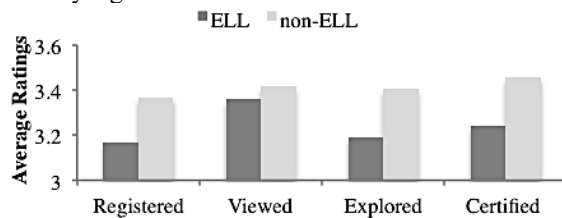differences between ELLs and non-ELLs remain statistically significant.

Figure 3. Average ratings on *lifelong learning* motivation item by course progression grouped by language fluency.

DISCUSSION

This study explored differences and similarities in motivation, behaviors and outcomes between ELL and non-ELLs in HarvardX courses. We hope the results will inspire L@S community to look closer at sub-learner groups in MOOCs and their potentially differing needs, as one-size-fits-all approach is not the ideal model of education.

Results showed that ELL and non-ELL registrants have somewhat different motivations on various dimensions. Although it is expected to find significant differences among groups with large sample sizes, this finding is still telling: users' motivations may differ based on their language proficiency, which usually relates to other demographics and socioeconomic status. It is evident that learners who are ELLs have more utilitarian goals when signing up for these courses. For instance, previous studies found that many learners do not seek credit toward any credential [6]. Our findings not only confirm this for non-ELLs, but also show that ELL users might be more motivated toward earning certificate than non-ELLs. This implies potential focus on helping ELLs attain certification and devising tracks similar to ID verification to give certificates more utility for ELL users. We also found that although ELL users were more motivated to participate in community than non-ELLs, their forum participation is lower than non-ELLs. This may indicate that ELLs need more support to encourage their participation in forums and other online learning communities.

While we recognize the importance of encouraging universities and colleges in non-English speaking counties to develop their own culturally targeted content in their own languages, we believe that goal driven interventions to make high quality instructional materials, which are usually developed in English, available to ELLs will lower barriers to democratize education through MOOCs.

We acknowledge generalizability issues for the sample we used for this paper, such as underestimated ELL prevalence (learners with especially-low English proficiency are less likely to fill out the survey) and inflated rates of course certification (learners who complete the survey are likely more committed to persistence), by comparing them to the full sample.

CONCLUSION

This work contributes to practical and scholarly knowledge of online instruction and learning environments. Results illuminate successes and challenges for ELLs in MOOCs, descriptively and through modeling statistically the relationships between learners' likelihood of certification and student-level behaviors and region-level characteristics. We hope that findings will inform future research and experimentation in digital learning environments to benefit both ELLs and non-ELLs globally and within the U.S.

REFERENCES

1. Tita Beaven, Anna Comas-Quinn, Mirjam Hauck, Beatriz De los Arcos, and Timothy Lewis. 2013. The Open Translation MOOC: creating online communities to transcend linguistic barriers. *Journal of Interactive Media in Education* 2013, 3, 18.
2. Sheryl Burgstahler. 2002. Distance learning: Universal design, universal access. *Educational Technology Review* 10, 1, 32–61.
3. Rainbow Tsai-Hung Chen, Sue Bennett, and Karl Maton. 2008. The adaptation of Chinese international students to online flexible learning: two case studies. *Distance Education* 29, 3, 307–323.
4. GuoKr. 2015. Available online: http://mooc.guokr.com
5. Grace Kena et al. 2015. The condition of education 2015. NCES 2015-144. *National Center for Education Statistics* (2015).
6. Steve Kolowich. 2013. The professors who make the MOOCs. *The Chronicle of Higher Education* 18.
7. Deborah L. Linebarger. 2001. Learning to read from television: The effects of using captions and narration. *Journal of Educational Psychology* 93, 2, 288–298.
8. Peter Shea. 2000. Leveling the playing field: A study of captioned interactive video for second language learning. *Journal of Educational Computing Research* 22, 3, 243–263.
9. Judith Uchidiuno, Amy Ogan, Kenneth R. Koedinger, Evelyn Yarzebinski, and Jessica Hammer. 2016. Browser Language Preferences as a Metric for Identifying ESL Speakers in MOOCs. In ACM Press, 277–280.
10. U.S. Census Bureau, 2011. Changes in Population Controls, American Community Survey Research Note.
11. Evelien Van Laere, Orhan Agirdag, and Johan van Braak. 2016. Supporting science learning in linguistically diverse classrooms: Factors related to the use of bilingual content in a computer-based learning environment. *Computers in Human Behavior,* 57, 428–441.
12. Yi (Leaf) Zhang. 2013. Power distance in online learning: Experience of Chinese learners in U.S. higher education. *The International Review of Research in Open and Distributed Learning* 14, 4.