Multiplatform MOOC Analytics:
Comparing Global and Regional Patterns in edX and Edraak

José A. Ruipérez-Valiente[1], Sherif Halawa[2], Justin Reich[1]
[1] Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge (MA), 02139, USA
[2] Edraak, 300 King Abdullah St, Amman, Jordan
jruipere@mit.edu, shalawa@qrf.org, jreich@mit.edu

ABSTRACT
While global massive open online course (MOOC) providers such as edX, Coursera, and FutureLearn have garnered the bulk of attention from researchers and the popular press, MOOCs are also provisioned by a series of regional providers, who are often using the Open edX platform. We leverage the data infrastructure shared by the main edX instance and one regional Open edX provider, Edraak in Jordan, to compare the experience of learners from Arab countries on both platforms. Comparing learners from Arab countries on edX to those on Edraak, the Edraak population has a more even gender balance, more learners with lower education levels, greater participation from more developing countries, higher levels of persistence and completion, and a larger total population of learners. This "apples to apples" comparison of MOOC learners is facilitated by an approach to multiplatform MOOC analytics, which employs parallel research processes to create joint aggregate datasets without sharing identifiable data across institutions. Our findings suggest that greater research attention should be paid towards regional MOOC providers, and regional providers may have an important role to play in expanding access to higher education.

ACM Classification Keywords
K.3.1 Computers and education: Computer Uses in Education: Distance learning; H.1.2 Information Systems: User/Machine Systems: Human factors; J.1 Computer Applications: Administrative Data Processing: Education

Author Keywords
MOOCs; Learning Analytics; Cross-institutional Collaboration; Large-scale Analytics; Cultural Factors.

INTRODUCTION
When the first university-based MOOC providers emerged in 2012, MOOC advocates argued that large-scale online learning could dramatically extend the reach of higher education institutions around the world, especially to places with limited access to elite universities [9]. However, some studies have suggested that instead of decreasing the 'educational gap', MOOCs have been primarily attracting learners with high educational attainment, thus widening educational disparities [12]. Among global providers, disparities in MOOC participation and persistence remain for females, learners with low levels of educational attainment, and among countries with low development levels, as measured by the United Nations Human Development Index [6]. Many (though not all) of the courses on the largest global MOOC providers, such as Coursera, edX or FutureLearn, are produced by Anglo-American universities. For learners not fluent in English or not familiar with Anglo-American higher education, language and cultural barriers may compound the challenges of learning independently online. Numerous studies have pointed to the importance of considering cultural factors when designing inclusive online learning experiences [16, 14, 11].

English language MOOC providers that have aimed to attract global audiences are complemented by a set of regional MOOC providers that have been emerging over the past years with a focus on serving learners in their regions. Some examples include the Ibero-American initiative MiriadaX [1], the Chinese initiative XuentangX [2], the Italian Federica.EU [3], the German OpenHPI [4], the French platform France Université Numérique (FUN [5]) or the Arab initiative Edraak [6]. Several of these initiatives (Edraak, XuentangX and FUN) use Open edX [7]–edX's collaborative open source initiative–as the underlying software to run MOOCs. By focusing on more specific communities with a more broadly shared language and culture, these initiatives may be able to play a complementary role to MOOCs aiming at a global, primarily English-language audience.

While several observers have noted that the global providers have recently put a greater focus on professional, post-Baccalaureate degrees and programs [18, 17, 3], these shifts

---

[1] https://miriadax.net/en/
[2] https://www.xuetangx.com/
[3] https://federica.eu/
[4] https://open.hpi.de/
[5] https://www.fun-mooc.fr/
[6] https://www.edraak.org/
[7] https://open.edx.org/

do not appear as pronounced in the regional providers, suggesting that regional MOOC providers may play a different role in higher education ecosystems than the larger, global, English-language entities. Understanding similarities and differences in learner behavior and experiences between global and regional platforms may help provide a more comprehensive picture of global learning with MOOCs.

At present, MOOC studies have focused on elite global providers over the regional providers. Prior studies focus on a detailed analysis of one or few MOOCs (e.g. [4]), a longitudinal study with many courses from one single MOOC provider (e.g. [6, 8], or a literature review of MOOC analytic studies [23]. Small MOOC studies have limited generalizability, longitudinal studies from one single institution do not capture variation across MOOC providers and literature reviews cannot make cross-study comparisons due to the different methodologies employed. In response to these challenges, we propose multiplatform MOOC analytics, which replicates the same analyses using a common script in different MOOC environments. This work aligns with calls for methodologies that promote open education science [22] and replication studies in the social sciences [2].

For this case study, we focus on learners in the Arab World, where equitable access to educational opportunity and overall educational achievement remain substantial challenges with potential for improvement [21]. Also, the Arab World has significant information technology and e-learning gaps compared to the developed world [1] and as a region, has lower participation and completion metrics in MOOCs compared to other regions [19]. Additionally, the Arab World countries rank in the "very low" category on the EF English Proficiency Index, a global measures of English proficiency [24]. A 2018 report from the International Labour Organization has brought up the low levels of female participation in labour markets, especially in regions such as the Arab States and Northern America, where unemployment rates for Arab women can be twice as high as men's [13]. Researchers have found correlations between the levels of unemployment of a region and MOOC demand [20].

To facilitate a comparison between MOOC learners from Arab states on edX (with data from MITx and HarvardX) and Edraak, we organize the data from these providers into a common framework analyzed with a shared body of analytic code. Our overarching study question explores the extent to which MOOC trends are universal versus context-dependent. To this end, we examine the following research questions:

1. Do regional MOOC providers help narrow educational and gender gaps for the local population?

2. Do regional MOOC providers have higher levels of engagement, participation and completion for the local population, compared to global providers?

The remainder of the paper is organized with a description of the methodology in next Section 2, a presentation of the results in Section 3 and we close the study with a discussion of future work in Section 4.

METHOD

This section describes the methodology in the context of MOOC providers Edraak and edX in Subsection 2.1, a description of the dataset, sample and measures in Subsections 2.2, 2.3 and 2.4 respectively. We end with an overview the multiplatform MOOC analysis methodology in Subsection 2.5.

Context

This work leverages commonalities between MOOC providers Edraak and edX: (1) both institutions aim to spread free education (although Edraak with a more regional focus), and (2) both providers use the Open edX platform as the underlying software for their Learning and Content Management Systems. The fact that both institutions use the same software environment to host their MOOCs facilitates comparisons in two important ways: (1) it is easier to establish a common data processing pipeline to put the data into the same format, and (2) both providers use a similar, Open edX-based learning management system (though Edraak is modified for right-to-left languages), limiting some of the variation in learner experience.

EdX was founded in 2012 by the Massachusetts Institute of Technology and Harvard University. More than 100 schools, companies and other institutions have joined this initiative as partners to teach free courses to a global audience. Open edX is being used by numerous institutions around the world to deliver their open or private courses, which has exponentially increased the impact of the initiative. From this platform, we focus on MITx and HarvardX, the institutional units that produce and teach the edX courses from each university.

Edraak was founded in 2013 by the Queen Rania Foundation for Education and Development. Edraak's inception was a response to the language barrier that prevents a substantial population in the Arab world from learning online in English. Edraak's mission is to fulfill these learners' needs by providing high-quality educational content in Arabic, although some courses are also available in English.

Edraak participated in the localization of Open edX for the Arab world by providing Arabic translation for the platform and enhancing the support for right-to-left (RTL) languages. Edraak produces all of its courses in Arabic, except for courses teaching foreign languages, and hosts them on its locally-adapted Open edX platform. Edraak's courses span multiple categories, including STEM, business and workforce development skills, health, arts, and language. Course content is designed in collaboration with regional experts from academia and industry. Edraak performs all remaining course development and management tasks internally, such as media production, content upload, course marketing, and course operations.

Dataset

MITx and HarvardX data contain 565 MOOC iterations, 12.67 million course registrations from over 5.63 million unique student accounts, and we are able to detect country of origin of 4.48 million accounts. Most of the courses taught in MITx and HarvardX are in English. Edraak's data contain 231 MOOC iterations with a total of 3.77 million registrations to these

courses from 1.5 million unique student accounts. We are able to identify the country of origin of 1.48 million of these student accounts. Most of the courses in Edraak are in Arabic (except for few courses licensed from edX).

We collected all MOOC data from MITx and HarvardX from 2012 to May 2018, and all data from Edraak from June 2014 to September 2018. The data from each provider include self-reported variables by students from site registration and pre-course surveys and the Open edX tracking logs, that contain every action performed in each course by each account. Part of the multiplatform MOOC research methodology that we establish for this case study involves having a common data format, as we explained in Subsection 2.5. Additionally, since both institutions are using the same software, the data processing to create the common format does not entangle any problematic transformations between the two environments. The data follows a person-course structure, where every row represents the data from the registration of a student to a course.

Analytic Sample
To enable comparison between cohorts across the study we define the following three populations of participants, where participants are defined as learners that registered for and then viewed a course at least once:

- `Edraak`: Participants who connected to Edraak from Arab countries (N = 589,817). The small number of learners who connected to Edraak from outside Arab countries in Edraak are not included, as our research questions are focused on the effect of regional providers on local populations (N = 12,203).

- `MITxHx Arab`: Participants who connected to MITx or HarvardX courses from Arab countries (N = 120,868).

- `MITxHx Rest`: Participants who connected to MITx and HarvardX courses from outside Arab countries (N = 3,267,199).

Noteworthy trends about the `Edraak` and `MITxHx Arab` cohorts include the relative size of each cohort: Edraak reaches almost five times more learners from the Arab World compared to MITx and HarvardX courses. In the 4 section, we synthesize study findings to form a hypothesis about the role of Edraak as an effective regional competitor and weigh in on the extent to which there are unobservable factors contributing to this trend.

Measures
For this study, the inputs to our multiplatform analytic infrastructure is a person-course file, where each row represents one course registration for one learner. The fields in this input file are as follows:

- User id: A unique identifier for each user account.

- Course id: A unique identifier for each course iteration in the course (note one course can be run multiple times).

- Nationality: Country of origin of each account computed based on the modal IP address from the tracking logs of the student.

- Level of education: Qualitative variable with the self-reported level of education of the account.

- Gender: Qualitative variable with the self-reported gender identification of the account.

- Viewed: Boolean variable that indicates if the account accessed the course at least once.

- Explored: Boolean variable that indicates if the account viewed at least half of the chapters of the course.

- Completed: Boolean variable that indicates if the account achieved a passing grade in the course.

We also also use the Human Development Index (HDI) provided by the United Nations (UN) in 2018 as a proxy for the country level of socioeconomic status [8].

Finally, we use a high level categorization of courses established by previous work [6] with the following four broad categories: Government, Health, and Social Science (GHSS), Humanities, History, Design, Religion, and Education (HHRDE), Computer Science (CS) and Science, Technology, Engineering, and Mathematics (STEM). The distribution of course category in each platform is as follows:

- Edraak (N=231): 31.17% GHSS, 47.19% HHRDE, 5.19% CS and 16.45% STEM.

- HarvardX and MITx (N=565): 26.55% GHSS, 27.07% HHRDE, 7.8% CS and 38.58% STEM.

Multiplatform MOOC Analytics
Much of MOOC research over the past five years has been conducted in studies of single higher education institutions, and even the largest studies have aggregated data from within a single MOOC provider. One of the most promising initiatives in this area is MORF framework [10], a platform that enables institutions to securely deposit their MOOC data and allows researchers to execute Docker containers for data analysis while maintaining full privacy of the data. In this study, we propose a methodology that we denote as multiplatform MOOC analytics, which leverages commonalities across MOOC learning and content management systems to allow research teams to create common data formats, agree upon analytic methods, and then generate aggregate data, produced through identical processes, that can allow for "apples-to-apples" comparisons between different MOOC platforms. The goal is to, as closely as possible, conduct the same analysis within the two MOOC environments. This is increasingly important in the field of the social sciences, since human factors can add a great variability in the outcomes of each study [2]. We used the following procedure:

1. Both institutions shape their data into the same format (see Subsection 2.4).

---

[8] http://hdr.undp.org/en/content/human-development-index-hdi

| | Year of Birth | |
Cohort	Median	Mode
Edraak	1992	1995
MITxHx Arab	1991	1994
MITxHx Rest	1988	1992

Table 1. Median and mode of the year of birth by cohort.

2. We collaboratively generate a common Jupyter notebook that is expecting the common data format. This script outputs aggregate data from both institutions that is merged together for the joint analysis. This greatly alleviates the logistical and privacy concerns of sharing student-level information.

3. We conduct the joint data analysis of both MOOC platforms together.

For MITx and HarvardX data we use *edx2bigquery* [15] framework to transform the raw edX logs into this person-course format. Edraak uses a simplified and adapted version of this process to transform the Open edX log files from their platform into this same person-course file. For analysis, we then input these person-course files into identical Jupyter notebooks, and then compare the aggregated output.

Our overarching study question examined the extent to which MOOC trends are universal versus context-dependent. We examined a host of metrics organized around two principal research questions. To answer research question 1–looking at the degree to which regional MOOC providers narrow educational and gender gaps–we compare the distribution of the level of education, year of birth and gender of each one of the cohorts. To answer research question 2–comparing engagement, participation and completion metrics among regional and global providers–we present participation metrics by country, perform an analysis of the "funnel of participation" of each cohort, compare engagement metrics of Edraak-produced and edX-licensed courses that has been offered in Edraak, and finally we report differences in engagement depending on the course category.

RESULTS

We organize the results by research question, based on summary statistics of the demographics of the participants in each cohort (research question 1) and the participation and completion, and retention trends (research question 2). In discussion section, we draw conclusions about the extent to which trends in the MOOCs examined are universal or context-dependent.

Regional Provider Edraak Reaches Regional Population with Lower Levels of Formal Education

On average, HarvardX and MITx participants from outside the Arab world have higher levels of educational attainment than fellow participants from within the Arab world. Participants from Edraak have lower average levels of education than both. Below, we present an overview of the level of education of participants in each cohort. Figure 1 shows on the left a boxplot visualization with the distribution of level of education by each cohort. The levels of education are 'Doctorate', 'Master', 'Bachelor', 'Associate', 'High School, Junior High School or

Elementary School' or 'Other'. Around 13% of the learners of Edraak have a Master or Doctorate, compared to the 19.6% of MITxHx Arab and the 25.5% of MITxHx Rest.

Regional and Global Providers Attract More Younger Female Participants from Arab Countries Compared to Younger Females from Other Countries

Our demographic analysis by gender revealed a higher participation of younger female Arab participants than older in both Edraak and MITxHx Arab, or an older population of both genders in MITxHx Rest.

In Figure 1 on the right, we display a boxplot visualization with the distribution by gender in each cohort. Edraak cohort reaches a higher population of female than male (50.67% female and 48.93% men), compared to MITxHx Arab (27.85% female and 71.89% male) and MITxHx Rest (34.09% female and 65.26% male).

In addition, Table 1 shows median and mode of the year of birth per cohort. With a median as year of birth of 1992 for Edraak, 1991 for MITxHx Arab and 1988 for MITxHx Rest, Edraak enrolls younger learners than both cohorts from edX.

Finally, in Table 2 we show the percentage of participants by cohort by cross-tabulating gender and age group. In Edraak there was only one course that was targeted to female participants, a course about Fashion Design. Removing this course only changes the gender gap by 1%, suggesting that the findings are not sensitive to course type.

Edraak Effectively Targets Local Learners and Participation is More Distributed in All Arab States

Examining Arab learners' participation across both platforms, we found a noteworthy difference in the representation of learners in Jordan in Edraak (15.27%) compared to MITxHx Arab (3.52%). In Table 3, we display the participation of the Arab World by country and cohort, showing the percentage of representation by country and number of learners per million inhabitants (which enables a better between-country comparison). The relatively higher representation of Jordan among Arab learners is attributed to three factors: 1) The stronger networked growth in Jordan (e.g. word of mouth, local press, as people consider it a national pride project as well), 2) Universities in Jordan have adopted Edraak as a blended learning platform quickly, unlike the other Arab countries, 3) There are some activation projects happening in Jordan to serve Edraak courses in Refugee camps with councils such as the Norwegian Refugee Council (NRC) and the Danish Refugee Council (DRC). There has been similar initiatives in the context of edX for example the MIT ReACT Hub, that aims to educate refugees [9].

In general, Edraak caters to a higher number of learners across all countries except for Somalia, which might be related to specific online learning and MOOC initiatives from MITx and HarvardX in the country. There are other slight differences in the level of participation by country, which might be related to the HDI level of each country. In fact, we find in
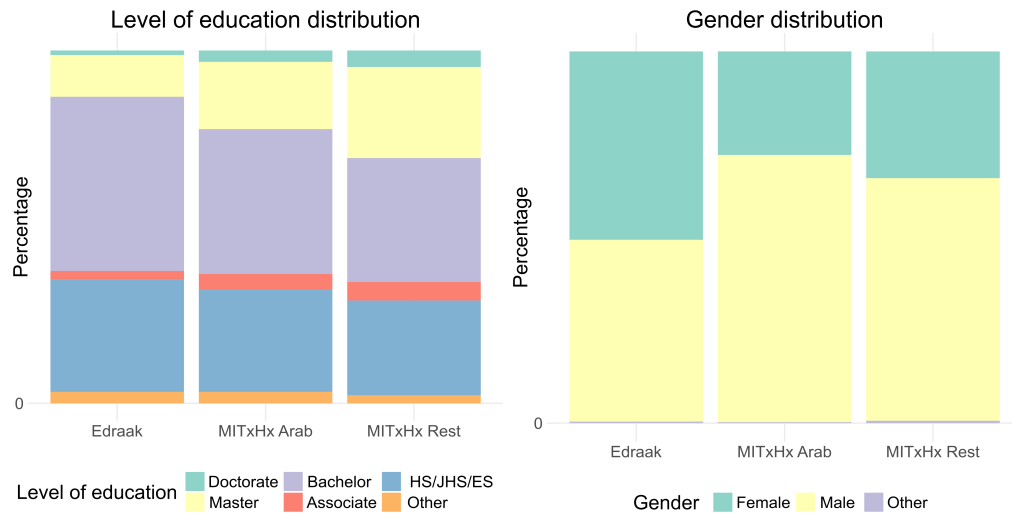
[9] https://react.mit.edu/

Figure 1. Distribution of gender and level of education by cohort.

Cohort	<18		>= 18 & <30		>= 30 & <45		>= 45 & <65		>= 65		Age
	Female	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*	**Gender**
Edraak	3.57%	2.24%	34.95%	30.35%	10.97%	14.06%	1.30%	2.40%	0.08%	0.07%	
MITxHx Arab	0.83%	1.90%	17.90%	45.74%	7.42%	20.65%	1.37%	3.94%	0.06%	0.19%	
MITxHx Rest	0.86%	1.69%	16.02%	33.55%	11.33%	21.98%	4.93%	7.43%	0.83%	1.38%	

Table 2. Demographic distribution of participants based on age intervals and gender by cohort.

`MITxHx Arab` cohort a moderate-high Pearson Correlation in of 0.69 ($p < 0.001$) between the HDI index of the country and the number of learners per million inhabitants, that declines in `Edraak` cohort with a value of 0.26 ($p = 0.23$). Another conclusion that we can get from this distribution is that the participation by country is more uniformly distributed in Edraak.

Edraak Participants Demonstrate Higher Completion Rates Compared to Their Arab Peers in EdX

Comparing completion rates between edX and Edraak reveals a more desirable "funnel of participation" [7] in `Edraak` cohort, where the percentage of courses completed triples the value of `MITxHx Arab` and doubles that of `MITxHx Rest`. As always when comparing completion rates, it is worth noting that learners take MOOCs for numerous reasons [5], and that the "funnel of participation" is not necessarily a proxy for MOOC quality or success. Nonetheless, comparisons of completion can help identify whether populations of learners have different goals or differential levels of success in meeting their goals. Table 4 presents completion trends in terms of percentage of courses explored and completed, by cohort. We see a similar, although attenuated trend, in the percentage of courses explored by cohort. Similar to what we saw in terms of correlations in Subsection 3.3, we find a moderate correlation of 0.55 ($p = 0.007$) by country within `MITxHx Arab` between the percentage of completed courses and the HDI, that becomes lower within `Edraak` with a value of 0.45 ($p = 0.03$).

Higher Engagement in Edraak-Produced Courses Compared to Local edX Licensed Courses

Additionally, there were eight courses taught in Edraak that were licensed directly from edX. Table 5 shows a comparison between those courses and the standard courses produced by Edraak in Arabic. The data show 11 times more completion in courses produced by Edraak compared to those licensed from edX, around three times more exploration and higher number of enrolments per course.

This contrast in engagement emerged despite Edraak's attempt to locally-adapt the edX-licensed courses as much as possible. Edraak translated all textual content in the courses to Arabic, including course information, announcements, and assessment questions. Images and visualizations were edited to change any English text to Arabic. The most challenging element to Arabize was the visual and audio content of videos. As a work-around, Edraak provided Arabic transcripts that learners can read alongside the videos. The challenge learners most commonly reported on the course forums is their split attention between watching the video and reading the transcript. This was not an issue in Edraak's home-produced content since the instructors speak Arabic and all video content is in Arabic.

Differences in Engagement Persist after Controlling by Course Category

The nature of the courses that are being offered by each platform can definitely have an impact in the patterns of engagement that we find (note course categories and distribution were presented in Subsection 2.4). We break down the results that we found in Table 4 by course categories draw from the

Edraak			MITxHx Arab		
Country	*Percentage of Arabs from total*	*Learners per million inhabitants*	*Country*	*Percentage of Arabs from total*	*Learners per million inhabitants*
Egypt	29.27%	1,737	Egypt	31.52%	383
Jordan	15.27%	9,093	Saudi Arabia	12.31%	443
Saudi Arabia	13.9%	2,444	United Arab Emirates	12.25%	1,552
Morocco	8.65%	1,410	Morocco	10.59%	354
Algeria	8.51%	1,195	Algeria	5.17%	149
United Arab Emirates	5.25%	3,244	Tunisia	3.77%	391
State of Palestine	5.08%	5,933	Jordan	3.52%	430
Syrian Arab Republic	4.26%	1,376	Lebanon	2.91%	578
Yemen	2.1%	429	Qatar	2.86%	1,284
Iraq	1.88%	282	Sudan	2.62%	76
Sudan	1.51%	214	Iraq	2%	61
Tunisia	1.44%	729	The Syrian Arab Republic	1.93%	127
Libya	0.91%	832	Kuwait	1.84%	531
Oman	0.5%	605	Oman	1.54%	385
Kuwait	0.39%	554	State of Palestine	1.22%	292
Lebanon	0.39%	379	Bahrain	1.15%	890
Qatar	0.26%	580	Somalia	1.09%	87
Bahrain	0.18%	694	Yemen	0.72%	30
Somalia	0.11%	42	Libya	0.6%	111
Mauritania	0.08%	109	Djibouti	0.25%	315
Comoros	0.02%	119	Mauritania	0.12%	32
Djibouti	0.02%	104	Comoros	0.01%	11

Table 3. Distribution of learners by country from the Arab World.

Cohort	Percentage explored	Percentage completed
Edraak	34.72%	10.66%
MITxHx Arab	15.86%	3.15%
MITxHx Rest	19.43%	4.85%

Table 4. Percentage of courses explored and completed from participants by cohort.

Licensed from edX?	Enrolments per course	Percentage explored	Percentage completed
No	16,363	35.28%	10.94%
Yes	15,009	9.25%	1.03%

Table 5. Participation and completion comparison of courses produced by Edraak (N = 223) from those courses licensed from edX (N = 8).

HarvardX/MITx annual reports [6] in Table 6. In general for both platforms, we see that Computer Science (CS) and Science, Technology, Engineering, and Math (STEM) courses have lower percentage of completion than Global Health and Social Science (GHSS) and Humanities, History, Religion, Design, and Education (HHRDE) course, suggesting the difficulty of STEM classes. Except for STEM category, we can see that for CS, GHSS and HHRDE the funnel of participation shows a better trend in `Edraak` than for both `MITxHx Arab` and `MITxHx Rest`. The only exception is STEM, where the trend of `Edraak` is better than `MITxHx Arab`, but not than the rest of the learners in `MITxHx Rest`.

Substantive subhead

Year-over-year learner retention, analogous to year-by-year client retention used in business anlaytics, can be used to evaluate the degree to which learners return to a platform for multiple courses. We define cohorts by the year of their first activity in the platform (i.e. registration year) and see how many of those learners remain in the platform during subsequent years. In Figure 2, we display the attrition rate, with the left plot for `Edraak` and `MITxHx Arab`. As an example, the '2014-2015' cohort of `Edraak` had 34,024 learners that participated in at least a course during the academic year '2014-2015', 32% of those learners participated in at least one course during '2015-2016', 22% in '2016-2017' and 16% during '2017-2018'. `Edraak` and `MITxHx Arab` have similar rates

Course category	Cohort	Percentage explored	Percentage completed
CS	Edraak	25.6%	6.91%
	MITxHx Arab	9.87%	1.20%
	MITxHx Rest	11.76%	2.74%
GHSS	Edraak	40.09%	13.40%
	MITxHx Arab	19.44%	5.59%
	MITxHx Rest	22.57%	6.41%
HHRDE	Edraak	35.08%	10.54%
	MITxHx Arab	17.42%	5.86%
	MITxHx Rest	21.59%	7.38%
STEM	Edraak	18.75%	3.61%
	MITxHx Arab	18.46%	2.33%
	MITxHx Rest	23.73%	4.22%

Table 6. Percentage of courses explored and completed by cohort and course category

of attrition. This trend, consistent with other recent MOOC studies [18], exemplifies a challenge that MOOC providers have in retaining learners year after year.

DISCUSSION

In this study we present "Multiplatform MOOC Analytics"– combining analysis based on a methodology to reproduce a research procedure with data from Edraak and two edX providers, MITx and HarvardX. The goal of the study was to analyze the how learners from the Arab states participate differently in these two platforms.

Our initial findings indicate that the educational gap that we find commonly in MOOCs, which is particularly salient among Arab learners in edX, is smaller in the regional setting of Edraak. This evidence suggests that Edraak is effectively at reaching less educated learners, which aligns with one of the original missions of MOOCs. We also found in Edraak a reduced gender gap, since the percentage of female learners (50.6%) is even higher than the number of male learners. Moreover, with high unemployment among women in the Arab World [13] and given that a previous study found a positive relationship between MOOC demand and unemployment [20], we hypothesize that this could be one of the reasons favoring a more equitable gender balance. The inclusion of more women in the workforce in the Arab world can have numerous positive outcomes in the development of the region [21].

It is noteworthy that the difference in gender distribution between Edraak and edX is not the result of gender-targeted marketing. Edraak performed targeted marketing to females in one course, but that course's contribution to the gender gap is than 1%. Previous work has reported lower enrolments from female population in courses related to CS and STEM, the lower proportion of courses in these topics when compared to HarvardX and MITx can also be one of the reasons of the better gender balance found in Edraak [6]. Understanding the cause behind this difference is an interesting direction for future study. The difference in education level distribution is another interesting future study direction. However, it may be partially attributed to Edraak's age-targeted marketing. Edraak focuses on the age bracket from 18 to 40 years old in marketing most of its courses. Another possible factor is the correlation between pursuing graduate studies and English proficiency. A perception survey study conducted by Edraak and the Queen Rania Foundation in 2017 has shown that learners with higher educational attainment levels are more familiar with English MOOC platforms including edX.

In terms of participation, we observed a much higher number of Arab learners in Edraak, we also found a strong, significant correlation between the number of learners per million inhabitants and the HDI of the country in `MITxHx Arab`, but this relationship did not hold in the `Edraak` cohort. These results might indicate the existence of barriers impeding Arab population accessing courses from global partners, we hypothesize that some of these barriers can be related to language, but also related to culture and pedagogy.
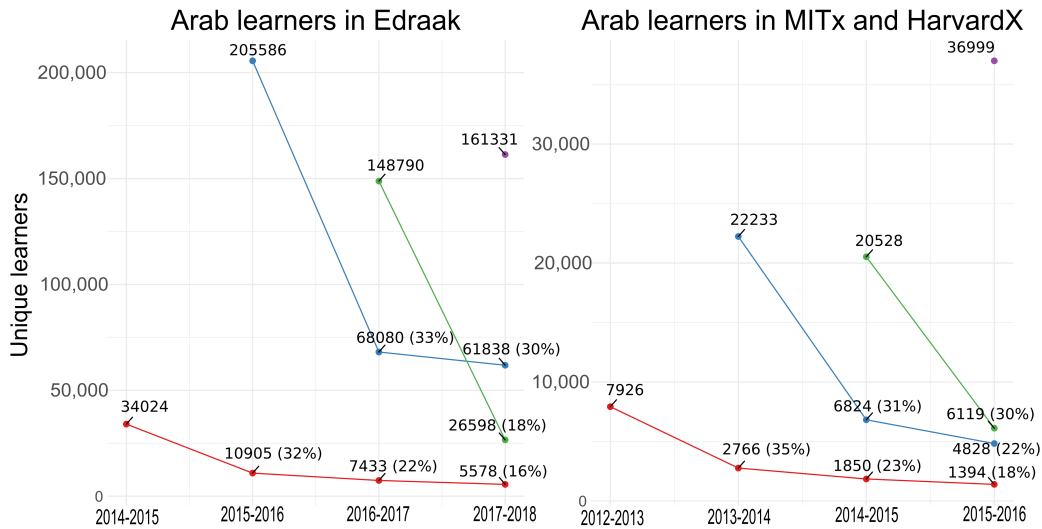
Figure 2. Retention of learners year by year with cohorts determined by registration year.

The completion trends differences were also noteworthy with a completion percentage 3 times higher in `Edraak` than for `MITxHx Arab` learners; and greater differences between the licensed courses from edX in Edraak (1% vs. almost 11% completion) and a higher number of enrolments in Edraak-produced courses. Therefore, the participation and completion levels are much lower for the Arab World for edX than for Edraak in the global context than in the regional one. These findings can be affected by some differences in the course category distribution of each platform, where MITx and HarvardX have a slightly higher percentage of courses on STEM and CS topics whilst Edraak has a slightly higher percentage of courses on GHSS and HHRDE. Therefore, as learners struggle more to complete technical courses, course categories might be having an effect on the overall trends of each platform. However, further and deeper qualitative analysis on the courses of each platform will be required to establish a better estimation of the importance of this difference.

Despite higher completion rates in Edraak courses, we found similar retention rates of learners across the years in both MOOC platforms. We might have anticipated that the platform with the higher completion rate would have higher year-over-year retention, as learners might have felt encouraged to continue completing courses in the platform. But, it may be that for many learners, even though who complete courses, individual courses are of more interest than developing an ongoing habit of learning within a single MOOC platform.

The contrast in engagement between Edraak's edX-licensed courses and its home-produced courses sheds light on an importance of localizing educational content. The difficulty that Edraak's learners in edX-licensed courses reportedly faced signals increased extraneous cognitive load due to split attention between the video and transcript. Educational content that does not require students to synchronize multiple separate sources of information naturally avoids this effect.

Our results are suggestive of the promise of local MOOC providers in achieving the goal of MOOCs to provide broadly accessible learning experiences for diverse learners, particularly for those learners with limited access to traditional higher education. The participants in Edraak are more gender-balanced, are more likely to have lower levels of education, are more likely to come from less developed countries, and are more likely to persist through and complete courses, and there are more total learners from the Arab world than on MITx and HarvardX. Multiple factors from marketing to course design to course rigor to the disciplinary makeup of course offerings to instructional quality could be responsible for some of these differences. Future work should explore the effect of Edraak's support for Arabic language (and for right-to-left languages more broadly), instructional materials created by Arab subject matter experts, and course design and marketing focused on Arab countries on MOOC learners' outcomes. While research and popular discussion has generally focused on the English-language, global providers, it may be that more attention should be paid to the regional, non-English providers and their efforts to expand online learning among cultural and language groups. Regional providers may be uniquely positioned to expand access to diverse learners.

One way to advance this line of work is to continue to expand research collaborations across multiple MOOC providers, extending the multiplatform MOOC framework demonstrated here from two platforms to many. As next steps we are developing a consortium of MOOC providers from across the globe to engage in a larger multiplatform project with an expanded list of several global and regional partners. The goal of this larger project is to identify global, longitudinal trends in MOOC usage and develop a deeper understanding of more universal and more context-dependent trends and to effectively compare MOOC learning in global and regional contexts.

REFERENCES

1. Kamal Abouchedid and George M Eid. 2004. E-learning challenges in the Arab world: Revelations from a case study profile. *Quality Assurance in education* 12, 1 (2004), 15–27.

2. Yehuda Amir, Irit Sharon, and others. 1990. Replication research: A" must" for the scientific advancement of psychology. *Journal of Social Behavior and Personality* 5, 4 (1990), 51.

3. Rose Baker, David L Passmore, and Brian Martin Mulligan. 2018. Inclusivity Instead of Exclusivity: The Role of MOOCs for College Credit. In *Enhancing Education Through Open Degree Programs and Prior Learning Assessment*. IGI Global, 109–127.

4. Lori Breslow, David E Pritchard, Jennifer DeBoer, Glenda S Stump, Andrew D Ho, and Daniel T Seaton. 2013. Studying learning in the worldwide classroom research into edX's first MOOC. *Research & Practice in Assessment* 8 (2013), 13–25.

5. Gayle Christensen, Andrew Steinmetz, Brandon Alcorn, Amy Bennett, Deirdre Woods, and Ezekiel Emanuel. 2013. The MOOC phenomenon: Who takes massive open online courses and why? (2013).

6. Isaac Chuang and Andrew Ho. 2016. HarvardX and MITx: Four years of open online courses–fall 2012-summer 2016. (2016).

7. Doug Clow. 2013. MOOCs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. ACM, 185–189.

8. Alexandra I Cristea, Ahmed Alamri, Mizue Kayama, Craig Stewart, Mohammad Alshehri, and Lei Shi. 2018. Earliest predictor of dropout in MOOCs: a longitudinal study of FutureLearn courses. (2018).

9. Tawanna R Dillahunt, Brian Zengguang Wang, and Stephanie Teasley. 2014. Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. *The International Review of Research in Open and Distributed Learning* 15, 5 (2014).

10. Josh Gardner, Christopher Brooks, Juan Miguel L Andres, and Ryan Baker. 2018. MORF: A Framework for MOOC Predictive Modeling and Replication At Scale. *arXiv preprint arXiv:1801.05236* (2018).

11. Elizabeth E Grandon, Khaled Alshare, and Obyung Kwun. 2005. Factors influencing student intention to adopt online classes: A cross-cultural study. *Journal of Computing Sciences in Colleges* 20, 4 (2005), 46–56.

12. John D Hansen and Justin Reich. 2015. Democratizing education? Examining access and usage patterns in massive open online courses. *Science* 350, 6265 (2015), 1245–1248.

13. International Labour Organization. 2018. *World Employment Social Outlook: Trends for Women*. Technical Report. Geneva. 7–8 pages. `https://www.ilo.org/global/research/global-reports/ weso/trends-for-women2018/WCMS`

14. Xiaojing Liu, Shijuan Liu, Seunghee Lee, and Richard J Magjuka. 2010. Cultural differences in online learning: International student perceptions. *Educational Technology & Society* 13, 3 (2010), 177–188.

15. Glenn Lopez, Daniel T Seaton, Andrew Ang, Dustin Tingley, and Isaac Chuang. 2017. Google BigQuery for Education: Framework for Parsing and Analyzing edX MOOC Data. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*. ACM, 181–184.

16. Catherine McLoughlin and Ron Oliver. 2000. Designing learning environments for cultural inclusivity: A case study of indigenous online learning at tertiary level. *Australasian Journal of Educational Technology* 16, 1 (2000).

17. Laurie Pickard, Dhawal Shah, and JJ De Simone. 2018. Mapping Microcredentials Across MOOC Platforms. In *2018 Learning With MOOCS (LWMOOCS)*. IEEE, 17–21.

18. Justin Reich and José A Ruipérez-Valiente. 2019. The MOOC Pivot. *Science* 363, 6423 (2019), 130–131.

19. José A Ruipérez-Valiente and Justin Reich. 2018. Participation of the Arab World in MOOCs. In *2018 Learning With MOOCS (LWMOOCS)*. IEEE, 47–50.

20. Tingting Tong and Haizheng Li. 2018. Demand for MOOC-An Application of Big Data. *China Economic Review* 51 (2018), 194–207.

21. United Nations Development Programme. 2016. *Arab Human Development Report*. Technical Report. New York. 74 – 78 pages. `http://arab-hdr.org/`

22. Tim van der Zee and Justin Reich. 2018. Open education science. *AERA Open* 4, 3 (2018), 2332858418787466.

23. George Veletsianos and Peter Shepherdson. 2016. A systematic analysis and synthesis of the empirical MOOC literature published in 2013–2015. *The International Review of Research in Open and Distributed Learning* 17, 2 (2016).

24. Wikipedia. 2019. EF English Proficiency Index. (2019). `https://en.wikipedia.org/wiki/EF`