Educational Data and Learning Analytics in KazNU MOOCs Platform

Alimzhanov Ye.S., University of International Business,
Almaty, Kazakhstan, E-mail: aermek81@gmail.com
Mansurova M.Ye., Al-Farabi Kazakh National University
Almaty, Kazakhstan, E-mail: mansurova01@mail.ru

The initial hype around massive open online courses (MOOCs) already subsided, but the number of new learners in MOOCs platforms is still growing. Due to low completion rates in the MOOCs compared to enrolled students it is important to establish and validate quality standards for these courses. Employing of educational data and learning analytics to improve lesson plans and course delivery become an innovative approach for teachers, curriculum developers and policy makers in education. Learning analytics of online courses can be also used for enhancement of classroom teaching by blending online and face-to-face learning models.

This work presents some observations about the behavior of students, obtained by analyzing the data generated during delivery of 13 MOOCs. Besides classification of learners by analysis their activity data, other interesting characteristics about platform learners like demographic, gender and level of education are described. The results indicate that the quality of interpersonal interaction within a course relates positively and significantly to student scores.

Key words: MOOCs, learning analytics, educational data, online learning, blended learning.

Әл-Фараби ат. ҚазҰУ-нің МООК платформасында білім беру мәліметтер мен ақпараттың сараптамасы

Әлімжанов Е.С., Халыкаралық бизнес университеті, Алматы, E-mail: aermek81@gmail.com
Мансұрова М.Е., эл-Фараби ат. Қазақ ұлттық университеті, Алматы,
E-mail: mansurova01@mail.ru

Жаппай ашық онлайн курстар (ЖАОК) алғашқы дүрлікпе төмендеді, бірақ ЖАОК платформасындағы жаңа студенттер саны әлі де өсіп келеді. Тіркелген студенттер санымен салыстырғанда, ЖАОК-ты аяқтаушылар біршама аз болғандықтан, бұл курстардыю сапа стандарттарын құрастыру және бекіту маөызды қадам болып табылады. Оқу жоспары мен оқу курстарын жетілдіру үшін білім беру деректерін және ақпараттың талдауды қолдану оқу бағдарламаларын әзірлеушілерге және білім беру саясатын құрастырушыларға инновациялық тәсіл болып табылады. Онлайн-курстың деректерін талдауды қашықтықтан және дәстүрлі оқыту модельдерін араластыру арқылы оқу процесін жетілдіру үшін пайдаланылуы мүмкін.

Бұл мақалада 13 ЖАОК-та жинақталған деректерді талдау арқылы алынған курс қатысушыларының іс-әрекеті туралы кейбір зерттеулер келтірілген. Нәтижелер курстың ішіндегі өзара қарым-қатынастың сапасы студенттер үшін оң және маңызды екенін көрсетеді.

Түйін сөздер: ЖАОК, ақпараттың сараптамасы, білім беру мәліметтер, онлайн оқыту, арарлас оқыту.

Образовательные данные и аналитика обучения на платформе МООК КазНУ им. аль-Фараби

Алимжанов Е.С., Университет международного бизнеса, Алматы, E-mail: aermek81@gmail.com
Мансурова М.Е., Казахский национальный университет им. аль-Фараби, Алматы,
E-mail: mansurova01@mail.ru

Начальный ажиотаж вокруг массовых открытых онлайн курсов (МООК) пошел на спад, но число новых учащихся на платформах МООК все еще растет. Из-за низких показателей завершения МООК по сравнению с зарегистрированными студентами, важным этапом является установление и утверждение стандартов качества для этих курсов. Использование образовательных данных и аналитики обучения для улучшения планов уроков и предоставления курсов станет инновационным подходом для учителей, разработчиков учебных программ и политики в области образования. Аналитика обучения онлайн курсов может быть использована для улучшения образовательного процесса смешиванием дистанционных и традиционных моделей обучения.

В данной работе представлены некоторые наблюдения о поведении слушателей курсов, полученные путем анализа данных, накопленных при проведении 13 МООК. Результаты показывают, что качество межличностного взаимодействия в рамках курса имеет положительный и существенный характер для учащихся.

Ключевые слова: МООК, анализ обучения, образовательные данные, онлайн обучение, смешанное обучение.

1 Introduction

In 2012, Massive Open Online Courses (MOOCs) made a real sensation in the higher education sector, providing open access through the Internet to the best courses from the best professors and universities of the world [1]. For the last 6 years the number of MOOCs and open education platforms has continuously grow around the world. These MOOCs platforms are developing together with universities evolving into a new market of higher online education providing massive online specializations, credentials and academic degrees [3]. If we look at the numbers, now there have been released more than 7,000 online courses (Figure 1) from above 750 universities and institutions, which are located in more than 40 MOOCs resources where up to 60 million users are enrolled [2]. This numbers are given only according to the data of the Class Central MOOCs aggregator where many other online courses and providers are not taken into account.

The Learning Management System (LMS) allows to collect detailed information about the users' activities and interactions with course content during the learning in the online course. These data are actively used by researchers to improve the quality of educational resources and improve the content of online courses, as well as a deeper understanding of the learning process in online format and other practical purposes (see e.g., [4], [5]). In addition, the accumulated data is sufficiently large to facilitate the development of intelligent LMS and new methods of active learning in the future.

One of the negative indicators of MOOCs is a large dropout rate [6]. But in many cases they do not take into account the fact that learners participate in the MOOCs with different initial intention and motivation [7]. If the traditional university courses are mainly attended by full-time students whose main activity is studying, then MOOCs participants are mostly employed people with tertiary education [8]. According to statistics it is known that for in MOOCs about half enrolled students never engage with any of the content [9]. Most of the students do not reach the end of the course due to lack of time or lack of digital and learning skills for studying by online courses [10]. Therefore, the classification signed up for MOOCs students in their initial motivation will help determine the exact causes of failure and to understand how to improve the course to achieve their goals.

In this paper we try to describe some finding about our MOOCs learners and classify them by their activities. Also we try to answer to the following questions:
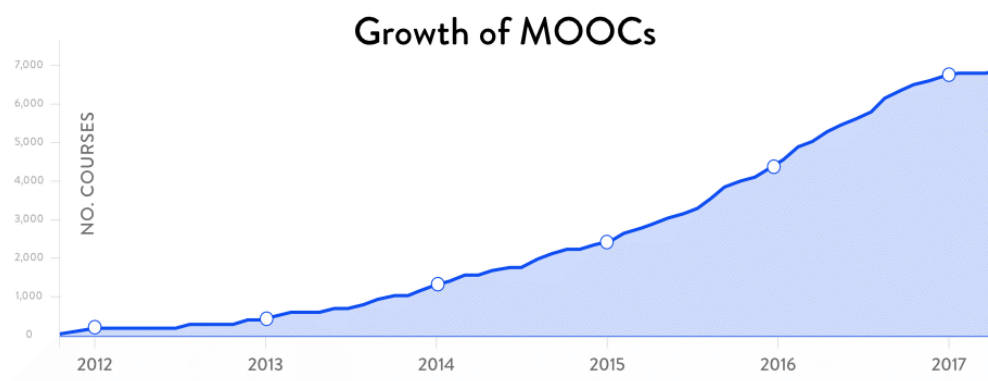
Figure 1: Growth of MOOCs number from 2012 according to Class Central data [3]

1. What is the motivation of each MOOCs learner?

2. How they interact with each other and teaching staff?

3. What they should do to successfully finish the course?

This findings and question answers can help to understand MOOCs developers and providers how improve course content, schedule and delivery methods, also policy makers and administration of universities can evaluate of MOOCs potential to include in academic process in appropriate blended learning model.

2 Literature Review

Several investigators (e.g. [11], [12]) expect that MOOCs can play an important role in future of global education system and even change it. The popularity of MOOCs has made a high volume of learner data available for analytic purposes. A number of scientists began to perform relative researches based on MOOCs data recently, which mainly focus on two aspects. The first is how to improve the MOOCs platform in personalization or to provide new features for both learners and instructors. For example, J. J. Williams and B. Williams [13] investigated how varying reminders and resources sent through emails to participants influence their use of course components like forums and their overall outcomes. C. Shi et al. [14] introduce VisMOOC, a visual analytic system to help analyze user learning behaviors by using video clickstream data from MOOC platforms. Kennedy et. al. [15] analyzed the relationship between a student's prior knowledge on end-of-MOOC performance.

The second aspect is to explore cognitive rules of learner by analyzing learning behavior and therefore to predict their following actions such as whether he will fall out the course. Predicting student performance in MOOCs is a popular and extensive topic. Kizilcec et al. [16] presented a simple, scalable, and informative classification method that identifies a small number of longitudinal engagement trajectories in MOOCs. Learner classification can be fulfilled by different criteria. Researchers from Stanford [17] divided learners into five categories by analyzing learning activities such as viewing a lecture and handing in an assignment for credit: Viewers, Solvers, All-rounders, Collectors, and Bystanders. Researchers

from MIT [18] divided learners into four types based on whether or not they participated in the class forum or helped edit the class wiki pages: passive collaborator, wiki contributor, forum contributor, and fully collaborative.

Many researchers' works based on Person-Course Dataset AY2013 [19] which is provided by HarvardX-MITx (e.g. [20]) and others used CAROL Learner Data [21] by Stanford University (e.g. [22]). In this work we used learners data which collected during providing online courses in the Al-Farabi KazNU's own MOOCs platform [23].

3 Material and methods

3.1 MOOCs by Al-Farabi KazNU. Data Description

In 2014 al-Farabi KazNU became the first from Kazakhstani universities, which have joined the MOOCs movement and began work on producing own online courses. Initially, as the target audience was selected prospective students: graduates from secondary schools, students of vocational schools and colleges. Since the project was an initiative, funds for the development and delivery of courses was not provided. Despite the high teaching load and other professional duties of teaching staff we found and revealed among them enthusiasts and volunteers, which agreed to create courses in the new format.

In 2015 al-Farabi KazNU launched the first MOOCs for high school and undergraduate students. Since then we have developed, tested and implemented in the educational process of the University more than 35 courses in Kazakh, Russian and English. Currently on our website for open education registered more than 12 000 users and over 7 000 of them are actively studying the provided courses. Over the past years to KazNU MOOCs was enrolled about 7 200 learners: 6 624 from Kazakhstan, 432 from other countries of CIS, 42, 58 and 43 from EU, Asia and other countries respectively.

In the table 1 there are demographic characteristics of the platform learners. As you can see in he table most of learners are females, under the age of 25, with bachelor or associate degree. Average age of learners is equal to 25.2 and median age is 21.

Table 1: Demographic characteristic of learners (None means not provided)

Age between	Male	Female	None	All
under 25	1 768	3 103	9	4 880
25 and 35	315	735	1	1 051
36 and 50	173	650	0	823
over 50	92	231	0	323
None	37	78	7	122
Level of education				
Doctorate	108	285	0	393
Master's degree	314	922	3	1 239
Bachelor's degree	1 187	2 458	1	3 646
High school	669	864	4	1 537
Other	107	268	9	384
Total	**2 385**	**4 797**	**17**	**7 199**

Before the launch of the course, various marketing events were held to gather as much as possible the audience of learners. Most of the courses were conducted in the framework of

programs for the refresher courses of teaching staff from universities and secondary schools. All certificates provided freely. That is why learners had very high motivation to get certificates and many courses have high completion rates than in usual MOOCs. In the table 2 the most popular courses are listed, where the number of successfully completed a course learners as well as external students from this number are indicated. Here external means MOOCs students from other institutions.

Table 2: Most popular online courses by Al-Farabi KazNU

Title of the course	Language	Enrolled	Completed	External
Management	English	699	508	212
Selected Issues of Inorganic Chemistry	Kazakh	450	169	90
Biophysics	Russian	186	32	32
Branding	Kazakh	405	206	109
Ethnography of the World Nations	Kazakh	326	177	41
Al-Farabi and Modernity	Kazakh	993	535	341
Constitutional Law of the RK	Russian	1028	650	318
Conflictology	Russian	129	68	45
Law Enforcement Bodies of the RK	Russian	370	107	46
Statistics	Kazakh	510	190	121
Probability Theory	Russian	553	125	25
Solving Physical Problems with prof. V. Kashkarov	Russian	725	68	12
Methods of Ethnological Research	Kazakh	454	276	109
Total	**1+6+6**	**6828**	**3111**	**1501**

The dataset collected from this courses (about 3 GB of JSON and CSV data) is used for analyzing users activity and classify them by their behavior. Initial row data is analyzed and reduced to 10% of original volume by dropping the insignificant attributes, personal data and the records with inconsistent and administrative information. Then we performed denormalization of the tables (*users*, *enrollments*, *certificates* and *tracking logs*) to get one universal table with the records where the most informative attributes collected. Below the attributes of cleaned dataset and their description are described: *user_id*: deidentificated id number of user; *course_id*: id of the courses; *viewed*: anyone who accessed the 'Courseware' tab; *explored*: anyone who accessed at least half of the chapters in the course content; *certified*: earned or not a certificate; *level of education*; *gender*; *year of birth*; *grade*: final grade of the course, ranged from 0 to 1; *start_time*: date of course registration; *last_time*: date of last interaction with course, blank if no interactions; *nevents*: number of the interactions with the course; *ndays_act*: number of unique days student interacted with course; *nvideo_plays*: number of play video events within the course; *nchapters*: number of chapters with which the learner interacted; *nforums*: number of posts to the Discussion Forum.

3.2 Methodology

Due to the great diversity of learners in age, education background, region, motivation and learning habits predicting of their successful course completing is a big challenge. Online survey is a good option to recognize learners' motivation, but most of them may not respond to an online survey. Therefore, learning activities may reflect a learner's motivation. The

detailed records of learning activities in MOOCs platforms give us a chance to analyze a learner's motivation. Learners have different goals when following a MOOC. These goals are reflected in their behaviour patterns when following the course. Hill [24] has identified five categories of learners' behaviour in a MOOC:

- **No-shows:** register but never log in to the course whilst it is active.

- **Observers:** log in and may read content or browse discussions but do not take any form of assessment followed after videos.

- **Drop-ins:** perform some activity (watch videos, browse or participate in the discussion forum) for a select topic within the course but do not attempt to complete the entire course.

- **Passive participants:** view a course as content to consume. They may watch videos, take quizzes and/or read discussion forums but generally do not engage with the assignments.

- **Active participants:** fully intend to participate in the MOOC and take part in discussion forums, the majority of assignments and all quizzes.

A recent study by Wang and Baker [25] has shown that participants who expected to finish a MOOC were more likely to do so than participants who did not think they would complete the course. This motivation in the category of "active participants" is a good predictor for completing a MOOC. Although this finding is in line with the findings of other studies, the authors concluded that further research is needed to gain more insight into the motivations of MOOC participants and how these relate to MOOC design, in order to provide a learning experience worthwhile for a large community of learners.

Figure 2 displays the learners which have some activity entries, mark above 0 and those who obtained a certification. The average certification rate of 13 courses is 22,7%.

4 Results and Discussion

Figure 3 shows the average total activities and average video watchings of two group learners with mark above and equal to 0. The difference between these two groups is apparently huge especially in total events and activities. The average activities of learners with mark above 0 are three times more than learners with no mark at least. If a learner wants to earn certificate, he/she will spend more time on this course.

Based on the above analysis, learners can be divided into different categories according to their activities. Learners in category *Active participants* have highest activities while learners in category *Observers* have lowest ones. An activity index value was proposed to measure the engagement of a learner. According to above statistics, if a learner spend more time (days) in one course or with higher activities especially video playing events, he/she should obtain a higher grade value, while if a learner enrolled on too many courses, the engagement in one course will be less.

Understanding the reasons behind dropout rates in MOOCs and identifying areas in which these can be improved is an important goal for MOOC development. Many widely-quoted
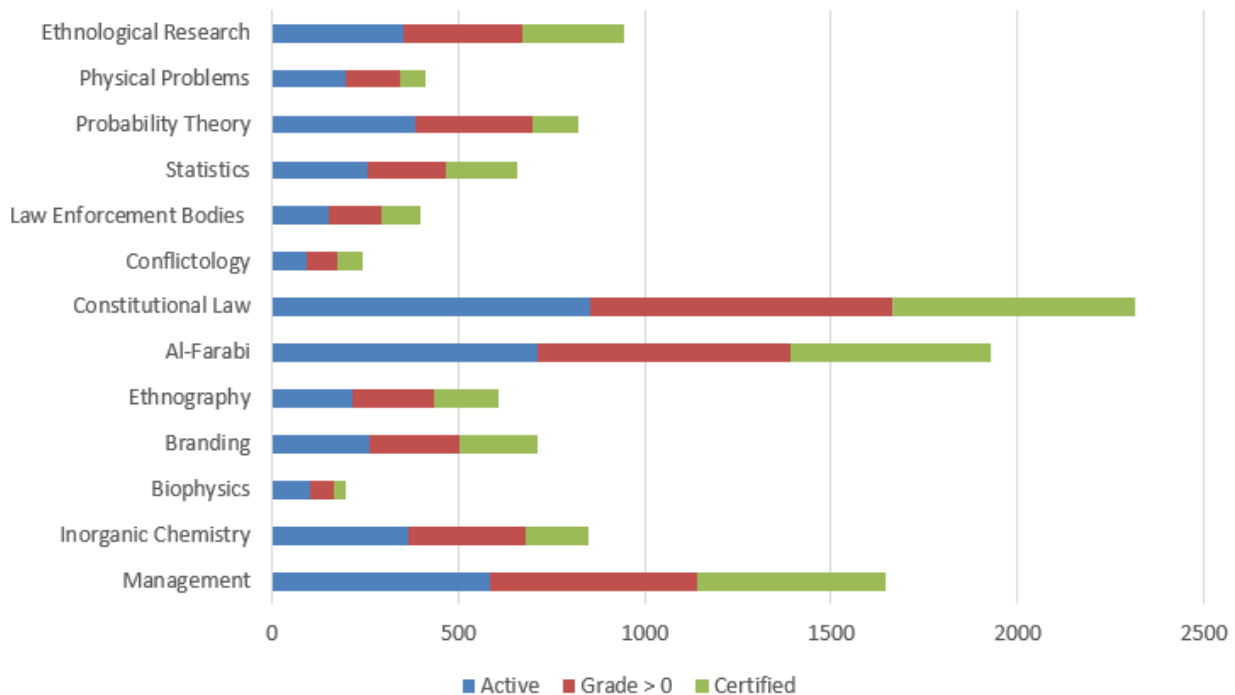
Figure 2: Course statistics on active, mark above 0 and certified learners

dropout rates are calculated from baseline numbers which include registrations by people who never engage with the course or who engage in their own way but without completing assessments. Despite this, it is clear that many of those who do wish to follow and complete a course are hindered by factors such as level of difficulty, timing and lack of digital and learning skills. These problems become even more acute when MOOCs are proposed as a replacement for traditional teaching (rather than just free, spare time activities) and particularly when they are suggested as the means to close gaps in education.

5 Conclusion

Employing of educational data and learning analytics to improve lesson plans and course delivery become an innovative approach for teachers, curriculum developers and policy makers in education. Learning analytics of online courses can be also used for enhancement of classroom teaching by blending online and face-to-face learning models. In the figure 4 weekly learners engagement in the course can be useful information for the teaching staff to apply motivating posts or emails to learners when learners become inert. Also learning analytics can help to identify flush of activity and reasons of occurrence which can help apply right conducting strategy during the course delivery.

In this paper, we first made an analysis about learning behaviors of learners in MOOCs and explored the differences and characteristic of learning behavior features between the learners with different grades. MOOCs learning is one kind of high-level behavior, learners may be influenced by many incentive factors of ultimate goal. The passing rate maybe not enough incentive for some learners. Survey research has bolstered the notion that effective
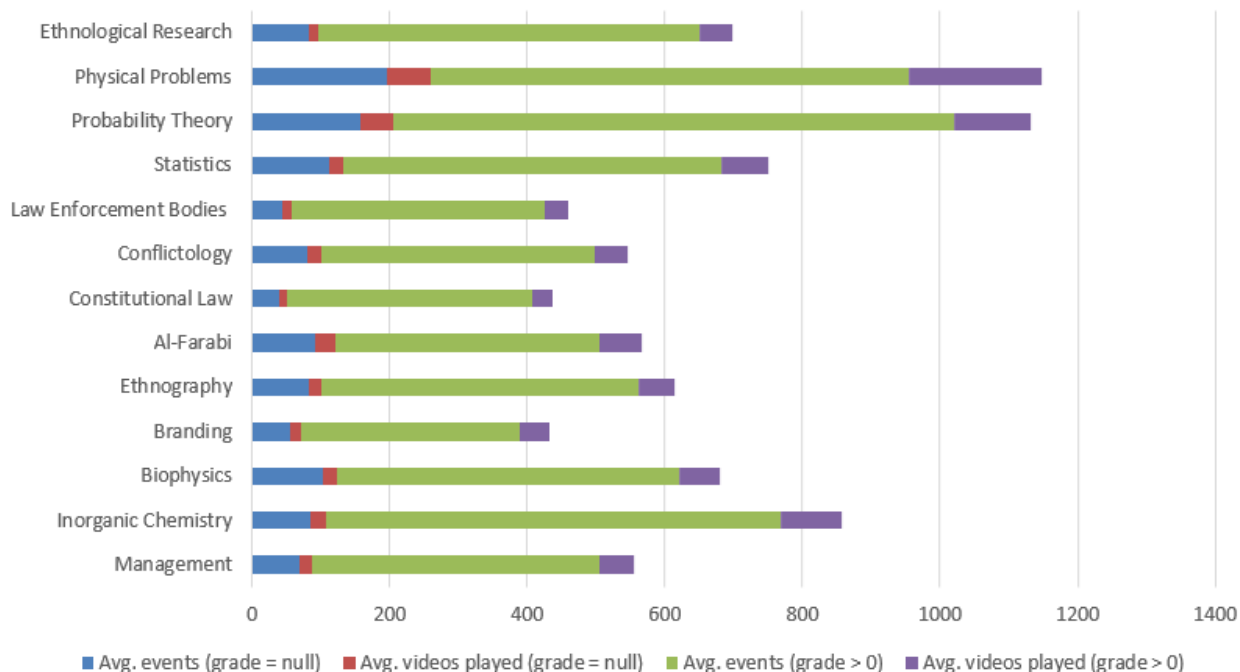
Figure 3: Average activities learners with grade above and equal to 0

learner–instructor and learner–learner interactions are critical to effective online learning and concluded that increased interpersonal interaction within the framework of the course, either with the instructor or with learner peers, positively affects student learning.
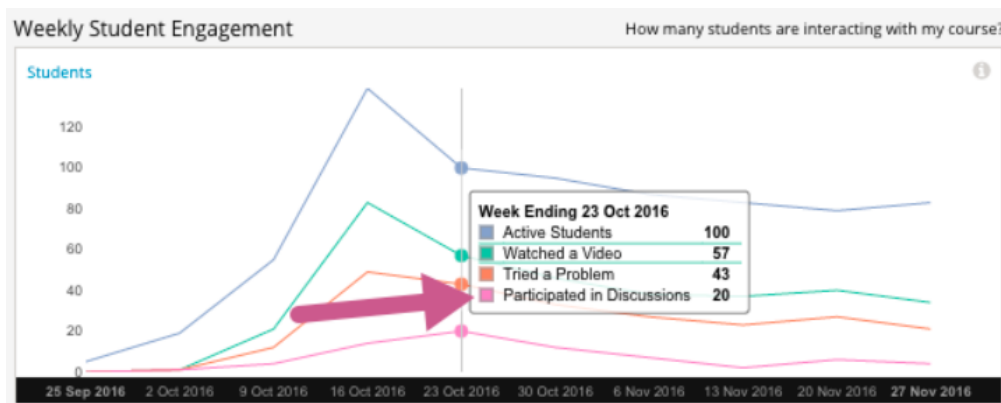
6 Acknowledgements

Figure 4: Weekly learners engagement chart of "Solving Physical Problems" online course

References

[1] "Pappano L. The year of the MOOC accessed June 15, 2018, http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

[2] "Shah D. MOOC Providers Target Degrees accessed June 15, 2018, https://www.class-central.com/report/moocwatch-16-mooc-providers-target-degrees/

[3] Shah D. *"EdSurge, Monetization Over Massiveness: Breaking Down MOOCs by the Numbers in 2016"*, (2016).

[4] Jordan K. Initial Trends in Enrolment and Completion of Massive Open Online Courses. *The International Review of Research in Open and Distributed Learning*, 15(1): 133–160, (2014).

[5] O'Reilly U.-M., Veeramachaneni K. Technology for mining the big data of MOOCs. *Research & Practice in Assessment*, 9(2): 29—37, (2014)

[6] Jordan K. *"MOOC completion rates: The Data*, accessed June 15, 2018, http://www.katyjordan.com/MOOCproject.html

[7] Reich J. *"MOOC Completion and Retention in the Context of Student Intent"*, accessed July 15, 2018, https://er.educause.edu/articles/2014/12/mooc-completion-and-retention-in-the-context-of-student-intent

[8] Christensen G., Steinmetz A., et al. "The MOOC Phenomenon: Who Takes Massive Open Online Courses and Why?"(2013). accessed July 15, 2018, http://dx.doi.org/10.2139/ssrn.2350964

[9] Newman J., Oh S. "8 Things You Should Know About MOOCs (2014), accessed July 15, 2018, http://www.chronicle.com/interactives/moocs_stats

[10] Onah D.F.O., Sinclair J., Boyatt R. Dropout Rates of Massive Open Online Courses: Behavioural Patterns. *In: EDULEARN14 Proceedings*, (2014):5825–5834, accessed July 15, 2018, https://warwick.ac.uk/fac/sci/dcs/people/research/csrmaj/daniel_onah_edulearn14.pdf

[11] Stein L.A. Casting a Wider Net, *Science*, 338(6113): 1422–1423, (2012)

[12] Patru M., Balaji V. Making Sense of MOOCs: A Guide for Policy-Makers in Developing Countries. *Paris, UNESCO* (2016), accessed July 15, 2018, http://unesdoc.unesco.org/images/0024/002451/245122E.pdf

[13] Williams J.J., Williams A. Using interventions to improve online learning. *In: Proc. of the Neural Information Processing Systems, Workshop on Data Driven Education*, (2013)

[14] Shi C., Fu S., et al. VisMOOC: Visualizing Video Clickstream Data from Massive Open Online Courses. *IEEE Pacific Visualization Symposium*, Hangzhou, China, (2015)

[15] Kennedy G., Coffrin C., et al. Predicting success: how learners' prior knowledge, skills and activities predict MOOC performance. *In: Proc. of the Fifth International Conference on Learning Analytics And Knowledge, ACM*, (2015):136-140.

[16] Kizilcec R., Piech C., Schneider E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. *In: Proc. of the Int. Conf. LAK '13*, (2013):170—179.

[17] Anderson A., Huttenlocher D., et al. Engaging with massive online courses. *In: Proc. of the Int. Conf. WWW'14*, (2014):687-–697.

[18] Taylor C., Veeramachaneni K., O'Reilly U.-M. "Likely to stop? Predicting stopout in massive open online courses accessed July 15, 2018, http://arxiv.org/abs/1408.3382

[19] "Harvard Dataverse: HarvardX-MITx Person-Course Academic Year 2013 De-Identified dataset, version 2.0 (2014), accessed July 15, 2018, http://dx.doi.org/10.7910/DVN/26147

[20] Northcutt C.G., Ho A.D., Chuang I.L. "Detecting and Preventing "Multiple-Account" Cheating in Massive Open Online Courses (2015), accessed July 15, 2018, https://arxiv.org/abs/1508.05699v3

[21] "CAROL Learner Data accessed July 15, 2018, http://datastage.stanford.edu/

[22] Ren Zh., Rangwala H., Johri A. "Predicting Performance on MOOC Assessments using Multi-Regression Models (2016), accessed July 15, 2018, https://arxiv.org/abs/1605.02269v1

[23] "Al-Farabi KazNU's MOOCs platform accessed July 15, 2018, http://open.kaznu.kz

[24] Hill P. "Emerging student patterns in MOOCs: A (revised) graphical view [Blog post] (2013), accessed July 15, 2018, http://mfeldstein.com/emerging-student-patterns-in-moocs-a-revised-graphical-view/

[25] Wang Y., Baker R. Content or platform: Why do students complete MOOCs? *MERLOT Journal of Online Learning and Teaching*, 11(1): 17–30, (2015), accessed July 15, 2018, http://jolt.merlot.org/vol11no1/Wang_0315.pdf