Demographic Indicators Influencing Learning Activities in MOOCs: Learning Analytics of FutureLearn Courses

Lei Shi *lei.shi@liverpool.ac.uk*
University of Liverpool
Liverpool, UK

Alexandra I. Cristea *alexandra.i.cristea@durham.ac.uk*
Computer Science, Durham University
Durham, UK

Abstract

Big data and analytics for educational information systems, despite having gained researchers' attention, are still in their infancy and will take years to mature. Massive open online courses (MOOCs), which record learner-computer interactions, bring unprecedented opportunities to analyse learner activities at a very fine granularity, using very large datasets. To date, studies have focused mainly on dropout and completion rates. This study explores *learning activities* in MOOCs against their *demographic indicators*. In particular, pre-course survey data and online learner interaction data collected from two MOOCs, delivered by the University of Warwick, in 2015, 2016, and 2017, are used, to explore *how learner demographic indicators may influence learner activities*. Recommendations for educational information system development and instructional design, especially when a course attracts a diverse group of learners, are provided.
Keywords: demographic analytics, learning analytics, massive open online courses, MOOCs.

1. Introduction

Since the launch of the first three major massive open online courses (MOOCs) platforms, Coursera, Udacity and edX, in 2012, the landscape has grown to reach a total of 57 MOOC platforms, 9,400 courses, more than 500 MOOC-based credentials, and approximately 100 million learners worldwide, by the end of 2017 [30, 31]. With the rapid advancement of eLearning technologies, MOOC platforms have been experiencing a massive increase in the amount of learner data collected. Along with the development of data analytics techniques, this brings unprecedented opportunities to explore learner behaviours and behavioural patterns, which in turn may help enhance MOOC platforms and their design, and ultimately improve learning experience and outcomes.

FutureLearn, founded in December 2012, is a joint initiative of the UK universities, backed by the UK government, created to alleviate the increasing domination by USA's MOOC platforms. The first FutureLearn MOOCs were launched in September 2013. As of June 2018, FutureLearn has 143 UK and international partners, including non-universities, and more than 7.9 million people have joined FutureLearn [9], which tops it as one of the five most popular MOOC platforms worldwide by registered users [30]. As a growing MOOC platform, FutureLearn has demonstrated their commitment to support partners on co-implementing effective solutions to improve research opportunities and, ultimately, the learner experience. FutureLearn MOOCs collect complete records of all learner activity data. The dataset used in this study was extracted from the FutureLearn platform, in particular, from six runs of two MOOCs delivered by the University of Warwick.

FutureLearn employs a social constructivist approach, inspired by Laurillard's Conversation Framework [13, 16], which, in brief, describes a general theory of effective learning through conversation. The aims include allowing for multimedia resources, collaborative learning, and opportunities for tutorial intervention and guidance [8]. However, one of the main challenges has been to keep learners motivated in performing desired learning behaviours and achieving learning goals [26]. Motivational theories, such as self-determination theory (SDT) [20, 23] and techniques, such as gamification and social interaction [22], have been influencing the improvement of the system development and the instructional and pedagogical design of MOOCs. Other techniques, such as open social user modelling [25], opening (visualising) learner data for the learners or for other parties, have also been inspiring the engagement strategy development in MOOCs. Since many techniques and strategies have been implemented in MOOCs, there is a strong need to examine how they influence learner activities.

To date, most studies have focused on dropout rates and completion rates of learners [7, 10, 19, 28]. This study was conducted at a relatively finer-grain level – investigating learner *demographic indicators*, including gender and age, against their *learning activities*, including following the courses, discussions in the forums, learning material visits and quiz attempts, on two MOOCs delivered by the University of Warwick in 2015, 2016 and 2017, respectively. The pre-course surveys were used to collect learner *demographic data*; whilst the system logs were used to collect learner *activity data*. These two datasets were linked together using the unique and anonymous Learner IDs, in order to anonymously associate learner *demographic indicators* with their *activities* in MOOCs.

The results of the study revealed statistically significant differences of learning activities among different groups of learners categorised by different ways using the demographic indicators. This paper reproduces the process of the study and discusses the results.

2. Related Work

Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the information system in which it occurs [29]. It combines expertise from different academic disciplines, such as predictive modelling [21]. It is overlapping with another two rapidly developing fields, i.e., educational data mining and academic analytics, yet learning analytics is concerned with enhancing aspects of learning [6].

Learning analytics is influenced by a wide range of disciplines, including education, psychology, philosophy, sociology, linguistics, learning science, statistics, intelligence and computer machine learning/artificial science [21]. Various tools and approaches have been in use in learning analytics, to provide educators and designers with quantitative intelligence, to make informed decisions about student learning. Data is collected from a broad range of sources, including behavioural data taken from online learning systems, such as discussion forums, activity completion, assessments, and functional data taken from student admissions systems and progress reports [27]. Learning analytics has been used in many application areas, such as modelling of user knowledge, user behaviour and user experience; user profiling; modelling of key concepts in a domain and modelling a domain's knowledge components; trend analysis; and adaptation and personalisation of user experience [14].

Learning analytics provides a method for identifying factors influencing retention, which enables MOOC providers to make improvements of the learning context, design and pedagogies, where appropriate; the large datasets collected in MOOC activities provide strong support for this type of method [7].

Bote-Lorenzo and Gómez-Sánchez [3] discussed the decrease of engagement indicators using learner activity data. Those indicators were derived for the main tasks carried out in a MOOC, including watching lecture videos, solving short and simple comprehension questions interspersed in the videos (called 'finger exercises'), and submitting assignments. The results supported the possibility of detecting disengaging learners in the MOOC. Khalil and Ebner [12] used clustering techniques to portray learner engagement in MOOCs. Their study clustered learners based on learners' engagement level. The results recommended adding intrinsic factors to improve future MOOCs. The study conducted by Kahan, *et al.* [11], characterised different types of learning activities in a MOOC using data mining techniques, which clustered learners based on their activities in relation to the main learning resources of the MOOC, including video lectures, discussion forums and assessments. The results supported the claim that MOOCs' influence should not be evaluated solely based on certification rates, but rather based on learning activities. In their study [15], Morris *et al.* argued that four learner demographic indicators were significantly associated with the degree of completion, namely age, online experience, educational attainment and employment status.

In this study, learning analytics techniques were used to collect, analyse, and report data about learner activities, to understand *how learner demographic indicators may influenced learning activities*, in the FutureLearn MOOC context. The learner demographic indicators considered include gender and age. Different from previous studies, instead of investigating dropoff and completion rates [5, 7, 10, 19, 28], this study focused on a finer-grain level – the influences of these learner demographic indicators on learning activities, including following a course, discussions in the forums (comments), learning material visits and attempts to answer questions in quizzes; instead of predicting learning performance [2, 4, 18], the results may be able to shed light on the importance and possibility of personalisation and early intervention in MOOCs.

3. The Method

3.1. Study Settings

FutureLearn MOOCs are organised in *weekly learning units*. They consist of a set of *learning blocks*, which may contain one or several *steps*, which are the basic learning items. *Steps* may include articles, images, videos, and quizzes. Fig. 1 shows the navigation page of a MOOC, where a learner can click on the WEEK button on the top to navigate to a *weekly learning unit* or click on the *step* title to navigate to the *step* page.

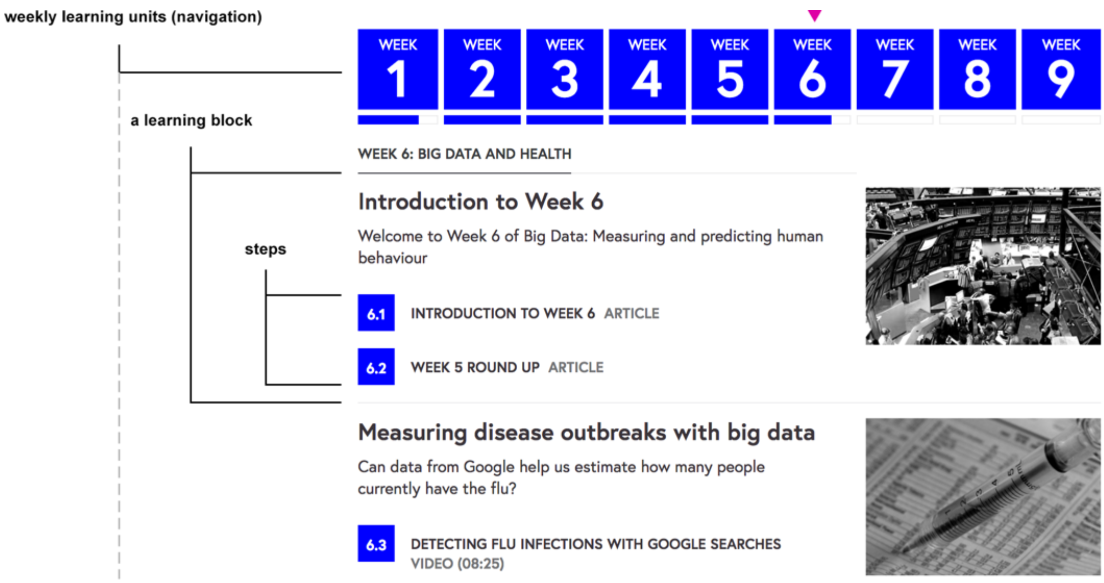

Fig. 1. The navigation page of a MOOC.

Fig. 2 demonstrates the interaction component on a *step* page (on the left). Using the interaction component, learners can navigate to the *last step* and the *next step*, by clicking on the arrows at the bottom of the *step* page; they can click on the button "Mark as complete", to claim that they have completed the current *step*. To read or submit comments (discussions), they firstly click on the pink "plus" button on the left, so that the comment component shows (Fig. 2 on the right). Learners can then also declare they "like" comments of their own or written by others (as in Facebook, Weibo and Zhihu).

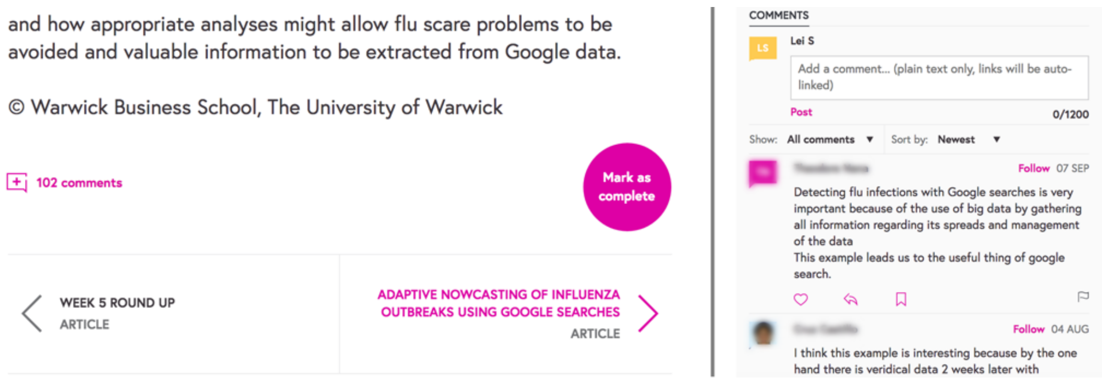

Fig. 2. The interaction component (left) and the comment component (right) on a *step* page.

The study was to explore the influence of learners' *demographic indicators* (gender and age) on their *activities* in two MOOCs delivered by the University of Warwick. The first MOOC, "Big Data: Measuring and Predicting Human Behaviour", aimed to introduce learners with an overview on the state of the art in 'big data' research across a range of domains, including economics, crime and health, as well as teach them basic practical skills for data science, including writing basic programs in R, creating basic data visualisations and carrying out simple analyses. This MOOC was broken down into 9 weekly learning units: 8 units of study with a break for reflection in the fifth unit. Each weekly learning unit contained a sequence of individual *steps* to complete. There were 11 *steps* in week 1, 12 *steps* in week 2, 15 *steps* in week 3, 11 *steps* in week 4, 4 *steps* in week 5, 11 *steps* in week 6, 12 *steps* in week 7, 12 *steps* in week 8, and 14 *steps* in week 9. In total, MOOC 1 had 102 *steps*. Learners were learning by watching videos, reading articles and taking part in discussion activities (writing and reading comments on *step* pages). Learners were asked to do a quiz in weeks 2, 3, 4, 6, 7, 8, 9, respectively. In each quiz, there were 5 questions, thus there were 35 questions in total within the MOOC.

The second MOOC, "The Mind is Flat", aimed to present how understanding human being's minds could help recognise some of the surrounding social and economic forces, from market booms and crashes, to the origin of communication and language, to human being's mysterious collective ability to construct societies. This MOOC was broken down into 6 weekly learning units, each of which consisted of several *steps*. There were 14 *steps* in week 1, 12 *steps* in week 2, 14 *steps* in week 3, 12 *steps* in week 4, 12 *steps* in week 5, and 18 *steps* in week 6. In total, MOOC 2 had 82 *steps*. Most *steps* contained videos, whilst a few contained only articles. Learners could use the comment component on *step* pages for discussions. There were 10 quizzes each week, thus there were 60 questions in total.

The *learner activities* that this study focused on included clicking on the button "Mark as complete", submitting a comment, and attempting to answer a question in a quiz.

The first MOOC was a STEM[1] course; whereas the second MOOC was a non-STEM course. The reason of choosing these two courses was thus to compare demographic indicators and learning activities between different disciplines of courses.

Table 1 shows the number of *weekly learning units* and the number of *steps* within both MOOC 1 and MOOC 2.

Table 1. The number of *weekly learning units* and *steps* within MOOC 1 and MOOC 2

	MOOC 1	**MOOC 2**
The number of weekly learning units	9	6
The number of steps	102	82
The number of questions in quizzes	35	60

[1] STEM stands for Science, Technology, Engineering and Mathematics. It is a term used to group academic disciplines, in order to address education policy and curriculum choices.

3.2. Data Collection

This study was conducted in accordance with the FutureLearn Code of Practice for Research Ethics[2]. All data was completely anonymous – thus individuals could not be identified by any means. Two data sources were used in this study: 1) responses from a pre-course survey, and 2) system logs populated by learners. Each record (either a response or a system log) had a unique Learner ID, linking both sources. The pre-course survey was sent by email, either when a learner first joined FutureLearn, in case of a new FutureLearn user, or first enrolled on a new course, in case of an existing FutuerLearn user. Learners might have also completed the survey by visiting the URL directly. The optional questions on the survey included their gender and age range. System logs were generated nightly by the FutureLearn platform from the course start until two weeks after it ended. System logs contained the information about learning activities, such as visiting a *step* page, clicking on the button "Mark as complete" (Fig. 2), submitting a comment, and attempting to answer a question on a *step* (quiz) page.

3.3. The Dataset

Each MOOC ran three times. For MOOC 1, Run 1 was in spring 2015; Run 2 was in spring 2016; and Run 3 was in spring 2017. For MOOC 2, Run 1 was in autumn 2015; Run 2 was in spring 2016; and Run 3 was in autumn 2016.

In MOOC 1 Run 1, there were initially 16,329 learners enrolled, yet 2,222 of them proactively unenrolled from the course. Thus, the number of remaining learners were 16,329 – 2,222 = 14,107. Additionally, the learners who did not visit any *step* pages were considered to be irrelevant and thus removed from this study. Therefore, in MOOC 1 Run 1, there were 6,631 learners considered in the study. Using the same method of filtering, 4,094 learners were considered in MOOC 1 Run 2, and 3,571 in MOOC 1 Run 3. Thus, in total, 6,631 + 4,094 + 3,571 = 14,296 learners from MOOC 1 were able to be considered in the study. The resulting relevant learner rate was thus 14,296 / 29,343=48.72%.

The same filtering process was applied to MOOC 2 (see Table 2) resulting in a relevant learner rate of 12,068/30,010=40.21%.

In summary, in total, 14,296 + 12,068 = 26,364 learners were considered in this study. The total relevant learner rate was (14,296 + 12,068) / (29,343 + 30,010) = 44.42% (Table 2).

Table 2. The numbers of learners being considered in the study.

	MOOC 1				**MOOC 2**			
	Run 1	Run 2	Run 3	**Total**	Run 1	Run 2	Run 3	**Total**
Enrolled	16,329	11,258	5,753	**33,340**	13,446	14,240	7,511	**35,197**
Unenrolled	2,222	1,355	420	**3,997**	2,087	2,030	1,070	**5,187**
Remaining	14,107	9,903	5,333	**29,343**	11,359	12,210	6,441	**30,010**
Considered in the study	**6,631**	**4,094**	**3,571**	**14,296**	**4,421**	**4,992**	**2,655**	**12,068**
Relevant learner rate	*47.0%*	*41.3%*	*67.0%*	***48.7%***	*38.9%*	*40.9%*	*41.2%*	***40.2%***

4. Analysis

4.1. Learner Demographic Indicators Influencing Following MOOCs

Note that the demographic analytics below is only relevant under the assumption that responding the pre-course survey is independent of the demographic indicators. For example, females and males are equally likely to respond to the survey.

Table 3 details learners' gender and age range, as per learners' responses to the pre-course survey. Most learners, i.e., more than 90%, did *not* answer the optional questions.

[2] Research Ethics for FutureLearn, https://about.futurelearn.com/terms/research-ethics-for-futurelearn

Table 3. The learner demographic indicators.

		MOOC 1				MOOC 2				Grand Total
		Run 1	Run 2	Run 3	Total	Run 1	Run 2	Run 3	Total	
Gender	Female	136	196	163	**495**	188	496	224	**908**	1403
	Male	167	251	154	**572**	113	244	93	**450**	1022
	Nonbinary	1	1	1	**3**		3	1	**4**	7
	Other	2	5		**7**	3	4	1	**8**	15
	Unknown	6,325	3,641	3,253	**13,219**	4,117	4,245	2,336	**10,698**	23,917
	Total	**6,631**	**4,094**	**3,571**	**14,296**	**4,421**	**4,992**	**2,655**	**12,068**	**26,364**
Age Range	<18	1	0	10	**11**	1	3	5	**9**	20
	18-25	55	32	26	**113**	48	128	58	**234**	347
	26-35	12	50	42	**104**	17	68	28	**113**	217
	36-45	53	127	79	**259**	49	92	34	**175**	434
	46-55	45	97	68	**210**	50	98	42	**190**	400
	56-65	56	72	49	**177**	63	178	71	**312**	489
	>65	69	62	37	**168**	74	163	82	**319**	487
	Unknown	6,340	3,654	3,260	**13,254**	4,119	4,262	2,335	**10,716**	23,970
	Total	**6,631**	**4,094**	**3,571**	**14,296**	**4,421**	**4,992**	**2,655**	**12,068**	**26,364**

4.1.1. Gender Indicator influencing Following MOOCs

Fig. 3 shows the gender distribution. Amongst the 26,364 relevant learners considered in the study, there were 2,447 learners answered the questions about their gender asked in the pre-course survey (1,077 from MOOC 1, and 1,370 from MOOC 2), with an overall response rate of 2,447 / 26,364 = 9.28%. In MOOC 1, 495 learners disclosed their gender as female, 572 as male, 3 as "nonbinary", and 7 as "other". As the "nonbinary" and "other" only represented a very small proportion (0.9%), to simplify the procedure, in the following analyses, we only take into consideration the "female" and "male" gender categories. The result shows that the female male ratio was 0.865 in MOOC 1. The gender gap was more prominent in MOOC 2: 908 learners disclosed their gender as female, and 450 as male, with the female male ratio of 2.018. A chi-square test revealed significant gender differences in choosing the two MOOCs (χ^2=102.697, $p < .01$). This suggests that in comparison with male learners, female learners are more underrepresented in STEM fields, and vice-versa, which has consistently been reported in the literature, e.g. [17].

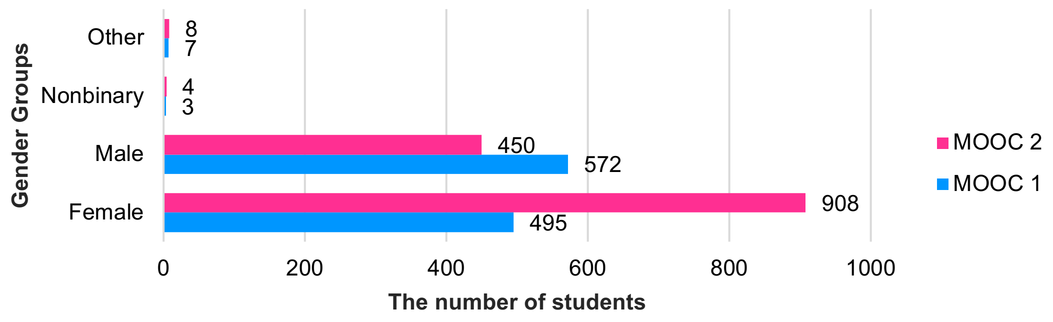

Fig. 3. Gender distribution.

4.1.2. Age Group Indicator influencing Following MOOCs

Fig. 4 shows the age group distribution. There were 2,394 learners who disclosed their age group (1,042 from MOOC 1, and 1,352 from MOOC 2), with an overall response rate of 2,394 / 26,364 = 9.08%. In MOOC 1, 11 learners claimed to be in age group <18, 113 in age group 18-25, 104 in age group 26-35, 259 in age group 36-45, 210 in age group 46-55, 177 in age group 56-65, and 168 in age group >65. In MOOC 2, 9 learners claimed to be in age group <18, 234 in age group 18-25, 113 in age group 26-35, 175 in age group 36-45, 190 in age group 46-55, 312 in age group 56-65, and 319 in age group >65. Interestingly, MOOC 2 attracted a larger proportion of older learners i.e. >55 (46.67%) in comparison to MOOC 2 (33.11%). For both MOOC 1 and 2, the age group <18 was the most underrepresented, with percentages of only 1.06% and 0.67%, respectively. A chi-square test was conducted, showing significant age group differences in choosing the two MOOCs (χ^2=105.745, p < .01).

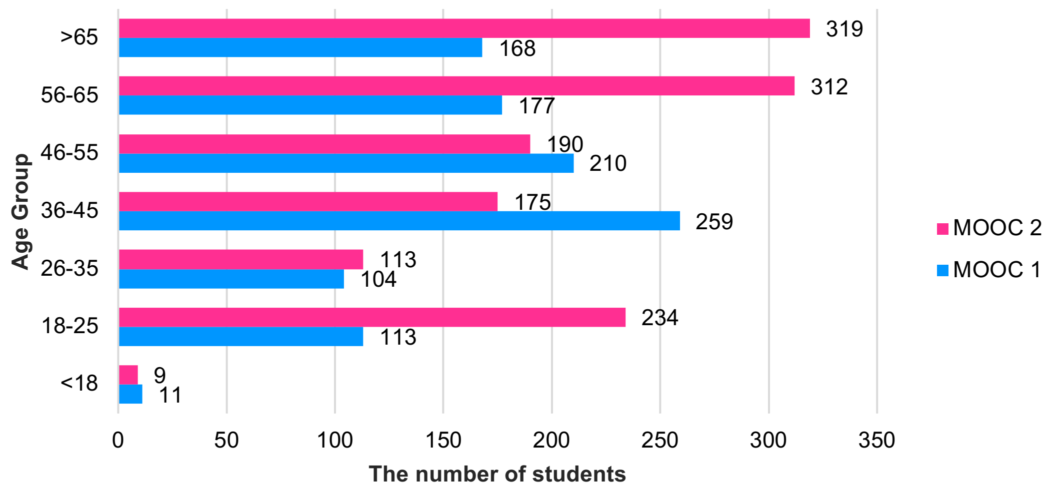

Fig. 4. Age group distribution.

4.2. Learner Demographic Indicators influencing Learning Activities in MOOCs

Table 4 summarises the activities performed by those 26,364 relevant learners (14,296 from MOOC 1, and 12,068 from MOOC 2). From MOOC 1, there were 317,882 distinct visits to *step* pages, 275,596 "completes" marked on distinct *step* pages, 18,938 comments (discussions), and 92,535 attempts to answer a question in a quiz, whilst from MOOC 2, there were 224,839 distinct visits to *step* pages, 200,228 "completes" marked on distinct *step* pages, 29,880 comments, and 179,227 attempts to answer a question in a quiz.

Table 4. The number of activities performed by learners.

Actions	MOOC 1				MOOC 2			
	Run 1	Run 2	Run 3	**Total**	Run 1	Run 2	Run 3	**Total**
Visits	159,488	102,912	55,482	**317,882**	76,021	98,497	50,321	**224,839**
Completes	137,763	91,486	46,347	**275,596**	66,100	88,460	45,668	**200,228**
Attempts	46,196	31,541	14,798	**92,535**	58,547	78,613	42,067	**179,227**
Comments	8,830	7,431	2,677	**18,938**	7,703	16,210	5,967	**29,880**
Total	*352,277*	*233,370*	*119,304*	***704,951***	*208,371*	*281,780*	*144,023*	***634,174***
Grand Total	**1,339,125**							

Visits denotes the number of distinct *step* pages visited; **Completes** denotes the number of *step* pages marked as "complete"; **Attempts** denotes the number of attempts to answer a question in a quiz; **Comments** denotes the number of comments submitted on *step* pages.

As stated in section 3.1, there were 102 *steps* and 35 questions in MOOC 1; 82 *steps* and 60 questions in MOOC 2. As they contained different numbers of *steps* and questions, to compare learner activities between them, we considered the "rates" instead of the actual numbers of *steps* and *attempts*. Here we define the following "rates":

$$R_v = V_s \div S_m \quad (1)$$
$$R_c = C_s \div V_s \quad (2)$$
$$R_a = A_q \div Q_m \quad (3)$$

where R_v represents the "visit rate"; V_s denotes the number of distinct visits to *steps*; S_m is the number of steps in MOOCm ($m \in \{1,2\}$); R_c is the "completion rate"; C_s is the number of steps marked as "complete"; R_a indicates the "attempt rate"; A_q is the number of attempts to answer questions in quizzes; Q_m means the number of questions in MOOCm; $R_v \in \{r|0 \le r \le 1\}$, $R_c \in \{r|0 \le r \le 1\}$, $R_a \in \{r|r \ge 0\}$. Note that the reason that R_a can be greater than 1 is because learners could attempt to answer the same question multiple times.

4.2.1. Gender Indicator influencing Learning Activities in MOOCs

Fig. 5 displays the comparison of the mean visit rate (R_v), mean completion rate (R_c), and mean attempt (to answer questions in quizzes) rate (R_a), between the two gender groups, i.e. female and male, in MOOC 1 and MOOC 2, respectively. Overall, for both MOOCs, all these rates were higher for male learners compared to female learners.

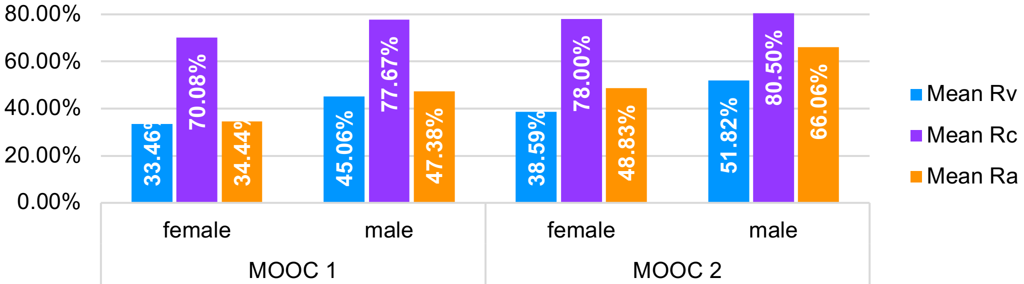

Fig. 5. Mean visit rate **(R_v)**, completion rate **(R_c)** and attempt rate **(R_a)** for female and male learners in MOOC 1 (on the left) and MOOC 2 (on the right).

A Mann-Whitney test shows that, in MOOC 1, the visit rates (R_v) of male learners (Median=27.45%) was significantly larger than that of female learners (Median=15.69%), *u=114,902.5*, *p<.001* (Table 5). Performing the same test for MOOC 2 confirmed the same trend, i.e., the visit rates of male learners (Median=48.78%) being significantly larger than that of female learners (Median=19.51%), *u= 240,604*, *p<.001*.

Table 5. Mann-Whitney tests results for visit rate **(R_v)**, completion rate **(R_c)**, attempt rate **(R_a)**.

		MOOC 1	**MOOC 2**
n		Female: 495, Male: 572	Female: 908, Male: 450
Visit Rate (R_v)	Median	Female: 87.50%, Male: 94.59%	Female: 19.51%, Male: 48.78%
	Mean Ranks	Female: 480.1, Male: 580.6	Female: 639.5, Male: 760.2
	U	114,902.5	240,604
	z	-5.31	-5.34
	p	***<.001***	***<.001***
Completion Rate (R_c)	Median	Female: 87.50%, Male: 94.59%	Female: 90.00%, Male: 96.18%
	Mean Ranks	Female: 489.9, Male: 572.1	Female: 735.8, Male: 651.6
	U	119,749.5	229,642.5
	z	-4.35	-3.73
	p	***<.001***	***<.001***
Attempt Rate (R_a)	Median	Female: 0, Male: 14.29%	Female: 5.00%, Male: 45.83%
	Mean Ranks	Female: 491.7, Male: 570.6	Female: 646.6, Male: 745.9
	U	120,643	234,169
	z	-4.17	-4.39
	p	***<.001***	***<.001***

Similarly, Mann-Whitney tests conducted for completion rates (R_c) and attempt rates (R_a) also show that male learners tended to complete significantly (*p<.001*) more steps and attempt to answer significantly (*p<.001*) more questions in quizzes.

In terms of comments (discussions), the Mann-Whitney tests conducted for MOOC 1 suggest that there are significantly (*p<.001*) more comments (discussions) from male learners

(Median=5.73) than from female learners (Median=4), *u=129,115.5*, *p=0.0066<.05*. In MOOC 2, on average, male learners (Mean=9.75, SD=26.08) tended to produce more comments (discussions) than female learners (Mean=8.26, SD=45.25), but the Mann-Whitney test performed did NOT show a significant difference (*u=214,951.5*, ***p=.0582>.05***).

4.2.2. Age Group Indicator influencing Learning Activities in MOOCs

Fig. 6 shows the comparison of the visit rate (R_v) among different age groups in MOOC 1 and MOOC 2. Fig. 7 shows the comparison of the completion rate (R_c) for different age groups in MOOC 1 and MOOC 2. Fig. 8 shows the comparison of the attempt rate (R_a) for different age groups in MOOC 1 and MOOC 2. Interestingly, overall, the older the learners were, the more activities they performed. From the Kruskal-Wallis test results for both MOOCs, we found statistically significant differences for all these three activity rates, i.e. the visit rate (R_v) (MOOC 1: *H=124.649*, *p<.001*; MOOC 2: *H=175.534*, *p<.001*), the completion rate (R_c) (MOOC 1: *H=60.691*, *p<.001*; MOOC 2: *H=107.799*, *p<.001*), and the attempt rate (R_a) (MOOC 1: *H=96.746*, *p<.001*; MOOC 2: *H=125.44*, *p<.001*), for all 7 age groups.

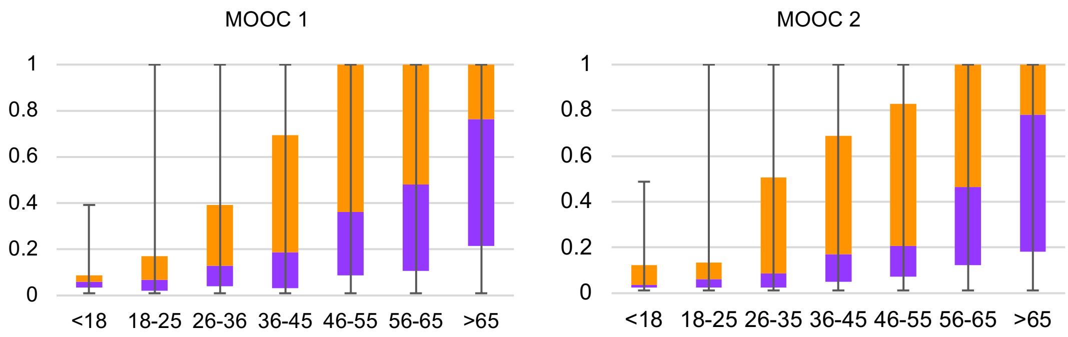

Fig. 6. Visit rate ($\boldsymbol{R_v}$) for different age groups in the two MOOCs.

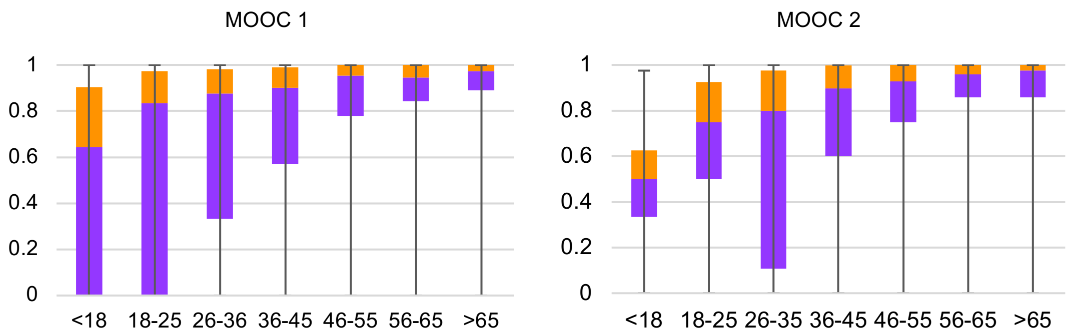

Fig. 7. Completion rate ($\boldsymbol{R_c}$) for different age groups in the two MOOCs.

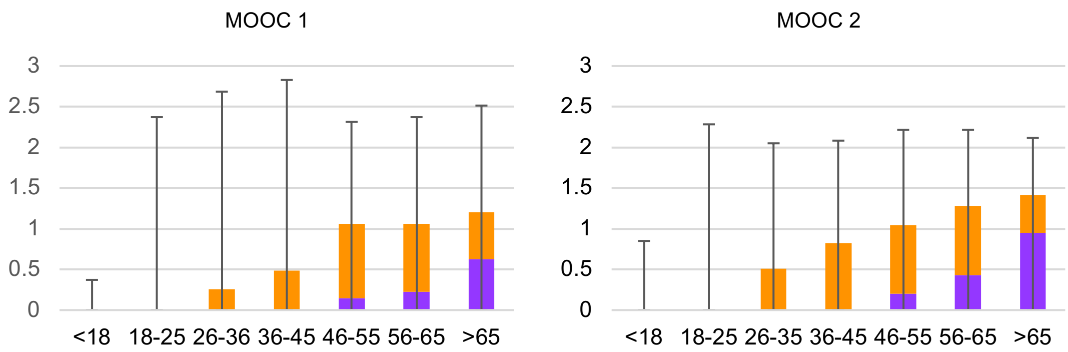

Fig. 8. Attempt rate ($\boldsymbol{R_a}$) for different age groups in the two MOOCs.

Fig. 9 shows the comparison of the mean numbers of comments (discussions) for different age groups in MOOC 1 and MOOC 2. Overall, in general, for both MOOC 1 and MOOC 2, the older the learners were, the more comments (discussions) they contributed. Additionally, the Kruskal-Wallis test result suggested that the difference between different groups were statistically significant, as per, MOOC 1: $H=86.489$, $p<.001$; MOOC 2: $H=140.817$, $p<.001$.

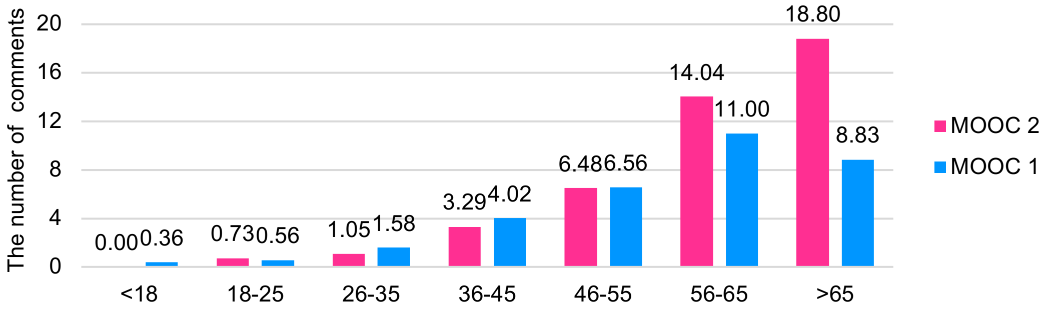

Fig. 9. Mean number of comments for different age groups in the two MOOCs.

5. Conclusions and Discussions

To conclude, this study has analysed and reported learner data collected from six runs of two MOOCs delivered by the University of Warwick. Whilst two courses may be too few to conclude that gender plays an important role, these two courses have been analysed over several runs and have reasonably large sample size. Analyses in section 4 show thus that both *gender* and *age group* indicators may have very strong influence on following a MOOC, visiting *step* pages, completing *steps*, attempting to answer questions, and writing comments (discussions). The results suggest that ***learners' demographic indicators may strongly influence their learning activities in MOOCs***.

Given the fact that MOOC learners originate from all around the world, with very different backgrounds and characteristics, when designing MOOCs, there is clearly a strong need for providing *personalised learning support when developing educational information systems and instructional design*. This means not only recommending learning content for learners to learn, based on their prior learning experience or knowledge, as some MOOC platforms can do, but also personalising the way they learn, such as adapting the learning path and supporting adaptive interventions.

Finding out, for example, even for a specific course, that a certain age group is more likely to complete the course than another, opens up possibilities for support offered for a new run of the same course, to the age group that is less likely to continue. They can be offered a version that runs at a different pace, or slightly streamlined materials, if it is a matter of time available, etc. Importantly, these findings allow *very early intervention*, starting immediately after registration, as these *demographic indicators* are known often even before the MOOC starts, as many learners register early. Thus, real-time (or close to real-time) interventions can be developed. FutureLearn tutors tend to have at least weekly wrap-up sessions which are recorded during the course run, as well as tutor assistants that monitor and answer questions – both of these methods can be used to specifically address learners that may struggle later on.

MOOCs are widespread, but in order to increase their success, the challenge remains to add the capability of adapting to learners' individual *demographic indicators*, such as gender and age, in order to suggest the most beneficial learning activities for every learner, at every moment during the learning. Current MOOCs often lack personalisation support. Still, most MOOCs break down learning materials into smaller units, which gives the chance to break away from the "one-size-fits-all" education. However, presently, this heavily relies on learners' effort to self-direct and self-determine their learning process, which is clearly not functioning well [1]. Therefore, there is a clear and strong need to understand *how learner demographic indicators may influence activities and learning experience in MOOCs, and,* *more importantly, to develop effective pedagogical strategies and information systems to support meaningful adaptation and interventions*.

This study used data from two MOOCs: one was "Big Data: Measuring and Predicting Human Behaviour" – a STEM MOOC (science/engineering); the other was "The Mind is Flat" – a non-STEM MOOC (social science/psychology), thus covering different disciplines. Nevertheless, the influence of learning demographic indicators on learning activities in terms of dependence on the MOOC discipline needs further investigation. Therefore, our future work will include investigating the dimension of the MOOC discipline.

It is noteworthy mentioning that, as clicking on the button "Mark as complete" on a *step* page is a self-claim, it is still unclear to which extent this represents 'real' completion of the *step*. This is specifically interesting in the context of the proportion of claiming learnt *steps* within all pages visited being very high. This is very similar to the observation from [24], yet the implications need further interpretation.

Another dimension to be considered in our future work includes the *time sequence*, e.g., a chronologically ordered set of learner activities. This may potentially help gain deeper insight into learning activities in MOOCs, thus allowing to efficiently cluster learners and provide real-time adaptation and personalisation, based on *learning patterns*.

References

1. Anders, A.: Theories and Applications of Massive Online Open Courses (MOOCs) : The Case for Hybrid Design. Int. Rev. Res. Open Distrib. Learn. 16 (6), (2015)
2. Barba, P.G. de, Kennedy, G.E., Ainley, M.D.: The Role of Students' Motivation and Participation in Predicting Performance in a MOOC. J. Comput. Assist. Learn. 32 (3), 218–231
3. Bote-Lorenzo, M.L., Gómez-Sánchez, E.: Predicting the Decrease of Engagement Indicators in a MOOC. (2017)
4. Cristea, A.I., Alamri, A., Kayama, M., Stewart, C., Alshehri, M., Lei, S.: How is Learning Fluctuating? FutureLearn MOOCs Fine-grained Temporal Analysis and Feedback to Teachers and Designers. In: 27th International Conference on Information Systems Development (ISD2018). Association for Information Systems, Lund, Sweden (2018)
5. Cristea, A.I., Alamri, A., Stewart, C., Alshehri, M., Shi, L., Shi, L.: Earliest Predictor of Dropout in MOOCs: A Longitudinal Study of FutureLearn Courses. In: 27th International Conference on Information Systems Development (ISD2018). Association for Information Systems, Lund, Sweden (2018)
6. Ferguson, R.: Learning Analytics: Drivers, Developments and Challenges. Int. J. Technol. Enhanc. Learn. 4 (5/6), 304 (2012)
7. Ferguson, R., Clow, D.: Examining Engagement: Analysing Learner Subpopulations in Massive Open Online Courses (MOOCs). (2015)
8. Ferguson, R., Sharples, M.: Innovative Pedagogy at Massive Scale: Teaching and Learning in MOOCs. In: Rensing, C., de Freitas, S., Ley, T., and Muñoz-Merino, P.J. (eds.) Open Learning and Teaching in Educational Communities. pp. 98–111. Springer International Publishing (2014)
9. FutureLearn: About FutureLearn, *FutureLearn*, https://www.futurelearn.com/about-futurelearn, Accessed: June 17, 2018
10. Hone, K.S., El Said, G.R.: Exploring the factors affecting MOOC retention: A survey study. Comput. Educ. 98 157–168 (2016)
11. Kahan, T., Soffer, T., Nachmias, R.: Types of Participant Behavior in a Massive Open Online Course. Int. Rev. Res. Open Distrib. Learn. 18 (6), (2017)
12. Khalil, M., Ebner, M.: Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories. J. Comput. High. Educ. 29 (1), 114–132 (2017)

13. Laurillard, D.: Rethinking University Teaching: a Conversational Framework for the Effective Use of Learning Technologies. RoutledgeFalmer, London ; New York (2002)
14. Mining, T.E.D.: Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief. In: Proceedings of conference on advanced technology for education. (2012)
15. Morris, N.P., Hotchkiss, S., Swinnerton, B.: Can Demographic Information Predict MOOC Learner Outcomes. Proc. EMOOC Stakehold. Summit. 199–207 (2015)
16. Pask, G.: Conversation Theory: Applications in Education and Epistemology. Elsevier, Amsterdam ; New York (1976)
17. Rebecca Yvonne Bayeck: Exploratory Study of MOOC Learners' Demographics and Motivation: The Case of Students Involved in Groups. Open Prax. 8 (3), 223–233 (2016)
18. Ren, Z., Rangwala, H., Johri, A.: Predicting Performance on MOOC Assessments using Multi-Regression Models. ArXiv160502269 Cs. (2016)
19. Romero, C., Ventura, S.: Educational Data Science in Massive Open Online Courses: Educational Data Science in Massive Open Online Courses. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 7 (1), e1187 (2017)
20. Ryan, R.M., Deci, E.L.: Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-being. Am. Psychol. 55 (1), 68–78 (2000)
21. Sclater, N., Peasgood, A., Mullan, J.: Learning Analytics in Higher Education. Lond. Jisc Accessed Febr. 8 2017 (2016)
22. Shi, L., Cristea, A.I.: Making It Game-like: Topolor 2 and Gamified Social E-Learning. In: The 22nd Conference on User Modeling, Adaptation and Personalization (UMAP 2014). pp. 61–64. , Aalborg, Denmark (2014)
23. Shi, L., Cristea, A.I.: Motivational Gamification Strategies Rooted in Self-Determination Theory for Social Adaptive E-Learning. In: Micarelli, A., Stamper, J., and Panourgia, K. (eds.) Intelligent Tutoring Systems. pp. 294–300. Springer International Publishing, Cham (2016)
24. Shi, L., Cristea, A.I., Awan, M.S., Stewart, C., Hendrix, M.: Towards Understanding Learning Behavior Patterns in Social Adaptive Personalized E-learning Systems. In: The 19th Americas Conference on Information Systems. pp. 1–10. Association for Information Systems, Chicago, Illinois, USA (2013)
25. Shi, L., Cristea, A.I., Hadzidedic, S.: Multifaceted Open Social Learner Modelling. In: Popescu, E., Lau, R.W.H., Pata, K., Leung, H., and Laanpere, M. (eds.) Advances in Web-Based Learning – ICWL 2014. pp. 32–42. Springer International Publishing, Cham (2014)
26. Shi, L., Cristea, A.I., Hadzidedic, S., Dervishalidovic, N.: Contextual Gamification of Social Interaction – Towards Increasing Motivation in Social E-learning. In: Popescu, E., Lau, R.W.H., Pata, K., Leung, H., and Laanpere, M. (eds.) Advances in Web-Based Learning – ICWL 2014. pp. 116–122. Springer International Publishing, Cham (2014)
27. Slade, S., Prinsloo, P.: Learning Analytics: Ethical Issues and Dilemmas. Am. Behav. Sci. 57 (10), 1510–1529 (2013)
28. Yuan, L., Powell, S., CETIS, J., others: MOOCs and Open Education: Implications for Higher Education. (2013)
29. 1st International Conference on Learning Analytics and Knowledge 2011 | Connecting the technical, pedagogical, and social dimensions of learning analytics, https://tekri.athabascau.ca/analytics/, Accessed: June 19, 2018
30. A Product at Every Price: A Review of MOOC Stats and Trends in 2017 - EdSurge News, https://www.edsurge.com/news/2018-01-22-a-product-at-every-price-a-review-of-mooc-stats-and-trends-in-2017, Accessed: January 28, 2018, (2018)
31. MoocLab Report: The Global MOOC Landscape - 2017, *MoocLab - Connecting People to Online Learning*, https://www.mooclab.club/resources/mooclab-report-the-global-mooc-landscape-2017.214/, Accessed: June 18, 2018