How is learning fluctuating? FutureLearn MOOCs fine-grained temporal Analysis and Feedback to Teachers and Designers

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

Mohammad Alshehri*   *mohammad.alshehri@durham.ac.uk*

Ahmed Alamri*   *ahmed.alamri@durham.ac.uk*

Mizue Kayama   *kayama@cs.shinshu-u.ac.jp*
Shinshu University, Shinshu, Japan

Craig Stewart   *craig.stewart@coventry.ac.uk*
Coventry University, Coventry, UK

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

Abstract

Data-intensive analysis of massive open online courses (MOOCs) is popular. Researchers have been proposing various parameters conducive to analysis and prediction of student behaviour and outcomes in MOOCs, as well as different methods to analyse and use these parameters, ranging from statistics, to NLP, to ML, and even graph analysis. In this paper, we focus on patterns to be extracted, and apply systematic data analysis methods in one of the few genuinely large-scale data collection of 5 MOOCs, spread over 21 runs, on FutureLearn, a UK-based MOOCs provider, that, whilst offering a broad range of courses from many universities, NGOs and other institutions, has been less evaluated, in comparison to, e.g., its American counterparts. We analyse *temporal quiz solving patterns*; specifically, the less explored issue on how the first number of weeks of data predicts activities in the last weeks; we also address the classical MOOC question on the completion chance. Finally, we discuss the *type of feedback a teacher or designer could receive* on their MOOCs, in terms of fine-grained analysis of their material, and what personalisation could be provided to a student.

Keywords: FutureLearn, MOOC, statistics, patterns, feedback.

1. Introduction

Online courses have been around for decades, yet often catered to a limited audience only. To address this scalability issue, massive open online courses (MOOCs) were developed. Tracing back to MIT's 2001OpenCourseWare initiative, MOOCs entered the modern age of successful commercialisation with Stanford's Coursera in 2011 [16]. MOOCs have become increasingly popular and their scale and availability make it possible to offer a diverse set of students from all over the world online courses. Thus, many MOOC providers, such as edX, Udacity, Coursera, FutureLearn and XuetangX, have started offering scalable online courses to the public. By the end of 2017, the number of MOOC providers has reached a total of 57, and the total number of MOOC students has become more than a hundred million[1].

Notwithstanding the unparalleled success of MOOCs, especially in terms of the thriving student enrolment, one of the more concerning aspects to date is the staggeringly low

---

[1] https://www.mooclab.club/resources/mooclab-report-the-global-mooc-landscape-2017.214/

participation and completion rate – a funnel with students "leaking out" at various points along the way of learning [4, 6]. Despite various studies conducted to investigate the links between behaviours and the completion [5, 18, 11], the race for finding predictors of completion, and, more importantly, early predictors, continues.

In this study, we take advantage of the fine-grain resolution of the clickstream data – single actions, such as visiting a page and marking it as completed, or answering a quiz, associated with student ID (the unique and anonymous ID signed to a student) and timestamps – to depict student behavioural patterns over entire courses, and how this may affect their future behaviours and the chance of completion of the courses. In particular, this paper presents the results of analysing a unique large dataset of FutureLearn MOOC students over several runs, investigating a large amount of user behaviour and completion, to extract early factors that can predict user behaviour and completion in the later part of a course. Specifically, the paper focuses on addressing, *at a large scale, or a large variety of courses on different topics*, the following set of umbrella research questions, around students' learning outcomes and behavioural patterns:

RQ1: *Is the behaviour in the first weeks influencing the behaviour of students in the last weeks of the course, regardless of course structure or topic?*

RQ2: *Is the behaviour in the first weeks influencing the completion chance of a student, regardless of course structure or topic?*

2.   Related Research

The area of analysing 'big data' and predicting relationships based on it is one of the hottest topics of web-related research [3, 1], encompassing statistics, machine learning, natural language processing, a.o.. Most researchers in this area have been focusing on social data analysis [12, 21], although other fields have also thriving communities (e.g., medicine [17], a.o.). Whilst, traditionally, educational research does not involve such numbers, with the advent of the MOOCs [2], the interest in analysing 'big data' in education increased, spouting the emerging fields of learner analytics and educational data mining. Learning analytics collects and analyses data about learners and their contexts, to understand and optimise learning and its environments [10], often providing a visual output for learners, educators, designers or administrators. Educational data mining, instead, applies computerised methods, such as machine learning and data mining, to the enormous volume of educational data [14].

Recently, work based on statistics, machine learning (ML) and visualisation has focussed on analyses and predictions directly related to our current paper, as below. Lu et al. [15] extract a large number of features (19) to predict dropout, based on ML methods and support vector machines (SVM), from 5 courses (similarly to us, although they only analyse 1 run each) on Coursera. Qiu et al. [19] extract factors of engagement on XuetangX (China, partner of edX), on 11 courses, predicting grades, certificate earning with different methods (LRC, SVM, FM, LadFG); their performance is evaluated with area under the curve (AUC), precision, recall, and F1 score. Features used include demographics (gender, age, education, etc.), forum (number of replies, etc.), learning behaviour (chapters browsed, deadlines completed, time spent on videos, doing assignments, etc.); they interestingly use also temporal correlations between their factors for the predictions. We are looking also at engagement in a large-scale study, but we ***analyse completion and behaviour***, not grades and certificates. Gardner and Brooks [9] discuss correct ways of applying predictive models in MOOCs; they evaluate models of student dropout; this is a longitudinal study, like ours, looking at 31 runs of 5 courses on Coursera, with 298,909 learners; they conclude that models utilising clickstream features consistently outperform those using forum and quiz features. Whereas here, we ***focus only on quizzes*** for our analysis. Robinson et al. [20] use NLP to predict dropout on only 1 HarvardX course; language features are selected via the lasso logistic regression model; performance is evaluated with AUC. Our current study explores a *different parameter* (i.e., quizzes), *in terms only of their quantity*, for *different categories of users*, and, importantly, for ***different study weeks***. Crossley et al. [8] do, in comparison to our current study, a much smaller-scale study (320 students, from only 1 course on Coursera), predicting completion (defined as an overall grade of >=70%); they do

cohesion analysis with the TASA corpus on the forum discussions; they compare completers and non-completers based on a wide set of parameters; evaluate performance with recall, precision, and the F1 score; they show that collaborating students have a higher chance to complete. These are all interesting areas to explore, on a much wider scale, for future research. In [7] we study 6 FutureLearn courses with 23 runs overall, and find statistically relevant periods for registration of students, which can predict the likelihood of course completion. Here, we are analysing different parameters, to *predict not only completion, but also **behaviour of students***. In [13] we investigate 2 FutureLearn courses with 6 runs in total, noticing that gender and education may influence students' behaviours, in terms of comments posted, questions attempted and steps (pages) completed. The latter study is our inspiration for the RQs below, as, beside looking only at a limited number of courses, [13] did not further research *how **behaviour influences behaviour***, or explore ***connections of behaviour to completion***, as is done in our current paper.

3. Methodology

3.1. Study Setting

MOOCs in FutureLearn are built upon *weekly learning units*[2]. Each of these units contains a number of *learning blocks*. These blocks can have one or more *steps* – the basic learning items. The latter can be articles, images, videos, which can include quizzes. Students can view (access) these steps and mark them as *completed*. Students can also add *comments* for each of these steps. Additionally, specifically for quizzes, students can have several *attempts* at each quiz containing several questions, till they arrive at the *correct answer*.

3.2. Data

All activities of the students are logged with their learner ID (of the student performing that activity) and a timestamp. We are analysing data from 5 MOOCs (all courses delivered via the FutureLearn platform by the University of Warwick from the start of their activity on FutureLearn (2013-2017), which have quizzes. These courses are of various subjects, ranging from literature to computer science to social sciences, as follows: 'Babies in mind', 'Big Data', 'Shakespeare and His World', 'Supply chains', 'The Mind is flat'. Each of these courses was delivered repeatedly, in consecutive years. These repeated deliveries are called *runs*, and we are analysing 21 runs in total. Overall, we analyse thus the activity of 218,795 learners, who accessed 3,007,789 materials, declared completed 2,794,578 steps, attempted 2,406,574 quiz questions, out of which 1,601,665 answers were correct.

3.3. Research Question Interpretation

Our umbrella research questions are intentionally kept broad, to cover the overall purpose of this research. For this paper, we interpret the *first weeks* or a course as the first half of a course (this means different things for different courses; e.g., it comprises weeks 1-2 for a 4-week course; weeks 1-4.5 for a 9-week course, etc.). Thus, the *last weeks* of the course represent here the other half of the course, although our main concern is the prediction of the last week, as well as the completion of the course.

Behaviour is referring to questions answered, where we include the total number of attempts, wrong attempts as well as correct answers.

Completion is analysed via 3 scenarios:

- Scenario 1: 100% of the steps are completed; this means that students actually press on the 'completed' button;
- Scenario 2: 100% of the steps are accessed;

---

[2] FutureLearn terminology is highlighted in *italics* in this section.

- Scenario 3: 80% of the steps are accessed.

Thus, based on the definitions above, the research questions can be rewritten as follows:

RQ1a/b: *Is the number of questions answered correctly in the **first half** of a MOOC influencing the number of questions answered correctly(a)/incorrectly(b) in the **last half**?*

RQ2.1a/b/c: *Is the number of questions answered correctly in the **first half** of a MOOC influencing the chance of a student to achieve 100% **completed** steps (a)/ 100% **accessed** steps (b)/ 80% accessed steps (c)?*

RQ2.2a/b/c: *Is the number of questions answered incorrectly in the first half of a MOOC influencing the chance of a student to achieve 100% **completed** steps (a)/ 100% **accessed** steps (b) / 80% accessed steps (c)?*

Due to the different nature of the MOOCs (e.g., some taking 4 weeks, others 9 weeks; some having quizzes every week, others skipping some of the weeks, quizzes being of different nature, difficulty and subject) across the 5 courses under investigation, we have considered it best to analyse the different courses in parallel, merging only the data from the different runs of each course. This way, we could both draw conclusions for all courses, as well as find specific characteristics for each course, importantly, allowing for predictions for future runs. Moreover, the methodology, as applied to the various subjects and diversity of MOOCs, is generic and can be applied elsewhere as well.

3.4.   Analysis

To address the research questions raised in section 1, we compute the mean value (μ) for the behaviour variables selected above (per run, per course, per weeks considered, per learner), their variance (σ^2), as well as for the number of people in the different subgroups identified. To establish if the data is normally distributed, we use the Pearson chi-squared test (establishing 'goodness of fit'). Depending on this, we then use a T-test or ANOVA for normally distributed data, or the Wilcoxon signed-ranked or the Kruskal-Wallis tests otherwise. The Pearson product-moment correlation coefficient is used for establishing relations between variables. The Bonferroni correction is used for compensation of multiple comparisons. To avoid bias in the results, students with no quizzes answered at all (neither correct nor incorrect) are removed. These are students who either only register and never access the course, which are dealt with elsewhere [2], or students who may have accessed the course, but have never answered any questions (56,289 or 26%).

4.   Results

This paper presents an analysis of quiz data from 5 courses taught across 21 runs on the FutureLearn platform. The analyses focus on rates of quiz questions attempted, correctly answered, and incorrectly answered, and how these rates change over the course of a term, for both students who complete and do not complete the course. Specifically, in the following, we show how we respond to the two research questions, *RQ1* and *RQ2*, at a per course-base, first, and then generally. None of our data is normally distributed, and our categorical data consisted only of two categories (completers and non-completers), so we use the Wilcoxon signed-ranked test throughout. The probabilities for significance are so low that the Bonferroni correction does not make any noteworthy changes to the final results (i.e., for $\forall p < 0.05$, also $p<0.05*(1/n)$ holds, where *n* is the number of overall tests performed).

To investigate *RQ1a/b*, we analyse correct and wrong answers for all weeks first. Subsequently, for *RQ2 and its sub-questions*, we hone in on the early weeks, when comparing to the completion chance. Fig. 1 displays the overall number of correct and wrong answers for each week in which a test is available (i.e., weeks W1 and W5 have no test and are not represented), for the 'Big Data' course, a Computer Science course, averaged for all its runs (here, 2 runs). We can see that the number of answers, either correct or wrong, of the students who don't complete, decreases gradually. Based on such an analysis, we decided to estimate the possibility to predict the completion, based on the correct or wrong answers in the early

weeks. For Week 2 correct answers, the mean of completing students is 4.92 (σ^2=0.38), just slightly higher than the mean of correct answers of non-completing students (μ=4.75; σ^2=1.3), and this difference is statically significant (Wilcoxon signed-rank test: p=0.0038); however, Week 2's wrong answers are not significantly different.
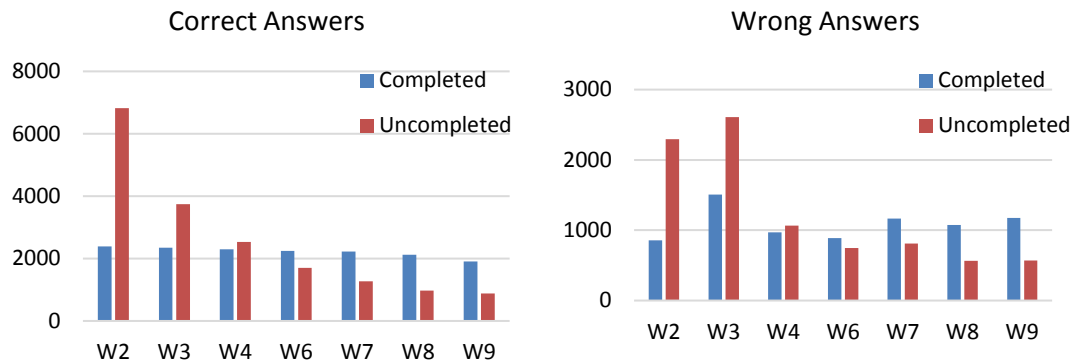

Fig. 1. Big Data (Computer Science): Correct versus Wrong answer evolution, per week (W*i*; *i*□*2,3,4,6,7,8,9}*); students completed (blue), or didn't (red), as per Scenario 3 (80% accessed).

This situation clearly improves, starting with Week 3 (see Table 1), from which we have statistical significance throughout. Fig. 2 and Fig. 3 show the boxplots of weeks 3 and 4, respectively, for correct and wrong answers (default whisker in R of 1.5 IRQ is used); one can see how these significant differences also increase in size, as well as how the groups are slowly more defined. Due to lack of space, we haven't represented boxplots for the remaining weeks, but the trend of significance and larger difference continues.

Table 1. Big Data, all runs: average correct and wrong numbers of questions per week, versus the completion status (Yes/No).

Week	Answers	Completed?	Mean (μ)	σ^2	p (Wilcoxon)
2	Correct	Yes	4.92	0.39	0.003803
2	Correct	No	4.75	1.13	
2	Wrong	Yes	1.76	5.89	0.8782
2	Wrong	No	1.60	4.23	
3	Correct	Yes	4.83	0.96	2.2e^{-16}
3	Correct	No	2.61	6.20	
3	Wrong	Yes	3.10	9.76	2.2e^{-16}
3	Wrong	No	1.82	7.48	
4	Correct	Yes	4.72	1.30	2.2e^{-16}
4	Correct	No	1.76	5.70	
4	Wrong	Yes	1.99	9.05	2.2e^{-16}
4	Wrong	No	0.74	3.96	

To analyse RQ1 for the 'Big Data' course, we studied the answers in early weeks, compared to the final ones. Week 2 is the earliest possible predictor for Week 9 (see Fig. 4). Interestingly, correct answers in the weeks compared are statistically significantly correlated for completers (Scenario 1; Pearson's: W2-W9: 0.3, p = 1.40e-10; W3-W9: 0.54, p < 2.2e-16; W4-W9: 0.69, p < 2.2e-16; W6- W9: 0.75, p < 2.2e-16; W7-W9: 0.86, p < 2.2e-16; W8-W9: 0.9, p < 2.2e-16). Thus, whilst predictions could be made starting Week 2, the precision is expected to increase in later weeks. A similar analysis for non-completers shows that only starting with Week 3 the correlations become significant, and thus likely candidates for prediction. Fig. 4 further indicates classes that could be identified for Week 2; for instance, the red dots on the x-axis of the left image represent students who have not given any correct answer in one or more weeks; similar red dots on the x-axis of the right image represent students without wrong attempts – possibly due to them making no attempts that particular week.
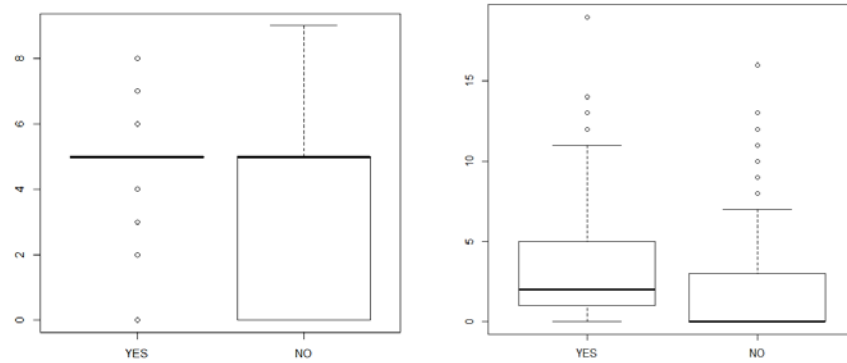
Fig. 2. Big Data: left to right: Week 3 correct answers; Week 3 wrong answers; (YES denotes completed, as per Scenario 3 (80% accessed); No denotes not completed).
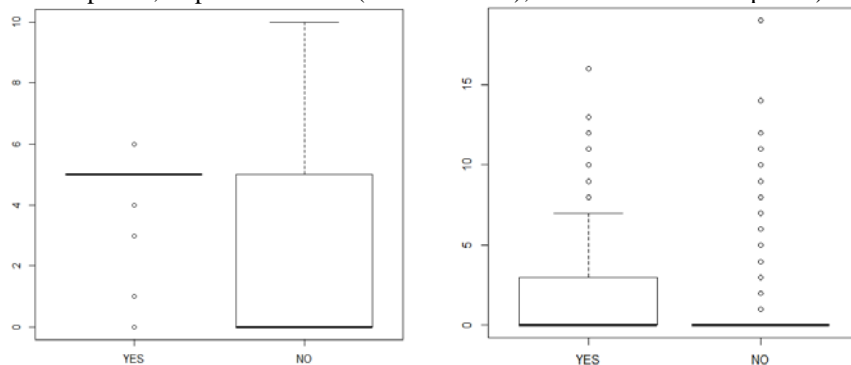

Fig. 3. Big Data: left to right: Week 4 correct answers; Week 4 wrong answers; (YES means completed, as per Scenario 3 (80% accessed); NO means not completed).
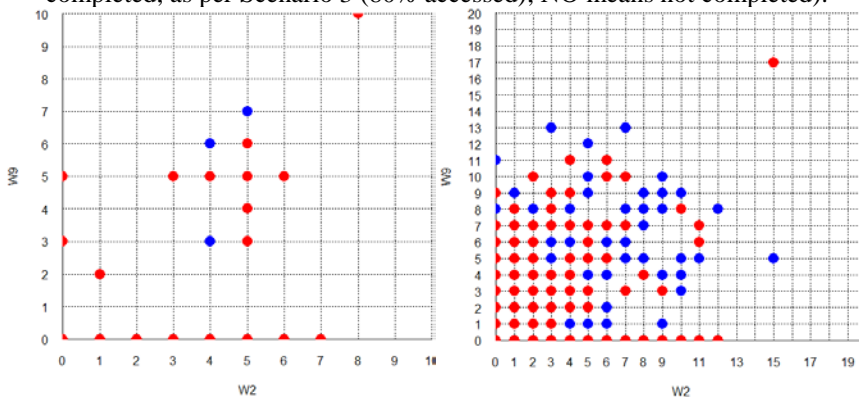

Fig. 4. Big Data: left (correct answers) to right (wrong answers), distribution of completers (blue) and non-completers (red) in Week 2 versus Week 9 (as per Scenario 3 (80% accessed).
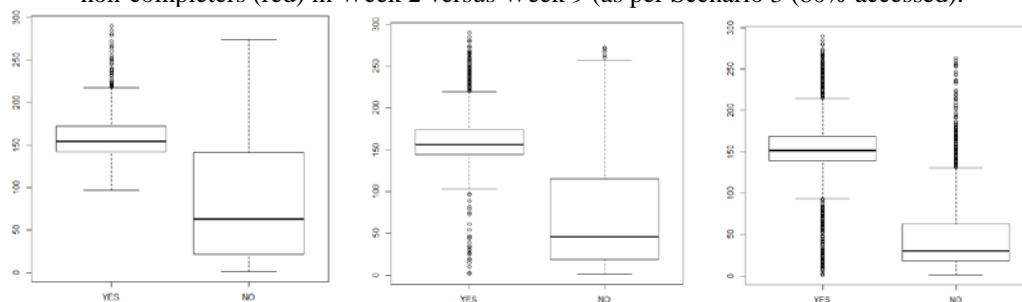

Fig. 5. Shakespeare and His World (Literature): illustrating the 3 scenarios for attempted quizzes: left to right, Scenario 1: completed ('YES') are only students who have clicked that they 'learned' all (100%) of the 'steps' (material): 1,117 students; Scenario 2: completed are students who accessed 100% of the steps: 3,678 students; Scenario 3: completed are students who accessed 80% of the steps: 7,137 students.

We performed similar analyses on all 5 courses, for all 3 scenarios. We illustrate this with a course at the other end of the spectrum, 'Shakespeare and His World' (Literature, 5 Runs). Fig. 5 shows how the 3 scenarios affect the classification of students as completers or not. Beside the difference in numbers of completers, the median for these is relatively constant for

the 3 scenarios, as is the box-size. What varies more is the median for non-completers, and the number of outliers. Due to the lack of space, we don't repeat this for the other 3 courses.

Fig. 6 shows the evolution for correct and wrong answers over the 10 weeks of the course, partitioned between students who completed (blue) and those who didn't (red). It can be seen that the number of answers is decreasing for those students who will not complete – regardless if they are correct or wrong answers. The number of correct answers for completers (blue, left side), on the other hand, remains relatively constant, whilst the numbers of wrong answers for completers (blue, right side) has more fluctuations, depending on the week. The figure shows also a marked similarity with Fig. 1, and this is consistent with our analyses of the rest of the courses (not displayed here, due to lack of space). This pattern is similar for the other three courses analysed – see Fig. 7. The figure also shows that the different courses had different number of weeks, as well as not all weeks did provide test and quizzes.
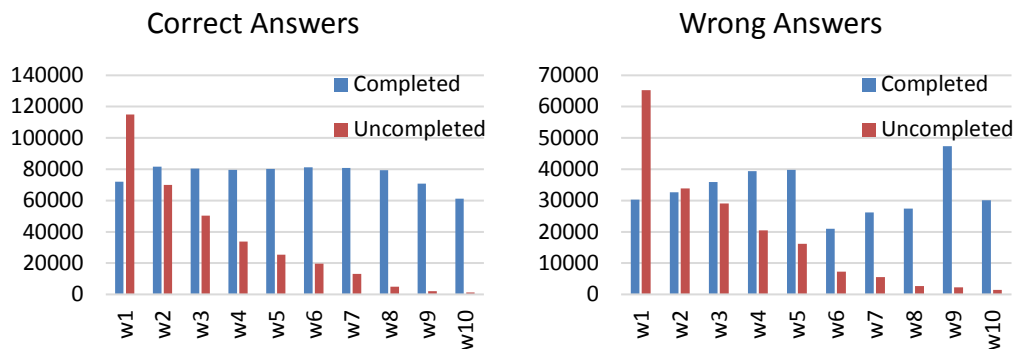

Fig. 6. Shakespeare and His World (Literature): Correct versus Wrong answer evolution, per week (Wi; i∈{1-10}); students completed (blue), or didn't (red), as per Scenario 3 (80% accessed).
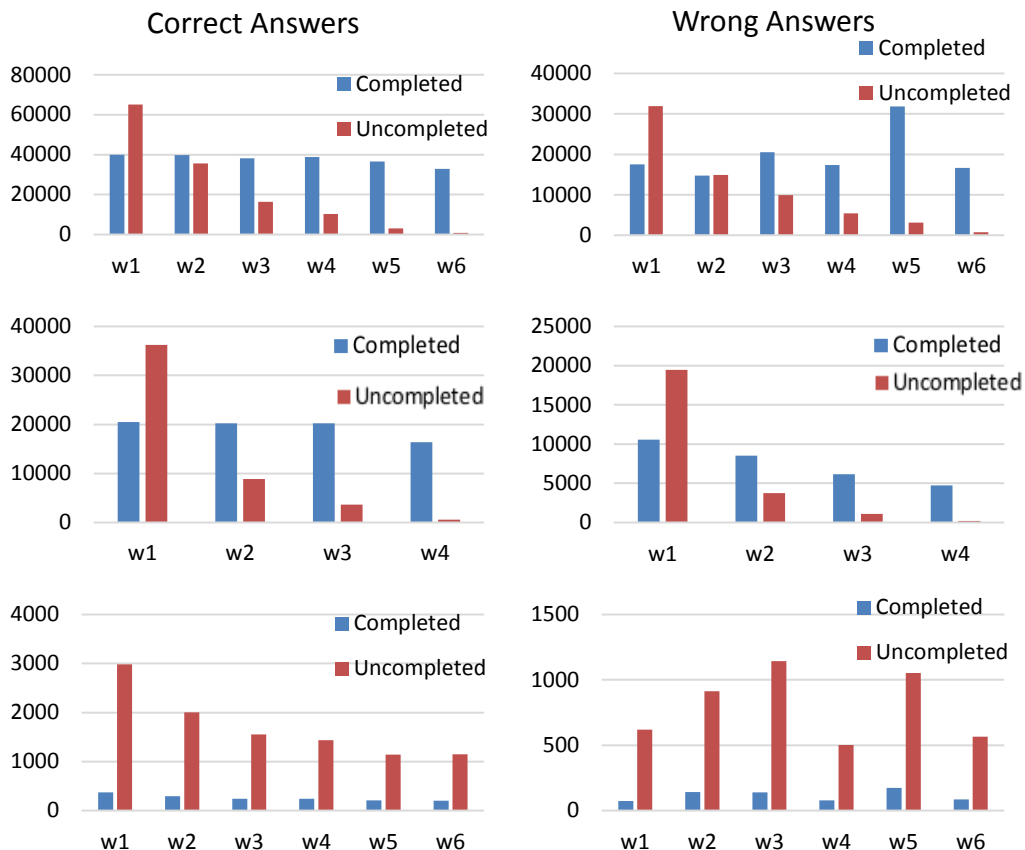

Fig. 7. The Mind is flat (Psychology) above, Babies in mind (Psychology2) middle, Supply chains (Business) below: Correct versus Wrong answer evolution, per week (Wi; i∈{1-10}); students completed (blue), or didn't (red), as per Scenario 3 (80% accessed).

Further analysing in details the 'Shakespeare and His World' course, to better visualise the weekly evolution of the distribution of correct and wrong answers for completers and non-completers, Fig. 8 shows that completers normally tend to answer all questions correctly for each week (see upper left side image), whereas non-completers clearly have a lower median for each week, as well as a greater variance for correct answers. For wrong attempts, completers have been very busy in the first weeks with very many wrong attempts, but slowly converge towards almost no wrong attempts. Non-completers follow a similar pattern, although they have fewer attempts in general, and 'give up' at an earlier stage (lower right side, Week 4). Unlike for the 'Big Data' course, the means for completers versus non-completers for correct and wrong answers for all weeks (W1-W10) are significantly different (Wilcoxon: p < 2.2e-16).
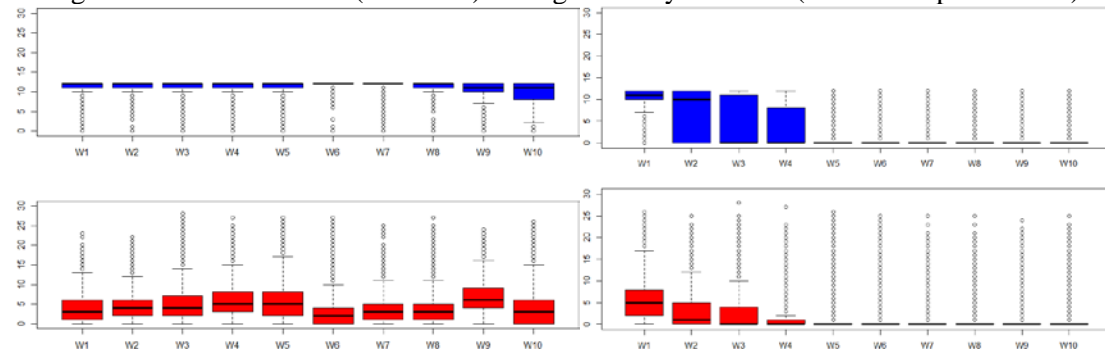

Fig. 8. Shakespeare and His World (Literature): left side: Correct versus right side: Wrong answer evolution, per week; blue are the students who complete, red who don't, as per Scenario 3 (80% accessed).

For behaviour prediction, as for the 'Big Data' course, we analyse the earliest quiz (here, in Week 1) versus the latest (here, in Week 10), visualised as scatterplot in Fig. 9. We also test correlation for correct versus wrong answers. The correlation becomes statistically significant (Pearson's: p < 2.2e-16) starting Week 2 (which is here the second week of tests). Comparable results (Pearson's: p < 2.2e-16, starting Week 2 or Week 1 for correlation) are found for the remaining 3 courses and 14 runs, clearly indicating behavioural prediction opportunities, in terms of number of quizzes solved correctly or incorrectly in the early weeks, which allow for the prediction of quizzes solved correctly or incorrectly in the later weeks.
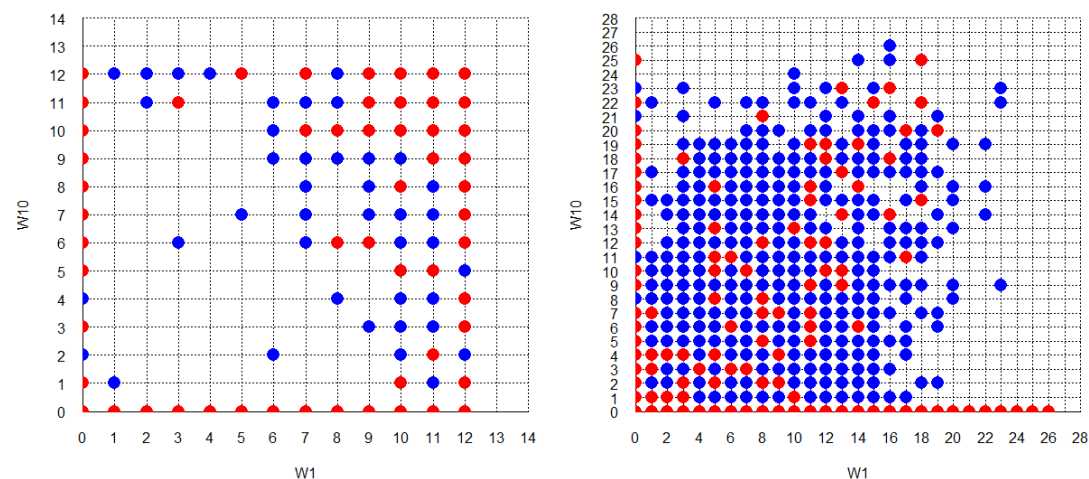

Fig. 9. Shakespeare and His World: left to right, distribution of completers in Week 1 versus Week 10 (blue dot means completed, as per Scenario 3 (80% accessed); red dots denote not completed).

Additionally, we have found statistically significant differences (based on Wilcoxon, as data distribution is non-normal) for completers versus non-completers in the general number of

attempts at quizzes, as well as number of quizzes for the different weeks – with interesting results in the early weeks (1, 2). The Bonferroni correction did not change the results, despite the multiple comparisons. When analysing the 5 courses together, specifically, first, second and third week versus the last week in terms of behaviour, as well as completion, we obtain surfaces as summarised in Fig. 10, for the three scenarios, showing the sparse nature of correctly answered questions, which are mostly either fully answered, or not at all (very few positive peaks – for correct answers; and negative peaks – for incorrect ones); e.g., the peek pointing downwards, at (0,0), is where the majority of students didn't answer questions either in week (1,2,3) or in the last week, and did not complete; the small positive peak (5,5) shows students who answered all questions in the early as well as late weeks and completed.

Table 2. P (Wilcoxon) results for behaviour (correct answers) on the first, second and third weeks, respectively, versus the last week in the three scenarios: 100% completed (Scenario 1), 100 Accessed (Scenario 2) and 80% accessed (Scenario 3).

Scenario (all courses)	Scenario 1	Scenario 2	Scenario 3
Week comparisons:	p-value		
First week versus last week	0.02524	0.004072	2.20E-16
Second week versus last week	0.02349	2.20E-16	2.20E-16
Third week versus last week	0.02145	0.0001778	2.20E-16

completion rate of 100% (Scenario 1).

access rate of 100% (Scenario 2).

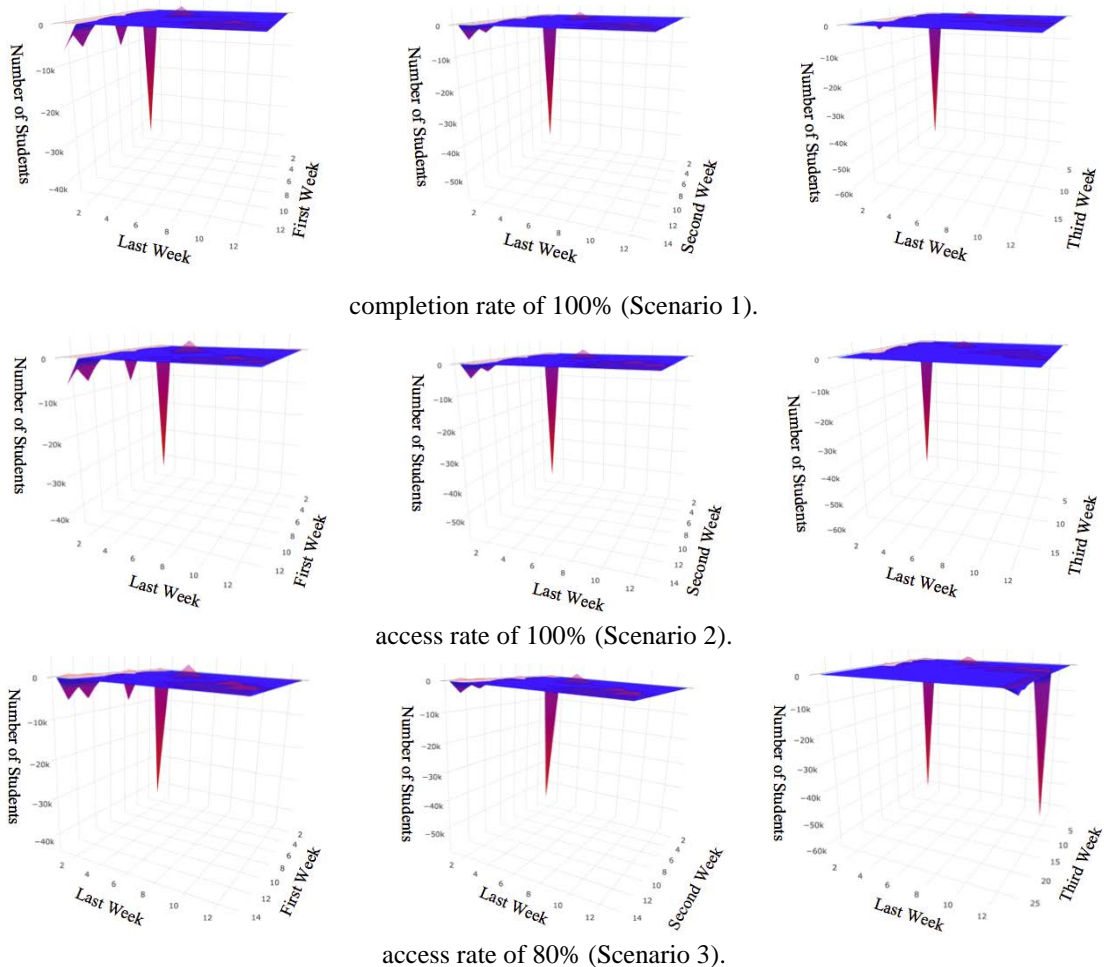access rate of 80% (Scenario 3).

Fig. 10. All 5 courses together: number of students who answered questions correctly (above)/ incorrectly (below) in the first, second or third week versus last week for the three scenarios; blue: completion; red: non-completion.

5.   Discussion

Overall, we have answered RQ1, by showing that completers have similar behaviour with respect to quiz answering in early weeks, as they have in later weeks; this statement is valid for correct answers (RQ1a), as well as for incorrect attempts (RQ1b). Similarly, we have answered RQ2: we can use behaviour variables, such as questions answered, to visualise the potential partitioning of completers versus non-completers. Importantly, we have found statistical significance in the difference in means between completers and non-completers, for the number of attempts in general, the correctly answered quiz questions (RQ2.1a/b/c), and, finally, incorrectly answered ones (RQ2.2a/b/c). Analysing the findings more in-depth, we can also remark the following. Fig. 1 shows that students who keep working and answering questions are more likely to complete the 'Big Data' course, regardless of the accuracy of their answers (follow completers (blue) for both correct answers (left) and wrong answers (right)). This observation is in line with prior work and is confirmed by all analyses of our 5 courses and 21 runs (see, e.g., Fig. 6 and Fig. 7, where completers have the majority of wrong answers for the latter half of the course). Table 1 shows the prediction becoming simpler, the closer the test results are to the predicted outcome – here, completion. However, what is interesting is that the first correct set of test results already could predict this outcome, as the difference between completers and non-completers is already significant – although, of course, the difference between means is still small, only in week 3 becoming large enough to possibly notice more. We removed students with 0 total answers from all our analyses of quizzes; however, it is important to note that the data remaining was still quite sparse: many students decided to only answer one question, e.g., in the early weeks, which is somewhat understandable, as students may have tried to attend, but then had to give up (due to the difficulty of the course, or simply due to other constraints, such as time). However, we also had students who only engaged with, e.g., questions in Week 4, which raises further questions on the ways the students in MOOCs learn and interact with these systems, and how their learning goals may not coincide with those of the **designers** (e.g., it is possible that only subjects in Week 4 were interesting for those students, and thus their learning goal had been achieved with that sparse activity). **Teachers and designers** can benefit from such analyses generating feedback for their course, for future improvements. It is clear from Fig. 6, for instance, that Week 6 was especially difficult for the 'Shakespeare and His World' course, which can be fed back to the **course designers**. Fig. 8 strengthens the conclusion that completers are more active, even if they get many questions wrong – but, interestingly, this happens more at the start of the course (upper right, weeks 1-4), after which students possibly get 'the hang of it'. Fig. 4 and Fig. 9 show the potential of finding separate classes, which can serve as predictors, or, even more importantly for a teacher, as groups of users for which a certain type of intervention is necessary (**manually** or done by **a personalisation rule-based system**). Thus, even if prediction may not be possible for all learners at a given stage, certain sub-groups could be given attention. For instance, learners who don't attempt any quizzes in a certain week, thus at risk of disengaging could be (**automatically** or via a **tutor**) encouraged to at least try. Moreover, learners who try but fail could be encouraged by **automatic** or **manual** messages stating that those who try many times are more likely to complete, etc. – depending on the temporal pattern of the respective week, and based on precise statistical data. With the analysis as part of an integrated platform, a **teacher** could get lists of students needing reminders, depending on the **design aims of the IS running the MOOC**.

Overall, we can see that courses with more runs can be better predictors, which is very useful. This means that the next time such a course is run, the teachers can have clear expectations in terms of student behaviour, and can take early measures against undesired outcomes (such as, but not exclusively, completion). However, the results also show that the self-declared completion (pressing the 'complete' button) is less of a reliable indicator of actual involvement in FutureLearn courses than the quizzes are. Thus, any predictor of completion should take the latter into account, and could ignore the former.

Summarising, the main contributions and findings of the study presented here are:

- We have described and visualised *novel detailed temporal patterns of quiz answering*, including pairwise comparisons of early weeks (1, 2, 3) and last week in terms of the number of correct/ incorrect answers as well as completion.
- *Completion is significantly correlated to behaviour*, as follows:
    - o The number of question attempts (both wrong and correct) decrease significantly over the term for students who do not complete the course.
    - o The difference in the number of correct and incorrect question attempts each week for completers versus non-completers is statistically significant as early as the 3rd week of the course. Those who do not complete the course have, on average, few question attempts (both correct and incorrect).
- *Novel temporal behaviour patterns of significance are found*: The correlation between number of question attempts in early weeks of the course, versus the last increases throughout the term, and is statistically significant from the third week onwards, for individual courses, and starting in week 1, for courses overall (Table 2).
- Specific examples for *teacher and designer feedback* and *adaptation* are discussed.

6. Conclusion and Future Work

In this paper we have analysed how learners are learning in MOOCs in general, and, specifically, in all FutureLearn courses of the University of Warwick. Important novel features of our research are the longitudinal aspect (of a truly long-term study of 21 runs of 5 courses), the systematic approach and analysis of the features, the focus on the early prediction with relatively simple variables, and the temporal aspect of our analysis. Additionally, our contributions include the discussion on personalisation rules which could be introduced, either automatically (via the **design of the IS**), or via a **teacher's feedback**, based on the analysis and visualisations provided. Scatterplots can provide a first insight into the clusters available. We also recommend analysis of courses separately, due to their clear differences in the way various variables are instantiated (i.e., test timings, test length, test difficulty, etc.).

Further work includes analysing students accessing courses but not answering quizzes, as their motivation may shed light on other reasons for non-completion. Similarly, we will analyse the rich comments exchanged by students, potentially with graph-based methods.

Acknowledgements

References

1. Agrahari, A., & Rao, D. D.: A Review paper on Big Data: Technologies, Tools and Trends (2017).
2. Atenas, L.: Model for democratisation of the contents hosted in MOOCs, Internat. Journal of Educational Technology in Higher Education 12(1), 3–14 (2015).
3. Bello-Orgaz, G., et al.: Social big data: Recent achievements and new challenges. Information Fusion, 28, 45-59 (2016).
4. Breslow, Lori, et al.: Studying learning in the worldwide classroom: Research into edX's first MOOC. Research & Practice in Assessment 8 (2013).
5. Castaño-Muñoz, J., et al.: Does digital competence and occupational setting influence MOOC participation? Evidence from a cross-course survey. Journal of Computing in Higher Education 29.1. 28-46 (2017).
6. Clow, D.: MOOCs and the funnel of participation. Proceedings of the Third International Conference on Learning Analytics and Knowledge. ACM, 2013.

7. Cristea, A. I., et al: Earliest Predictor of Dropout in MOOCs: A Longitudinal Study of FutureLearn courses, ISD'18 (2018).

8. Crossley, S.,et al.: Predicting Success in Massive Open Online Courses (MOOCs) Using Cohesion Network Analysis. CSCL'17, 1, Int. Soc. Learning Sciences (2017).

9. Gardner, J., Brooks, C.: Toward Replicable Predictive Model Evaluation in MOOCs, EDM 2017, 354-355 (2017).

10. Ferguson, Rebecca. The State of Learning Analytics in 2012: A Review and Future Challenges. Technical Report. KMI, Open University, UK (2012).

11. Hansen, J. D., & Reich, J.: Socioeconomic status and MOOC enrolment: enriching demographic information with external datasets. LAK'15, 59-63. ACM (2015).

12. Lazer, D., et al.: The parable of Google Flu: traps in big data analysis. Science, 343(6176), 1203-1205 (2014).

13. Lei, S., & Cristea, A. I., Demographic Indicators Influencing Learning Activities in MOOCs: Learning Analytics of FutureLearn Courses, ISD'18 (2018).

14. Liñán, C. L., et al.: Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. RUSC. Universities, Knowledge Soc. J., 12(3), 98 - 112 (2015).

15. Lu, X., Wang, S., Huang, J., Yan, Z.: What Decides the Dropout in MOOCs?, Database Systems for Advanced Applications. DASFAA'17 Int. Workshops Suzhou, China, March 27-30, 316-327, (2017).

16. Ng, A., Widom, J.: Origins of the Modern MOOC (xMOOC), MOOCs: Expectations and Reality: Full Report, 34-47 (2014).

17. Özdemir, V., et al.: Personalized medicine beyond genomics: alternative futures in big data—proteomics, environtome and the social proteome. Journal of Neural Transmission, 124(1), 25-32 (2017).

18. Pursel, Barton K., et al.: Understanding MOOC students: motivations and behaviours indicative of MOOC completion. J. of Comp. Assist. Learning 32.3 (2016): 202-217.

19. Qiu, J. et al.: Modeling and Predicting Learning Behavior in MOOCs, WSDM 2016, Proc. of the 9th ACM Int. Conf. on Web Search and Data Mining, 93-102 (2016).

20. Robinson, C. et al.: Forecasting Student Achievement in MOOCs with Natural Language Processing, LAK '16, ACM, 383-387 (2016).

21. Sivarajah, U., et al.: Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, 263-286 (2017).

22. Shi, L., et al.: Towards Understanding Learning Behavior Patterns in Social Adaptive Personalized E-Learning Systems. Proc. of the 19th Americas Conference on Information Systems (AMCIS), 1-10 (2013).