Data Analytics Informing MOOC Continuous Improvement

John Vulic[1], Mahsa Chitsaz[1], Ganga Prusty[1,2] & Robin Ford[1,2]

[1] UNSW Sydney Australia, Office of Pro-Vice Chancellor (Education)
[2] UNSW Sydney Australia, Faculty of Engineering
j.vulic@unsw.edu.au, m.chitsaz@unsw.edu.au,
g.prusty@unsw.edu.au, robinford1@a1.com.au

ABSTRACT. In 2016 UNSW Australia (The University of New South Wales) designed and developed the Massive Open Online Course (MOOC) 'Through Engineer's Eyes: Engineering Mechanics through experiment, analysis and design' (TEE). Two iterations of TEE were run that year on the FutureLearn (FL) platform. The data generated from student engagement with the MOOC was examined after the first course offering, and this informed various design changes aimed to improve learner experience in the second and future offerings of the course. This paper provides useful and usable insight into MOOC design, development and ways that data analytics can inform the continuous improvement over time.

Keywords: MOOC; FutureLearn; Data Analytics; Engineering; Education

1 Introduction

This paper examines the evolution of an engineering mechanics MOOC offered on the FL platform across two course offerings. MOOC platforms have offered access to courses to anyone in the world with an internet connection and an interest in learning. MOOCs have traditionally attracted large enrolment numbers, usually in the tens of thousands (Agarwal, 2014; Jordan, 2014). This has enabled the sharing of knowledge-making between people in varied geographical locations and of differing educational backgrounds, based on a common interest (Vigentini et al 2016). This presents challenges to both MOOC developers and educators. TEE offers insights into how data analytics informed and helped optimise the design of this MOOC to improve learner experience.

FutureLearn data: what we currently have, what we are learning and how it is demonstrating learning in MOOCs. Workshop at the 7th International Learning Analytics and Knowledge Conference. Simon Fraser University, Vancouver, Canada, 13-17 March 2017, p. 63-73.
Copyright © 2017 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

2 The Aim of the MOOC

A key aim of TEE was to introduce learners to the world-view of an engineer by demonstrating how engineers use analysis to understand their surroundings and to predict the behaviour of the things they design. TEE course content was designed to be anchored in practical reality and to provide learners with experience on which to base their studies of classical analysis. The course was designed to be accessible to a global audience, this informed the design of experiments that learners could conduct in their own time. The experiments aimed to spark learner interest in the topics and ground these in physical reality. Experiments demonstrated the use of commonly available items, such as rubber bands, cardboard, string and toy vehicles to explain complex engineering concepts. Analysis activities helped to explain the experiments and lead learners through the design process.

2.1 Initial Design Considerations

Initial design considerations of TEE centred on what the potential target audience would be for the course. A previous MOOC on a related subject suggested the demographic for the course would cover an age range from 16 to over 65, a range generally consistent with other related MOOCs. This provided a challenge in how to make TEE accessible and interesting to this broad and eclectic global cohort. One early decision was to focus TEE on teaching basic engineering mechanics, with a style that would be friendly, authoritative and fun. Learners would however need knowledge of basic trigonometry and algebra. Although this distinction has been challenged (Lukes 2012; Conole 2014), there are two well recognised types of MOOCs: cMOOCs - or connectivist MOOCs (Siemens 2005) which focus on community and peer interaction, and xMOOCs (McAuley et al. 2010; Rodriguez, 2012), normally driven by content and knowledge, often using automation of activities in order to accommodate large number of learners. TEE was designed to sit somewhere between these two types of MOOCs. This complimented the selection of FL as the platform to host the course. Of particular interest was the focus of the platform on narrative-led, collaborative and conversational learning, which was seen to compliment both the style of the MOOC and the broad demographic.

2.2 Structure of the MOOC

The course was modularised following a seven-week structure that covered the topics in Table 1.

Learners were led through the course by short 1-4 minute videos (over 50 in total), accompanied by supporting text resources. Each week an introductory video set the scene, followed by a video/s on the week's experiment. If learners decided not to physically attempt the experiments themselves, they nevertheless could identify with the activities because of the familiar nature of the equipment used.

Table 1. Structure of the TEE MOOC

Topic	Experiment	Analysis	Design
Elastic properties	Load-deflection of a rubber band	Stiffness	No design activity
Forces that act at a point	Measuring forces that act at a point	Adding forces that act at a point	Cables for suspending a loudspeaker
Forces on a rigid body	Moments, forces on a rigid body	Equilibrium in two dimensions	Connections for a folding washing line
Centre of gravity	Finding cg by suspension and balancing	cg of a composite body	Specifying the ballast weight for a model glider
Friction	Basic friction model, tip/slide, rope around a bollard	Basic friction model, tip/slide, rope around a bollard	Belt drive for a model car
Work and energy	Rolling resistance, aerodynamic drag/lift	Rolling resistance, aerodynamic drag/lift, work and power	Design evaluation of electric vehicles
Impulse and momentum	Shove ha'penny	Impulse/momentum	No design activity

The course design also incorporated several tools and resources into the FL platform. These were designed to promote collaboration and sharing amongst learners, provide rich interactive and adaptive courseware and promote learning consolidation. One of these tools included ‘Padlets’, which were added to each experiment. These were virtual walls that allowed learners to share images, videos and descriptions of any experiments that they attempted. Links to these could also be added to the discussion forums to elicit further discussion amongst learners. The structure of the course was also complimented with the inclusion on-line SmartSparrow Adaptive Tutorials. SmartSparrow is a learning design platform that enabled the incorporation of rich, interactive and adaptive e-learning courseware (Ben-Naim & Prusty 2010; Prusty et al, 2011). Another addition was the inclusion of ‘Retro Tutorials’. These consisted of downloadable PDF format exercises typically found supporting tutorials in university level courses, and were designed to assist learners in consolidating their learning each week. The inclusion of these tools and resources seamlessly blended with existing tools and resources available in the FL platform.

2.3 Wrangling MOOC Data

Large amounts of data were generated from learners' interactions, both with the course and with fellow learners. TEE learner data was sourced from both the FL platform and SmartSparrow. The FL platform is a pioneer in providing near real-time data of its published courses. The data sets are updated daily, and this creates an opportunity to analyse learner interaction and behaviour while a course is active. The available data sets for the TEE course included campaigns, comments, enrolments, question response, step activity and team members. The purpose of each file is described in Table 2. These data sets are downloadable as CSV (Comma Separated Values) files.

There are two sources of demographic information in FL: a profile survey that asks learners for their basic information such as age, gender and level of education, and a pre-course survey that focuses on learner motivation to enrol and goals. As both surveys are optional, the information gleaned should be used with caution as the responding sample (approximately 10% in both iterations) might not be fully representative. This demographic information does however provide a useful portrait of learners.

Table 2. FutureLearn Datasets

File	The purpose of the file
Campaigns	Information about the referral used to advertise a course is stored in this file, following the number of enrolments and active learners for each referral.
Comments	Information about learners' contributions to the discussion section in each step is stored in this file. It includes the text of the comment and the timestamp corresponding to when the comment was made.
Enrolments	This file provides basic information regarding the enrolled learners. It also includes demographic information of learners derived from the profile survey.
Question Response	This file holds information about the quiz activity of learners. It stores learners' responses, its correctness and the timestamp associated when answering a question of any quiz.
Step Activity	This file stores information regarding step activity from learners in the course, e.g. the time when a step is first visited, and the last time a step is marked as completed.
Team Members	Information about organization staff such as their ids and names are stored in this file.

A MOOC dashboard was created at UNSW Australia for courses published in FL platform (Chitsaz, Vigentini, & Clayphan, 2016). Raw data from the abovementioned sources was converted to a visual context using R and Python programming languages.

The dashboard provided numerous ways to conveniently analyse MOOC analytics in near real-time. Some of the data visualization options are shown in Table 3.

Table 3. MOOC Dashboard

Heading	Data Visualisation Description
Adaptive Tutorials (SmatSparrow)	Grade on Lesson: A histogram of the earned grades among all learners
	Time Spent on Lesson: A histogram of the time spent on each lesson among all learners
Demographics	Different types of visualisations to show the geographical distribution, gender distribution, gender vs. employment status, gender vs. age range, education Distribution, and employment area distribution.
Activity	Having multiple visualisations to analyse the step activities of learners. For example, the percentage of time spending on each week, finding the number of leavers at any step or any date of the course, a heat map to draw the step completion progress of learners, and transition networks between available materials of the course by step type or week number.

2.4 Learners

Approximately 7000 learners registered for the first run of the course with 40% actively engaging with the course at some point while it was open. Similar to the patterns already identified with the funnel of participation (Clow, 2013), a much smaller proportion (7%) of these ‘completed’ the course. The figures are slightly lower in the second run (4337 learners, 36% active and 2.5% completing).

'Active learners' are defined by FL as those who actively engage with some content while the course is open, and 'completing' refers to those who self-mark at least 90% of the steps in the course as complete. Due to the nature of the platform, active learners may have visited and completed learning activities, but may have not self-marked the step as completed, indicating a potential for under-estimation of the number of completers in the course. As anticipated in the design stage of the course, only a small proportion of learners obtained a paid certificate.

From the sample of survey responses, the typical learner in the TEE MOOC was male (71% of respondents), aged between 18-25 (26%), in full time employment (34%) and with an undergraduate degree (40%). The summary table below (Table 4) provides an overview of the distributions. This second run of the course revealed similar responses, with the typical learner being male (62% of respondents), aged between 18-25 (34%), in full time employment (33%) and with an undergraduate degree (40%).

Table 4. Overview of learners' characteristics based on survey responses (N=119), Demographic, response rates and Confidence Intervals (CI)*

demographic	**response rates and CI***	**Category distributions**
TEE – 1st Iteration		
Gender	9.26% ± 0.68%	Male (71%), Female (28%), Other (1%)
Age range	9.09% ± 0.67%	<18 (6%), 18-25 (26%), 26-35 (26%), 36-45 (13%), 46-55 (11%), 56-65 (9%), >65 (10%)
Employment	9.22% ± 0.68%	Full time worker (34%), Full time student (20%), Retired (10%), Looking for job (10%), Self-employed (9%), Part time worker (7%), Unemployed (5%), Not working (4%)
Highest level of Education	9.29% ± 0.68%	Undergrad degree (40%), Secondary (25%), Master degree (14%), Tertiary (12%), Less than sec. (6%), PhD degree (4%), Professional (5%), Apprenticeship (1%)
TEE – 2nd Iteration		
Gender	3.94% ± 0.01	Male (62%), Female (37%), Other (1%)
Age range	2.1% ± 0.01	<18 (9%), 18-25 (34%), 26-35 (23%), 36-45 (14%), 46-55 (11%), 56-65 (6%), >65 (3%)
Employment	2.1% ± 0.01	Full time worker (33%), Full time student (26%), Retired (3%), Looking for job (11%), Self-employed (9%), Part time worker (7%), Unemployed (5%), Not working (7%)
Highest level of Education	2.1% ± 0.01	Undergrad degree (40%), Secondary (19%), Master degree (19%), Tertiary (9%), Less than sec. (5%), PhD degree (4%), Professional (3%), Apprenticeship (2%)

2.5 Learner time spent in the MOOC

From the logs of interaction with the platform it is possible to identify several trends. Learners who engaged with the content spent between 90 minutes to two hours on average per week in the course. This equates to roughly 5-10 minutes per step. FutureLearn uses the concept of 'step' which can incorporate a variety of artefacts including articles, video, discussion, quiz, exercises etc. Table 5 provides a summary overview of the time spent in the course by active learners for both iterations.

Table 5. Average time spent per step and per week with actual distributions in 1st and 2nd iterations

1st Iteration of TEE			
Week	**N Steps**	**Avg mins to complete a step**	**Avg mins spent in week**
1	13	7.70	58.61
2	16	8.98	99.28
3	20	7.55	114.78
4	18	6.20	92.52
5	17	6.16	83.83
6	16	6.65	89.39
7	10	5.98	45.28
2nd Iteration of TEE			
Week	**N Steps**	**Avg mins to complete a step**	**Avg mins spent in week**
1	13	8.05	61.65
2	16	8.47	85.80
3	18	6.66	95.24
4	15	5.63	73.01
5	14	6.04	68.41
6	13	5.76	66.45
7	10	4.07	33.16

2.6 Continuous Improvement

In the first iteration of the course, learners spent more time in week 3 (114.78 average minutes) than in other weeks of the course (Table 5). This also correlated to a steeper drop in engagement during the first three weeks of the first iteration. The number and percentage of the leavers at any week is shown in Table 6 for both iterations. Typically, a large proportion of learners leave the course in the first week of any MOOC. The reasons for this will vary and require further research. Reasons may possibly relate to learner expectations not being met, or factors such as personal commitments hindering continuation and completion of a course.

Table 6. The number (percentage) of leavers at any week

Week	TEE – 1st Iteration	TEE – 2nd Iteration
1	1761 (68%)	1098 (73%)
2	447 (17%)	261 (17%)
3	207 (8%)	77 (5%)
4	54 (2%)	21 (1%)
5	37 (1%)	10 (1%)
6	40 (2%)	15 (1%)
7	60 (2%)	29 (2%)

Qualitative data in the form of discussion forum comments from learners in week 3 of the first iteration showed they experienced difficulty with some of the activities in this week. The risk with problems being too easy is learners may lose interest quickly; conversely problems that are too difficult may potentially place strain on the working memory on novices (Kirschner et al. 2006). Week 3 was considered an important week in the course and various changes were made to this week to improve the course for the second iteration.

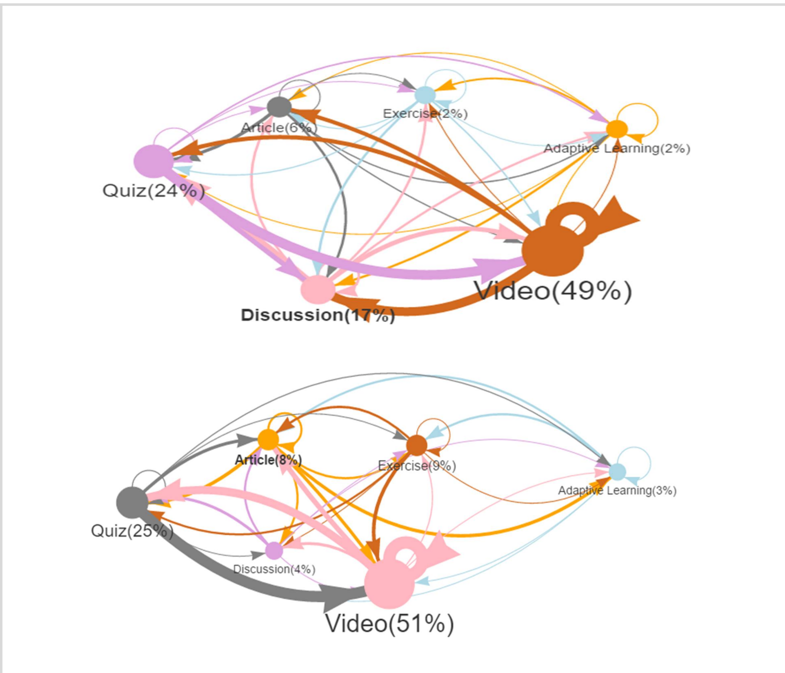

Fig. 1. Transition by Type – iteration 1 (top) and 2 (bottom) showing the transition of learners among materials of all weeks based on the step type

The story-line for week 3 was streamlined, and the design step for this week was also divided into two parts. The overall structure of the course was simplified by merging a majority of discussion forums that were originally separate steps into the step that they related to as ‘talking points’. The intention behind this was twofold. Firstly, it was hoped that fewer steps would make tasks in the course appear less intimidating to learners, secondly, a reduction in steps simplified the job of instructors by reducing the number of places they had to monitor in the course. This resulted in a general change in the transition of learners among materials of all weeks in the second iteration of the course (Figure 1).

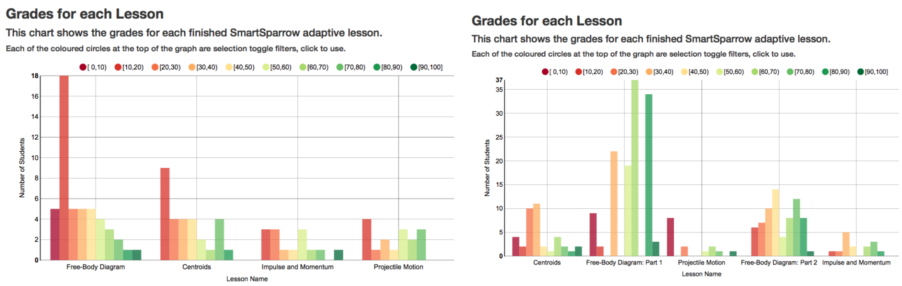

Fig. 2. Grades for each SmartSparrow Adaptive lesson - iteration 1 (top) and 2 (bottom)

The Adaptive Tutorials were also revised for week 3. On a technical level the UI was improved in the second iteration by enhancing the accessibility of the adaptive tutorials for mobile devices such as iPads, creating several new drag-and-drop activities, improving the adaptive feedback, and adding better quality LaTex equations. General information screens were added at the beginning of each Adaptive Tutorial lesson to help orient learners to the features found in the adaptive tutorials. Qualitative feedback gleaned from the discussion forums in these steps also suggested too much complex information in some Adaptive Tutorials. In response, some Adaptive Tutorials were chunked, such as the Free-Body Diagram, into two smaller learning segments to make them easier for learners to understand and complete in less time. The result of these changes included learners spending less time in week 3 of the course in the second iteration as compared to the first (as seem in Table 5). Splitting the Free-Body Diagram Adaptive Tutorial also resulted in more learners achieving higher scores in the tutorial, as seen in Figure 2.

3 Conclusion and Future Directions

The TEE MOOC has reinforced for us how important it is to analyse the learning experiences of the courses we offer as part of the cycle of continuous improvement. Offering this MOOC has enabled thousands of learners to have access to a free course in the fundamentals of Engineering Mechanics, but this brings with it correspondingly increased responsibility to do it well.

By leveraging on the data we can make informed choices about the design of the course and thereby improve the learning experience of a global cohort of learners. In this way we have created a data-driven course development process that provides learners with the best learning experiences possible – wherever they are in the world. There are still challenges in accommodating broad and large demographics of learners. For example, mathematics was intentionally kept as simple as possible, however basic algebra and simple trigonometry challenged a number of learners, as evidenced in the discussions. The changes made to the course were implemented after the first iteration. A challenge lies in how agile this process can be, such as whether near real-time data can also be leveraged to inform course design changes in near real-time.

The overall aim in offering TEE has been simple: to offer to a wide range of people an understanding of engineering mechanics through experiments, analysis and design, whether for general interest or in preparation for an engineering future. We are offering them all a chance to see the world "Through engineers' eyes".

4 REFERENCES

1. Agarwala, M. (2014). A Research Summary on MOOC Completion Rates | EdLab. Retrieved April 23, 2015, from http://edlab.tc.columbia.edu/index.php?q=node/8990
2. Ben-Naim, D. and Prusty, B.G. (2010) Towards a Community of Practice concerning the Use of Adaptive Tutorials in Engineering Mechanics, AaeE 2010, Sydney, 5-8th Dec 2010, Australia.
3. Clow, D. (2013). MOOCs and the Funnel of Participation. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 185–189). New York, NY, USA: ACM. http://doi.org/10.1145/2460296.2460332
4. Conole, G. (2014). A new classification schema for MOOCs. INNOQUAL - International Journal for Innovation and Quality in Learning, 2(3). Retrieved from http://papers.efquel.org/index.php/innoqual/article/view/164
5. Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. The International Review of Research in Open and Distributed Learning, 15(1). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1651
6. Kirschner, P.A., Sweller, J. & Clark, R.E., (2006). Why Minimal Guidance During Instruction Does Not Work : An Analysis of the Failure of Constructivist , Discovery , Problem-Based, Experiential, and Inquiry-Based Teaching. Educational Psychologist, 41(2), 75–86.
7. Lukeš, D. (2012, August 14). What is and what is not a MOOC: A picture of family resemblance (working undefinition) #moocmooc. Retrieved from http://researchity.net/2012/08/14/what-is-and-what-is-not-a-mooc-a-picture-of-family-resemblance-working-undefinition-moocmooc/

8. McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). The MOOC model for digital practice. Retrieved from http://www.davecormier.com/edblog/wp-content/uploads/MOOC_Final.pdf
9. Prusty, B. G., Ben-Naim, D., Ho, S. and Ho, O. (2011) Online Adaptive Tutorials Targeting Fundamental Concepts of Mechanics Courses in Engineering, Engineering Education - An Australian Perspective, Multi-Science Publishing Co Ltd., Australia.
10. Rodriguez, C. O. (2012). MOOCs and the AI-Stanford Like Courses: Two Successful and Distinct Course Formats for Massive Open Online Courses. European Journal of Open, Distance and E-Learning. Retrieved from http://eric.ed.gov/?id=EJ982976
11. Siemens, G. (2005). Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning, 2(1), 3–10.
12. Vigentini L, McIntyre S, Mirriahi N, Alonso, D. (2016) Exploring the real flexibility of learning sequences: Does course design constrain students behaviours or do students shape their own learning? in ElAtia S; Ipperciel D, Zaiane O (eds.). Data Mining and Learning Analytics in Educational Research, 2015, Wiley and Blackwell
13. Chitsaz, M., Vigentini, L., & Clayphan, A. (2016). Toward the development of a dynamic dashboard for FutureLearn MOOCs: insights and directions. Presented at the 33rd International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education (ASCILITE), Adelaide, Australia.