Information and Learning Science

Participating by activity or by week in MOOCs

Alok Baikadi, Carrie Demmans Epp, Christian D. Schunn,

Participating by activity or by week in MOOCs

Alok Baikadi
Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
Carrie Demmans Epp
Department of Computer Science, University of Alberta, Edmonton, Canada, and
Christian D. Schunn
Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Received 29 April 2018
Accepted 23 June 2018

Abstract

Purpose – The purpose of this study was to provide a new characterization of the extent to which learners complete learning activities in massive open online courses (MOOCs), a central challenge in these contexts. Prior explorations of learner interactions with MOOC materials have often described these interactions through stereotypes, which accounts for neither the full spectrum of potential learner activities nor the ways those patterns differ across course designs.

Design/methodology/approach – To overcome these shortcomings, the authors apply confirmatory and exploratory factor analysis to learner activities within three MOOCs to test different models of participation across courses and populations found within those courses.

Findings – Courses varied in the extent to which participation was driven by learning activities vs time/topic or a mixture of both, but this was stable across offerings of the same course.

Research limitations/implications – The results call for a reconceptualization of how different learning activities within a MOOC are designed to work together, to better allow strong learning outcomes even within one activity form or more strongly encourage participation across activities.

Originality/value – The authors validate new continuous-patterns rather than a discrete-pattern participation model for MOOC learning.

Keywords Participation, MOOCs

Paper type Research paper

Introduction

The appearance of massive open online courses (MOOCs), their potential to democratize learning and the media attention they initially received have more recently given way to disillusionment over their inability to fulfil the promises of their early proponents. Much of this disillusionment centered on their high attrition rates and the demographics of those who succeeded – predominantly educated men living in North America or Europe (Kizilcec and Halawa, 2015). Arguments around the importance of considering learner intent emerged to explain why so few learners completed MOOCs (Kizilcec *et al.*, 2013). However, user intent is only part of the story, and simple explanations that stereotype users as completers and topic shoppers inhibit our understanding of how people interact with this technology. We need a better understanding of the nuanced ways in which learners choose to interact with freely available educational resources (Anderson *et al.*, 2014; Baikadi *et al.*, 2016; Gasevic *et al.*, 2014) if we are to improve them by adding new features, designing better content or

Information and Learning Science
© Emerald Publishing Limited
2398-5348
DOI 10.1108/ILS-04-2018-0033

designing better interactions among learners, their peers, support staff, the system and the content. For example, it is possible to turn MOOCs that are offered through the EdX platform into systems that recommend supplementary learning content to individual students, when more nuanced approaches are used to understanding student preferences and knowledge (Pardos *et al.*, 2017).

This work provides a more nuanced understanding of learner interactions with MOOCs by modeling different learner behaviors and allowing learners to exhibit a combination of behaviors, rather than assigning a single stereotype, even if data-derived, to each user. This reformulation means that learners are no longer considered as belonging to an exclusive user type. Rather, the variability in user activities is captured through their engagement with different actions within the system. The more refined view of learner-system interactions that is proposed in this paper and recommended by Sinha *et al.* (2014) also accounts for differences in how learners interact with the system over time. Moreover, this new approach to analyzing user interactions within MOOCs includes all users who completed any activities rather than only including those who complete the course, a demonstrably non-representative population.

To provide this more nuanced modeling of user-system interaction, theory-testing factor analysis was used to identify the coherent clusters of behaviors that learners exhibit when participating in a given MOOC. From the literature (describe below), one theoretical model assumed a week-by-week structure to the factors (e.g. a factor modeling the extent to which participants completed Week 1 activities, a factor modeling the extent to which participants completed Week 2 activities). A second theoretical model assumes activity-centric factors (e.g. a factor modeling the extent to which participants watched lectures, a factor modeling the extent to which participants completed quizzes). A given participant is then conceptualized as having different levels on each factor (e.g. high on Week 1 activities, moderate on Week 2 activities in the first model; or moderate on video watching, high on quiz taking, moderate on forum posting in the second model). Each of the two models can be fit to the data for a given MOOC offering, and the best fitting model can be determined.

This factor-analytic approach reveals how common interaction patterns vary across MOOCs while also allowing each user to exhibit his or her own combination of interaction patterns. This exploration is important to building a better understanding of how course design elements and system features encourage engagement from students so that MOOCs can be designed to support a broader range of learners through the inclusion of new features and socio-technical processes.

Keeping this higher-level goal in mind, data from multiple offerings of MOOCs that aimed to support language learning, planning and content learning in the health sciences were analyzed. This analysis of data from 106,143 learners across nine MOOC offerings used exploratory and confirmatory factor analyses to test the following hypotheses:

H1. Participation in a given MOOC will be better fit by one of two different theoretical models of coherent factors of user behaviors (based on weeks vs based on activities).

H2. The best fitting factor model will be consistent from one offering of a course to another.

H3. The best fitting model will vary across courses based on coherence of content in the MOOC.

Participation in online learning environments
Courses have been offered via online tools for over 20 years (Selwyn, 2014). Initial offerings aimed to enable post-secondary access for traditionally underserved populations by providing structured learning experiences to students registered in established university programs. These courses often targeted distance learners and professionals by delivering content through an online learning environment (OLE), either one that also supports face-to-face courses, such as Blackboard, Bright space, Canvas or Moodle, or through a specialized OLE that focuses on online-only courses, such as EdX and Coursera. OLEs in general were seen as an improvement over prior distance-education approaches because they allowed for students to interact with the content, one another, and course staff (Kenny, 2002).

A considerable amount of recent research into the use of OLEs has explored student attrition and retention (Guo and Reinecke, 2014; Rosé and Ferschke, 2016; Rovai and Wighting, 2005; Sinha *et al.*, 2014) and student use of specific system features that include video replay, reading or writing forum posts, system log ins and the use of social-media-like features (Brooks *et al.*, 2011; Dawson *et al.*, 2008; Nelson, 2015; Phirangee *et al.*, 2016). This and other work has shown that the manner in which OLEs structure materials and activities can cause difficulties for students (Marshall *et al.*, 2015) or result in students feeling poorly supported (Phirangee *et al.*, 2016; Wilcox *et al.*, 2016). These shortcomings partially explain the high attrition rates (Park and Choi, 2009) and increased demand for instructor support (Marshall *et al.*, 2015; Rosé and Ferschke, 2016) that are often observed in online courses.

Participation in massive open online courses
Initial attempts at allowing open access to quality learning resources and experiences outside of a formal educational program was believed to revolutionize education (Kovanović *et al.*, 2015). However, the high student attrition rates observed in MOOCs (Coffrin *et al.*, 2014; Gasevic *et al.*, 2014; Kizilcec and Halawa, 2015; Kizilcec *et al.*, 2016; Siemens, 2013; Wen *et al.*, 2014) and the poor completion rates of already underrepresented groups (Breslow *et al.*, 2013; Guo and Reinecke, 2014; Kizilcec and Halawa, 2015) indicate this new framing of online learning suffered from many of the same challenges as that of previous online-education efforts: students felt disconnected or socially isolated (Yang *et al.*, 2013) and poorly supported (Rosé and Ferschke, 2016), lacked self-regulated learning skills (Kizilcec *et al.*, 2016), lacked motivation (de Barba *et al.*, 2016) and were not actively engaging in learning activities (Guo and Reinecke, 2014; Koedinger *et al.*, 2015; Ramesh *et al.*, 2014).

Initial research into how students engaged with the major MOOC platforms (e.g. Coursera and Edx) focused on student attrition and the reasons behind that attrition: student preparedness as a function of prior education (Breslow *et al.*, 2013; Kizilcec and Halawa, 2015), student intent (Gasevic *et al.*, 2014) and varied social factors such as structural attributes of students' social networks, lack of social presence or social comparison and the timing of their interactions with fellow students (Davis *et al.*, 2017; Gasevic *et al.*, 2014; Rosé *et al.*, 2014) were among those reasons. This line of research then gave way to an exploration of how students interact with the MOOC from the perspective of how and when they used different types of MOOC resources and features (Pursel *et al.*, 2016; Van der Sluis *et al.*, 2016).

Overall, these explorations of student use of MOOC resources and features fail to provide a nuanced understanding of how students interact with MOOC resources. They instead identify stereotypical patterns in student learning activities that are formulated as personae describing a user type (Anderson *et al.*, 2014; Corrin *et al.*, 2017; Gasevic *et al.*, 2014; Guo and Reinecke, 2014; Kovanović *et al.*, 2016; Liu *et al.*, 2016). For example, Anderson *et al.* (2014) identified five types of users based on student interactions with video lectures and graded assignments. Rather than providing nuanced analyses of user behaviors, these user types characterized learners as conforming to specific patterns of usage that included:

- mostly watching videos;
- mostly submitting assignments;
- both watching videos and doing assignments;
- registering but doing very little; and
- collecting resources by downloading them.

Similarly, Swan *et al.* (2016) identified three patterns (i.e. acquisition, participation and self-direction), which were characterized by the pedagogical philosophies enacted within the MOOCs they studied. Explorations of these user engagement personae have even gone so far as to characterize how students will morph from one personae to another between different offerings of the same MOOC (Kovanović *et al.*, 2016) or how they will cheat the system by creating multiple profiles to ensure certification (Ruiperez-Valiente *et al.*, 2016). These types of analyses align with claims that there are a small number of ways in which students interact with MOOC resources (Kizilcec *et al.*, 2013). However, the nature of the analysis and its output ensure the development of these group-level descriptors, despite the diversity present in MOOC courses and the potential for highly variable student behaviors. We instead take a variable-centered statistical approach to detail the potential for highly variable activity patterns, which enables a more refined understanding of user engagement without losing the explanatory power of cruder analysis approaches, such as personae.

Methodology

We conduct a mixture of exploratory and model-testing analyses of learner activities within several MOOCs. These analyses aimed to identify learner activity patterns that are common across courses and those that vary across courses, while allowing each learner to exhibit a different balance of activity patterns.

Massive open online course platform

We conducted our analyses using data from MOOCs offered through Coursera, which handles registration, accessing of lecture videos, quizzes, surveys, discussion boards, assignment submission and exams. Course assignment can also include papers, design-projects or the creation other artefacts. These more open-ended tasks are graded by the students' peers who are expected to follow a grading rubric. See Leontyev and Baranov (2013) and Siemens (2013) for additional discussion of the features and feature affordances that were available at the time.

Selected massive open online courses

To enable the description of learner activities from a range of course types, we selected three courses from two different domains: health sciences (Nutrition for Health and Clinical Terminology) and public health (Disaster Preparedness). Each course was offered multiple times. Each offering followed the same curriculum, using the same course activities with a different student population. For this work, we analyzed nine offerings across the three courses.

The selected courses were offered by the University of Pittsburgh, a large public research-intensive university in the USA. These courses were offered through the session-based version of Coursera, which means that they had pre-defined start and end dates. All courses took place between January 2013 and December 2015.

Disaster Preparedness helped learners to develop their disaster readiness and survival planning competencies. Nutrition for Health provided evidence of how nutrition and physical activity influence health through lower morbidity, increased longevity and increased quality of life. Finally, Clinical Terminology introduced learners to the clinical terms and abbreviations commonly used in US hospitals.

All three courses lasted six weeks, and they organized the course activities similarly: the core materials for each week consisted of a series of video lectures and a quiz. There was variability in course design and the manner in which courses approached supporting user learning. This variability provides a generalizability test for the observed patterns of learner behaviors as well as an opportunity to examine whether course structure or participant demographics would systematically influence patterns of learner behaviors.

In all three courses, a particular set of activities was assigned to each week: several lecture videos, some short quizzes and discussion forums (Figure 1). The number of each activity type varied from week to week, somewhat independently of each other (i.e. there were not always the same number of lectures, quizzes and discussion forums in a given week). In addition, there were sometimes discussion forums that were not associated with a particular week, often as a place to discuss general course issues.

One primary way in which the courses differed was the extent to which each week was conceptually modular vs sequentially building upon one another. The Clinical Terminology syllabus was primarily designed around modular components, with few connections between modules. By contrast, the other two courses had more conceptual interconnections between topics across weeks. Such a difference in modularity could influence how participants select components to complete – more modular courses are easier to consume as isolated topics.

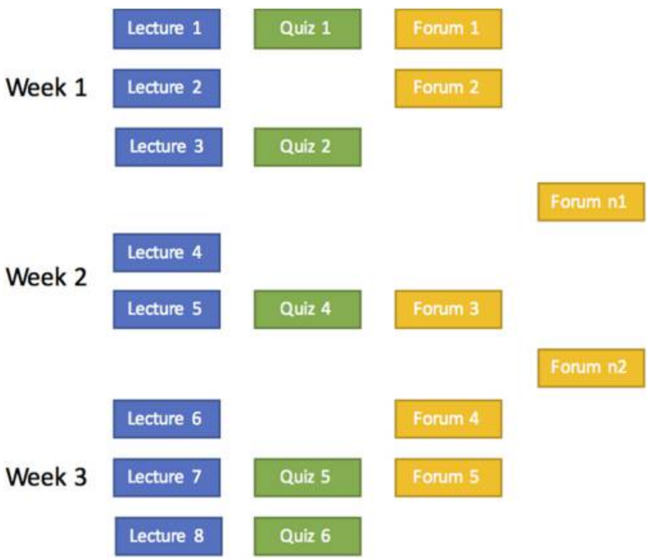

Figure 1. Prototype structure of lecture, quiz, and discussion forum activities within the weekly structure of the session-based MOOCs

Participants

The analyzed courses collectively have 246,245 enrolled learners and 106,143 who performed at least one activity (i.e. posted to the discussion forum, watched a video lecture or submitted a quiz or assignment). As shown in Table I, course enrollment decreased over time possibly reflecting some saturation of the pool of learners interested in these exact topics. This decrease in enrollment matches overall MOOC registration trends (Jordan, 2014).

The demographic information from the 25,529 learners who completed the demographic survey is provided in Table II. The frequency of participation in the demographic survey from offering to offering of a course was similar (Table I). As is typical in MOOCs, participation in each course was broad on every dimension, including having many learners coming from industries not obviously linked to the contents of the course. Further, there was variation in distributions by course, as would be expected by courses drawn from different disciplines (e.g. Clinical Terminology had younger learners and more females).

Data cleaning and analysis

Table III shows the amount of activity performed per learner from each offering of the selected courses; only the activities that account for at least 1 per cent of the actions performed by learners within that course are included in the analysis and table. While very large data sets like this could involve even lower thresholds (i.e. 0.1 per cent) and still include enough data for factor analysis, a much lower threshold would produce factors of little pragmatic value given how few participants engaged in those activities. However, early explorations of the data in one of the courses revealed similar results with both higher and lower thresholds.

For most courses, we used the raw activity counts for each forum, lecture video and quiz. When an instructor has used the quiz feature to conduct pre- or post-course surveys about learner perceptions, those data were removed. An additional pre-processing step was needed for the clinical terminology course since there were two versions of the customized multimedia learning modules that delivered core course content. In this case, the usage of both versions was summed: if a learner used the first version of Module B once and the second version of Module B twice, then that learner's activity count for Module B would be 3.

Once the data had been pre-processed, we used factor analysis to determine the relationships between the activities. Factor analysis attempts to describe the variance in

Table I. The number of learners who registered (enrolled), performed an activity (active) or provided demographic information (demog.) by course and offering

Course	Offering	Enrolled	Active	Demog.
Clinical terminology	All	55,631	28,895	4,460
	1	20,700	8,415	2,202
	2	12,395	7,066	1,068
	3	15,198	9,442	907
	4	7,338	3,972	283
Disaster preparedness	All	48,425	19,510	6,273
	1	29,103	11,329	4,006
	2	9,021	3,855	1,181
	3	10,301	4,326	1,086
Nutrition for health	All	142,189	65,738	14,796
	1	78,804	34,979	8,890
	2	63,385	30,759	5,906

Learner demographics	Clinical terminology (%)	Disaster preparedness (%)	Nutrition for health (%)
Female	58	49	52
Native/Native-like English Speakers	44	70	52
Age			
Under 20	4	3	2
20-29	46	32	37
30-39	24	28	27
40-49	14	18	15
50 +	12	20	18
Current residence			
North America	38	42	38
Europe	29	26	33
Asia	18	15	15
Other	16	17	14
Education level achieved			
Secondary or lower	14	5	9
Some post-secondary	12	12	10
Bachelor	34	41	41
Postgraduate	40	42	40
Current industry			
Engineering related	9	*20*	18
Health related	*46*	*18*	*12*
Education related	10	11	12
Business related	13	21	19
Other	19	28	40

Note: Italics indicate an industry that is directly relevant to the course domain

Table II. Learner demographic information by course

the underlying data in terms of several latent factors of behavior. It is an ideal match to the general conceptual framework being tested in this paper: that individuals have continuously varying preferences along multiple dimensions of participation. These factors can either be inferred from the data, in an exploratory factor analysis (EFA), or be designed *a priori* in a confirmatory factor analysis (CFA). To test *H1, H2* and *H3*, we used CFAs to compare whether a pure weekly approach or a pure activity-centered approach better explained the underlying data. As a follow-up, we then applied an EFA to determine whether some MOOCs involved a hybrid model between the two theoretical models or whether a model with entirely different factor structure emerged. For both CFAs and EFAs, a separate factor analysis was conducted for each offering of each course to test the stability of the resulting patterns (as explicitly the focus of *H2* and *H3*).

The CFA analyses evaluated four different models of learner behavior (a baseline, two focal theoretical models and a hybrid, emergent model). The *Participation* model considered all learner activity to be driven by a single factor of general overall levels of participation. Under this model, any activity the learners performed would be equally indicative of engagement with the course material. This model serves as the baseline for comparison. For the *Weekly* model, the activities for each course were partitioned by the week within which those activities were expected to occur (see Figure 1). Some activities, such as posting to the "General" or "Technical Support" forums, were expected to persist throughout the course

Table III. Median, interquartile range, min and max of learner activities (number of lecture views, number of quiz attempts, number of forum posts) by course offering, when performed by at least 1% of learners

	Lectures watching				Quizzes completion				Forum posts			
Course	Mdn	IQR	Minimum	Maximum	Mdn	IQR	Minimum	Maximum	Mdn	IQR	Minimum	Maximum
Clinical terminology												
1	11	36	0	521	25	97	0	766				
2	9	31	0	753	15	62	0	634				
3	7	26	0	1,434	11	48	0	735				
4	7	24	0	394	9	46	0	720				
5	6	20	0	202	5	42	0	973				
Disaster preparedness												
1	8	32	0	787	1	4	0	34	0	0	0	170
2	6	27	0	717	1	3	0	66	0	0	0	61
3	5	24	0	986	1	3	0	31	0	0	0	139
Nutrition for health												
1	5	17	0	545	0	4	0	569	0	0	0	56
2	4	12	0	944	1	3	0	114	0	0	0	79

rather than being tied to any given week's content. These activities were separated into an additional factor. For the *Activity* model, each activity was assigned a factor based on whether it was a Lecture, Quiz, or Forum post.

In the EFA, the number of factors was determined by inspecting the scree plot and by considering the per cent of variance explained by each factor. Only the factors that accounted for at least 5 per cent of the variance were retained.

Results

Hypothesis testing results

To evaluate model fit in the CFAs, we used the comparative fit index (CFI; larger values indicate a better fit). The CFI measures the explanatory power of the model with respect to the underlying data (Hu and Bentler, 1999). We did not expect the theoretical models to provide a very strong fit to data, as they were relatively simple models; rather, the CFI provides a measure of relative fit (i.e. the extent to which each model pattern is found in the data). Figure 2 shows the CFI values across the CFA models on each offering of the included courses. Several important patterns are clear. First, a single overall participation model is not the best fit or even close to the best fit for any course offering. Second, the best fitting model is completely consistent across offerings within a course. Third, different courses have different best fitting models. In two of the courses, Nutrition for Health and Disaster Preparedness, the *Activity* model outperforms the *Weekly* model. In Clinical Terminology, however, the reverse is true. Fourth, the relative fits of the *Activity* and *Weekly* models are often close, and neither is a strong fit to the data (i.e. above 0.9) on its own. Therefore, it may be that a combination of the two models either best fits the course offering overall or captures different subsets of participants.

Emergent factors

EFAs investigate whether a completely different model (i.e. not based in weekly structure of activity type) provides a better fit. The best fitting models within each course were sometimes a hybrid of weekly and activity type (Figure 3). Most commonly, the emergent factors were largely separated by type of activity: forum use, lecture viewing and quiz submissions. Within Disaster Preparedness and Nutrition for Health, there was also a single early activity factor which covered the first two weeks. In Clinical Terminology, however, there were additional factors for early, middle, and later weeks, as well as for the two different types of quizzes: formative (those that are intended as practice) and summative (those that contribute to student grades). Thus, the models that emerged from the EFAs were only slight elaborations to the simple models tested in the CFAs.

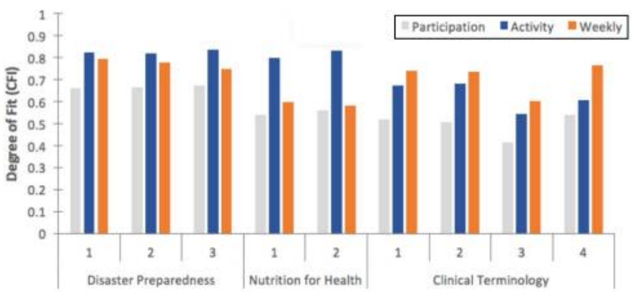

Figure 2. Comparative Fit Index values from the confirmatory factor analyses with overall participation, activity-based, or weekly-based models within each course offering (larger values indicate a better fit)

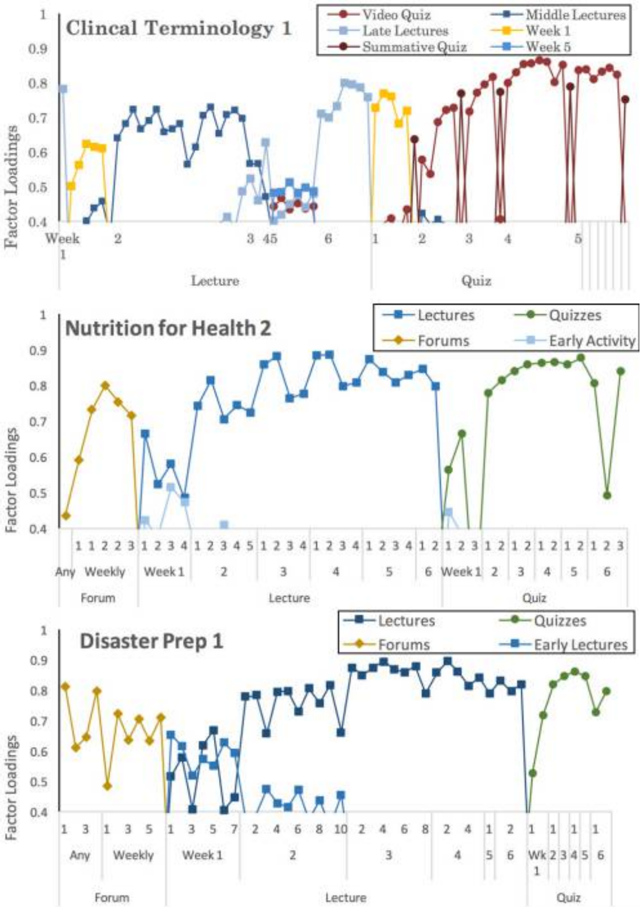

Figure 3.
Significant factor loadings for each activity in one offering of each course

Discussion

The consistency of the early activity factor structure lends some support to the hypothesis that MOOC participation is partially described by a relatively rapid initial dropout phenomenon described by time. This phenomenon is widely reported (Rosé *et al.*, 2014; Yang *et al.*, 2013; Jordan, 2014), and this type of student behavior has even been used to begin developing early predictors of student withdrawal from the course (Sinha *et al.*, 2014) or the next resource a student should use. However, all three courses also had some separation between types of activities in the EFA analyses suggestive of activity-based participation in addition to week-based participation. This type of activity-based participation has also been reported (Kizilcec *et al.*, 2013). But the combined presence of time-based and activity-based behavioral factors has received little if any attention.

Student behaviors in the studied courses build on our understanding of how students access course resources. This increased understanding comes from the interesting differences in activity patterns that were observed across these courses. Clinical Terminology consistently displayed a strong weekly (topical) browsing pattern. This time-based pattern continued beyond the first two weeks of temporal behavior observed in other courses and brings into question the generalizability of some of the activity-based stereotypes (Kizilcec *et al.*, 2013) and non-linear sequence-based activity patterns (Guo and Reinecke, 2014) that have been reported. Beyond these initial two weeks, we see further distribution of lectures between Weeks 2-4 and Weeks 5 and 6. Conversely, Nutrition for Health consistently displayed an activity-driven pattern that was similar to that reported by Kizilcec *et al.* (2013). Nevertheless, the current analysis allowed students to exhibit behaviors from multiple activity types, which was not previously done. Disaster preparedness also had a majority of students following the activity-based analyses, but some subsets of learners displayed a topic shopping behavior, which confirms the need to support non-linear navigation, as recommended by Guo and Reinecke (2014), in some learning contexts.

Considering these varying patterns in learner behavior across courses, the structure and design of each course may have impacted how learners engage with the content. The Clinical Terminology syllabus was designed around modular components, with few connections between them. This modular structure would have made it easier to meaningfully participate in later weeks' activities without first participating in earlier weeks' activities. By contrast, Nutrition for Health, which displays the weakest connection to the weekly topic-shopping paradigm, has sequences of module content that are often not contained within a single week. In Disaster Preparedness, several later modules build upon the information in prior modules. However, there are some modules that are more isolated. Thus, the hybrid of weekly and activity-based participation could be explained by this mixed modular structure.

Caveats

More courses need to be examined to see what design factors influence participation modes more generally. Most useful would be to examine two different course designs for the same content and similar learner populations, although this is unlikely to be found with a close match in both content and learner populations.

Implications for research

We have provided new characterizations of participation in MOOCs and online learning environments. Some of our identified factors (forum use, lecture viewing and quiz submissions) are consistent with prior research that used simplifying personae-based approaches. Others are not. These newly found characterizations indicate that existing personae-based analyses are inadequate because they fail to capture the more nuanced modes of participation that are evident within and across online courses. In contrast, our more nuanced approach allows students to exhibit multiple behavioral patterns to varying degrees. Future research can now take up the issue of what factors within courses and within learners leads to these different modes of participation.

Implications for practice

Given that learners choose to preferentially use specific features (discussion, quizzes and lectures) rather than consistently use all features, MOOC activities will need to provide multiple pathways to achieving the same goal. One avenue for achieving this system-design objective may be the integration of external tools that meet user needs. In some cases, learners are already initiating this process by creating their own Facebook groups to support MOOC learning (Kasunic *et al.*, 2016). MOOC designers can now take inspiration from these learners and use the above results to better understand how users engage in MOOCs. This understanding can then be used to plan and test appropriate pathways through MOOCs and enable better experiences for all learners, thus addressing calls for improved pedagogy within MOOCs (Ferguson *et al.*, 2016).

References

Anderson, A., Huttenlocher, D., Kleinberg, J. and Leskovec, J. (2014), "Engaging with massive online courses", *Proceedings of the 23rd International Conference on World Wide Web, ACM, New York, NY*, pp. 687-698, available at: https://doi.org/10.1145/2566486.2568042

Baikadi, A., Schunn, C.D., Long, Y. and Demmans Epp, C. (2016), "Redefining 'what' in analyses of who does what in MOOCs", *9th International Conference on Educational Data Mining (EDM 2016), International Educational Data Mining Society (IEDMS), Raleigh, NC*, pp. 569-570.

Breslow, L.B., Pritchard, D.E., DeBoer, J., Stump, G.S., Ho, A.D. and Seaton, D.T. (2013), "Studying learning in the worldwide classroom: research into edX's first MOOC", *Research & Practice in Assessment*, Vol. 8, pp. 13-25.

Brooks, C., Demmans Epp, C., Logan, G. and Greer, J. (2011), "The who, what, when, and why of lecture capture", *Learning Analytics and Knowledge*, ACM Press, Banff, pp. 86-92, available at: https://doi.org/10.1145/2090116.2090128

Coffrin, C., Corrin, L., de Barba, P. and Kennedy, G. (2014), "Visualizing patterns of student engagement and performance in MOOCs", *Learning Analytics and Knowledge*, ACM, New York, NY, pp. 83-92, available at: https://doi.org/10.1145/2567574.2567586

Corrin, L., de Barba, P.G. and Bakharia, A. (2017), "Using learning analytics to explore help-seeking learner profiles in MOOCs", *Proceedings of the Seventh International Learning Analytics & Knowledge Conference, ACM, New York, NY*, pp. 424-428, available at: https://doi.org/10.1145/3027385.3027448

Davis, D., Jivet, I., Kizilcec, R.F., Chen, G., Hauff, C. and Houben, G.-J. (2017), "Follow the successful crowd: raising MOOC completion rates through social comparison at scale", *Proceedings of the Seventh International Learning Analytics & Knowledge Conference, ACM, New York, NY*, pp. 454-463, available at: https://doi.org/10.1145/3027385.3027411

Dawson, S., McWilliam, E. and Tan, J.P.-L. (2008), "Teaching smarter: how mining ICT data can inform and improve learning and teaching practice", *Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education, Melbourne*, pp. 221-230.

de Barba, P.G., Kennedy, G.E. and Ainley, M.D. (2016), "The role of students' motivation and participation in predicting performance in a MOOC: motivation and participation in MOOCs", *Journal of Computer Assisted Learning*, Vol. 32 No. 3, pp. 218-231, available at: https://doi.org/10.1111/jcal.12130

Ferguson, R., Scanlon, E. and Harris, L. (2016), "Developing a strategic approach to MOOCs", *Journal of Interactive Media in Education*, Vol. 2016 No. 1, pp. 21, available at: https://doi.org/10.5334/jime.439

Gasevic, D., Kovanovic, V., Joksimovic, S. and Siemens, G. (2014), "Where is research on massive open online courses headed? A data analysis of the MOOC research initiative", *The International Review of Research in Open and Distributed Learning*, Vol. 15 No. 5, pp. 134-176, available at: www.irrodl.org/index.php/irrodl/article/view/1954

Guo, P.J. and Reinecke, K. (2014), "Demographic differences in how students navigate through MOOCs", *Proceedings of the First ACM Conference on Learning @ Scale Conference*, *ACM, New York, NY*, pp. 21-30, available at: https://doi.org/10.1145/2556325.2566247

Hu, L. and Bentler, P.M. (1999), "Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 6 No. 1, pp. 1-55, available at: https://doi.org/10.1080/10705519909540118

Jordan, K. (2014), "Initial trends in enrolment and completion of massive open online courses", *The International Review of Research in Open and Distributed Learning*, Vol. 15 No. 1, pp. 133-160, available at: http://dx.doi.org/10.19173/irrodl.v15i1.1651

Kasunic, A., Hammer, J., Kraut, R., Massimi, M. and Ogan, A. (2016), "A preliminary look at MOOC-associated facebook groups: prevalence, geographic representation, and homophily", *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, *ACM*, *New York, NY*, pp. 205-208, available at: https://doi.org/10.1145/2876034.2893415

Kenny, A. (2002), "Online learning: enhancing nurse education?", *Journal of Advanced Nursing*, Vol. 38 No. 2, pp. 127-135.

Kizilcec, R.F. and Halawa, S. (2015), "Attrition and achievement gaps in online learning", *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, *ACM*, *New York, NY*, pp. 57-66, available at: https://doi.org/10.1145/2724660.2724680

Kizilcec, R.F., Pérez-Sanagustín, M. and Maldonado, J.J. (2016), "Recommending self-regulated learning strategies does not improve performance in a MOOC", *Third Annual ACM Conference on Learning at Scale (L@S)*, *ACM Press*, *Banff*, pp. 101-104.

Kizilcec, R.F., Piech, C. and Schneider, E. (2013), "Deconstructing disengagement: analyzing learner subpopulations in massive open online courses", *Learning Analytics and Knowledge (LAK)*, ACM, New York, NY, pp. 170-179, available at: https://doi.org/10.1145/2460296.2460330

Koedinger, K.R., Kim, J., Jia, J.Z., McLaughlin, E.A. and Bier, N.L. (2015), "Learning is not a spectator sport: doing is better than watching for learning from a MOOC", *Learning @ Scale*, ACM Press, Banff, pp. 111-120, available at: https://doi.org/10.1145/2724660.2724681

Kovanović, V., Gašević, D., Joksimović, S., Siemens, G. and Hatala, M. (2015), "MOOCs in the news: a european perspective", *Proceedings of "WOW! Europe embraces MOOCs*, *Rome*, available at: www.researchgate.net/publication/293172665

Kovanović, V., Joksimović, S., Gašević, D., Owers, J., Scott, A.-M. and Woodgate, A. (2016), "Profiling MOOC course returners: how does student behavior change between two course enrollments?", *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, *ACM*, *New York, NY*, pp. 269-272, available at: https://doi.org/10.1145/2876034.2893431

Leontyev, A. and Baranov, D. (2013), "Massive open online courses in chemistry: a comparative overview of platforms and features", *Journal of Chemical Education*, Vol. 90 No. 11, pp. 1533-1539, available at: https://doi.org/10.1021/ed400283x

Liu, Z., Brown, R., Lynch, C.F., Barnes, T., Baker, R., Bergner, Y. and McNamara, D.S. (2016), "MOOC learner behaviors by country and culture; an exploratory analysis", *Educational Data Mining (EDM)*, International Edcuational Data Mining Society, Raleigh, NC, pp. 127-134.

Marshall, F., Jiang, W. and Dennen, V. (2015), "An examination of the usefulness of a learning community system and blackboard in both online and face-to-face courses", *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pp. 1136-1141.

Nelson, K. (2015), "Using K-means clustering to model students' LMS particiation in traditional courses", *Issues in Information Systems*, Vol. 16 No. 4, pp. 102-110.

Pardos, Z.A., Tang, S., Davis, D. and Le, C.V. (2017), "Enabling real-time adaptivity in MOOCs with a personalized next-step recommendation framework", *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, *ACM*, *New York, NY*, pp. 23-32, available at: https://doi.org/10.1145/3051457.3051471

Park, J.-H. and Choi, H.J. (2009), "Factors influencing adult learners' decision to drop out or persist in online learning", *Educational Technology & Society*, Vol. 12 No. 4, pp. 207-217.

Phirangee, K., Demmans Epp, C. and Hewitt, J. (2016), "Exploring the relationships between facilitation methods, students' sense of community and their online behaviours", *Online Learning*, Vol. 20 No. 2, pp. 134-154, available at: http://dx.doi.org/10.24059/olj.v20i2.775

Pursel, B.K., Zhang, L., Jablokow, K.W., Choi, G.W. and Velegol, D. (2016), "Understanding MOOC students: motivations and behaviours indicative of MOOC completion: MOOC student motivations and behaviors", *Journal of Computer Assisted Learning*, Vol. 32 No. 3, pp. 202-217, available at: https://doi.org/10.1111/jcal.12131

Ramesh, A., Goldwasser, D., Huang, B., Iii, H.D. and Getoor, L. (2014), "Learning latent engagement patterns of students in online courses", *Twenty-Eighth AAAI Conference on Artificial*

Intelligence, pp. 1272-1278, available at: www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8571

Rosé, C.P. and Ferschke, O. (2016), "Technology support for discussion based learning: from computer supported collaborative learning to the future of massive open online courses", *International Journal of Artificial Intelligence in Education*, Vol. 26 No. 2, pp. 660-678, available at: https://doi.org/10.1007/s40593-016-0107-y

Rosé, C.P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P. and Sherer, J. (2014), "Social factors that contribute to attrition in MOOCs", *Proceedings of the First ACM Conference on Learning @ Scale Conference*, *ACM*, *New York, NY*, pp. 197-198, available at: https://doi.org/10.1145/2556325.2567879

Rovai, A.P. and Wighting, M.J. (2005), "Feelings of alienation and community among higher education students in a virtual classroom", *The Internet and Higher Education*, Vol. 8 No. 2, pp. 97-110, available at: https://doi.org/10.1016/j.iheduc.2005.03.001

Ruiperez-Valiente, J.A., Alexandron, G., Chen, Z. and Pritchard, D.E. (2016), "Using multiple accounts for harvesting solutions in MOOCs", *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, *ACM*, *New York, NY*, pp. 63-70, available at: https://doi.org/10.1145/2876034.2876037

Selwyn, N. (2014), *Distrusting Educational Technology: critical Questions for Changing Times*, Routledge, Taylor & Francis Group, New York, London.

Siemens, G. (2013), "Massive open online courses: innovation in education?", in McGreal, R. Kinuthia, W. and Marshall, S. (Eds), *Open Educational Resources: innovation, Research and Practice*, Common wealth of Learning (COL), Vancouver, pp. 5-16, available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.403.7881&rep=rep1&type=pdf#page=31

Sinha, T., Li, N., Jermann, P. and Dillenbourg, P. (2014), "Capturing 'attrition intensifying' structural traits from didactic interaction sequences of MOOC learners", *Proceedings of the EMNLP 2014 Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses*, pp. 42-49.

Swan, K., Day, S. and Bogle, L. (2016), "Metaphors for learning and MOOC pedagogies", *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, *ACM*, *New York, NY*, pp. 125-128, available at: https://doi.org/10.1145/2876034.2893385

Van der Sluis, F., Ginn, J.V. and Zee, T. (2016), "Explaining student behavior at scale: the influence of video complexity on student dwelling time", *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, *ACM*, *New York, NY*, pp. 51-60, available at: https://doi.org/10.1145/2876034.2876051

Wen, M., Yang, D. and Penstein-Rosé, C. (2014), "Sentiment analysis in MOOC discussion forums: what does it tell us?", Presented at the Educational Data Mining, San Francisco, CA, pp. 130-137, available at: http://educationaldatamining.org/conferences/index.php/EDM/2014/paper/view/1360/1326

Wilcox, D., Thall, J. and Griffin, O. (2016), "One canvas, two audiences: how faculty and students use a newly adopted learning management system", *Society for Information Technology & Teacher Education International Conference*, Association for the Advancement of Computing in Education (AACE), Chesapeake, VA, pp. 1163-1168.

Yang, D., Sinha, T., Adamson, D. and Rosé, C.P. (2013), "Turn on, tune in, drop out: anticipating student dropouts in massive open online courses", *Proceedings of the 2013 NIPS Data-driven Education Workshop*, Vol. 11, p. 14.

Corresponding author
Christian D. Schunn can be contacted at: schunn@pitt.edu