Redefining “What” in Analyses of Who Does What in MOOCs

Alok Baikadi[1,2], Christian D. Schunn[1], Yanjin Long[1,2], Carrie Demmans Epp[1,2]
[1]Learning Research and Development Center
[2]Center for Instructional Development and Distance Education
University of Pittsburgh
{baikadi, schunn, ylong, cdemmans}@pitt.edu

ABSTRACT

To advance our understanding of learning in massive open online courses (MOOCs), we need to understand how learners interact with course resources. Prior explorations of learner interactions with MOOC materials have often described these interactions through stereotypes, which does not account for the full spectrum of potential learner activities. A focus on stereotypes also limits our ability to explore the reasons behind learner behaviors. To overcome these shortcomings, we apply factor analysis to learner activities within four MOOCs to identify emergent behavior factors. The factors support characterizations of learner behaviors as driven heavily by types of learning activities and secondarily by time/topic; regression revealed demographic factors (especially country and gender) associated with these activity and topic preferences. Both factor and regression analyses revealed structural variability in learner activity patterns across MOOCs. The results call for a reconceptualization of how different learning activities within a MOOC are designed to work together.

Keywords

MOOCs; learning analytics; online learning; factor analysis

1. INTRODUCTION

With the increasing popularity of massive open online courses (MOOCs), the need to investigate the relationships among learner characteristics, learner-selected activities, and learning outcomes has become critical. Determining these relationships can help us understand how people learn within MOOCs and inform MOOC design and pedagogy. Prior work identified different learning-activity patterns [1, 3] and investigated the relationship between certain types of learning activities and outcomes [7]. Many of these studies were conducted in the context of a single domain or MOOC (e.g., [1]). Furthermore, little work has investigated how demographic variability could lead to different behavioral patterns in MOOCs, leaving an open question: Can the identified patterns be generalized across instructional domains and populations?

Until recently, studies of learning within MOOCs focused more on the number of learners being served than pedagogy [6]. This focus on their size has left many facets of MOOCs underexplored and poorly understood [1]. These aspects include a need to understand how learners engage with MOOCs [1], their behavior patterns, and their motivations [3]. Understanding these factors may allow us to design courses that support the learning activities and outcomes that learners want.

We investigate learning patterns in four MOOCs based on learner activities across courses from different disciplines. We used the activity-centered data reduction technique of factor analysis to identify the underlying course activities that describe learner activity patterns within each offering of the selected MOOCs. The factor analyses applied to 10 MOOC offerings enabled us to identify 1) factors that are common to most of these MOOCs and 2) factors that are less common. Regression analyses were then used to examine the relationship between learner demographic variables and their participation on each factor. These analyses support the distinctions between factors and the presence of varied factors across MOOCs.

This investigation is among the first to identify and compare activity patterns and demographic influences across learning domains. The results improve our understanding of learner behaviors across contexts and could inform the design of more domain-sensitive learning experiences.

2. LEARNER ACTIVITIES IN MOOCS

Research into MOOCs has spanned a range of topics, with recent discussions becoming more nuanced. Work that has investigated how learners interact with a MOOC [5] found that their behaviors can be characterized through a set of trajectories rather than the commonly used completion and attrition model. These trajectories through graded assignments and lecture videos within computer science MOOCs characterize how different types of learners used some of the course materials to support their learning activities [1]. The identified usage patterns included those who mostly watched lectures, mostly submitted assignments, performed some combination of these activities, downloaded course resources, or registered but did very little.

Some researchers have taken the next step by linking these types of activities (watching video lectures, submitting assignments, and discussion forum activity, types of questions asked) to course performance (certificate earned, learning outcomes and gains, course completion) [2, 7]. To obtain a better understanding of how these and other factors influence learner success within MOOCs, the relationships among socio-demographic variables, student activities, and learner success have been explored. The most common predictors of certificate earning and completion were prior education [2], sex [4], and country of origin [4].

3. MOOC CORPUS

Data from the 132,324 learners who performed at least one action (taking a quiz, posting to the forum, or watching a video) in 4 of the University of Pittsburgh's Coursera MOOCs were used. To describe learner activities within a range of course types and explore generalizability across disciplines, courses from different domains were chosen: health sciences (nutrition for health and clinical terminology), education (accountable talk), and public health (disaster preparedness). Data from multiple offerings (Jan. 2013 – Dec. 2015) of the same course were used when available.

The courses lasted 6 or 7 weeks. The core materials for each week consisted of video lectures and a quiz. Some weeks included assignments, disaster preparedness used peer-assessment, and accountable talk had a project. Clinical terminology incorporated multimedia modules that enabled the learner to interact with learning resources. Since these modules presented core content, they were labeled as lectures. Only the Clinical Terminology instructors explicitly encouraged discussion forum use and provided study tips. This variability provided a cross-section of course formats that enables us to identify learner activities that apply across courses and that are specific to a course. We used the activity counts for each forum, quiz, and lecture video.

4. RESULTS

4.1 Learner Activities

Factor analysis with varimax rotation was used to reduce the dimensionality of the data and identify learners' underlying behavioral tendencies. Course activities that at least 1% of active learners performed were used. To test the stability of the patterns, a separate factor analysis was conducted for each course offering. Factors that accounted for at least 5% of the variance were kept.

In 3 of the 4 courses, activities were largely grouped into 4 factors: lecture activity, quiz activity, forum participation and participation in activities from weeks 1 and 2. In contrast, clinical terminology shows more depth in weekly content: lecture activity is represented by 4 factors, each capturing a 1-2 week span. For quizzes, we see three factors: summative quizzes presented at the end of each module, early quiz activities, and later quiz activities.

4.2 Predicting Activities Using Demographics

We calculated a factor score for each learner, which indicates a tendency towards the behavior described by that factor. For example, a learner with a high score for the lectures factor would have viewed more lectures than one with a low score. A general linear model (GLM) was used to predict learner factor scores from learners' socio-demographic characteristics. Only those (n = 2963) with individual demographic profiles were included. We applied GLM to courses that had contrastive factor structures: the second offering of nutrition for health represented those with media-based factors and the first offering of clinical terminology represented those with time-based factors.

For clinical terminology, we aggregated early lecture factors, late lecture factors, and quiz factors to create factors that were comparable to the other courses. We then ran a generalized linear model predicting each of these aggregated factors.

Each factor is influenced differently by learner demographics and are contrasted between the two courses. For example, the early lecture watching factor from nutrition for health was more strongly associated with female learners than males. This was not the case for clinical terminology. Late lecture watching activity was predicted by learner age for both courses. However, a difference in factor scores for the younger and older populations for those in the middle age groupings is visible between the courses. Within clinical terminology, we also see that some age groups are more active earlier in the course than later. Additional differences in how demographic variables predict factors are visible when considering learners' quiz participation and their continent of residence. Similar factor scores are seen for those who live in Asia and North America when considering learner activities within clinical terminology. This similarity does not hold across courses; learners from Asia and Europe appear to be more similar in their quiz taking habits when considering the data from nutrition for health.

5. CONCLUSION

Our factor and regression analyses across multiple offerings of the same course show that learner behaviors are relatively consistent across time. However, differences in factors across courses suggest that design and domain affect how learners select learning content and activities, which requires further study.

Our work is among the first applications of exploratory factor analyses across learner activities within MOOCs from different domains. Prior work has focused on a person-based approach that describes the behavior patterns of individuals by assigning them to canonical groups. This work, therefore, provides a new lens to examine the full range of learner behaviors in MOOCs.

6. REFERENCES

[1] Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J. 2014. Engaging with Massive Online Courses. In *Proceedings of the 23rd International Conference on World Wide Web*, ACM, 687–698. DOI = http://doi.org/10.1145/2566486.2568042

[2] 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* 8: 13–25.

[3] 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* 15, 5.

[4] Kizilcec, R.F. and Halawa, S. 2015. Attrition and Achievement Gaps in Online Learning. In *Proceedings of the 2nd ACM Conference on Learning @ Scale*, ACM, 57–66. DOI = http://doi.org/10.1145/2724660.2724680

[5] Kizilcec, R.F., Piech, C., and Schneider, E. 2013. Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (LAK)*, ACM, 170–179. DOI = http://doi.org/10.1145/2460296.2460330

[6] Kovanović, V., Gašević, D., Joksimović, S., Siemens, G., and Hatala, M. 2015. MOOCs in the News: A European Perspective. In *Proceedings of "WOW! Europe embraces MOOCs."*

[7] Wang, X., Yang, D., Wen, M., Koedinger, K., Rosé, C. P. (2015). Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM)*.