

# Using learning analytics to explore help-seeking learner profiles in MOOCs

Linda Corrin  
University of Melbourne  
Level 6, 111 Barry Street  
Melbourne, VIC 3010, Australia  
lcorrin@unimelb.edu.au

Paula G. de Barba  
University of Melbourne  
Elisabeth Murdoch Building  
Melbourne, VIC 3010, Australia  
paula.de@unimelb.edu.au

Aneesha Bakharia  
University of Queensland  
Learning Innovation Building  
St Lucia, QLD 4072, Australia  
a.bakharia1@uq.edu.au

## ABSTRACT

In online learning environments, learners are often required to be more autonomous in their approach to learning. In scaled online learning environments, like Massive Open Online Courses (MOOCs), there are differences in the ability of learners to access teachers and peers to get help with their study than in more traditional educational environments. This exploratory study examines the help-seeking behaviour of learners across several MOOCs with different audiences and designs. Learning analytics techniques (e.g., dimension reduction with t-sne and clustering with affinity propagation) were applied to identify clusters and determine profiles of learners on the basis of their help-seeking behaviours. Five help-seeking learner profiles were identified which provide an insight into how learners' help-seeking behaviour relates to performance. The development of a more in-depth understanding of how learners seek help in large online learning environments is important to inform the way support for learners can be incorporated into the design and facilitation of online courses delivered at scale.

## CCS CONCEPTS

•Applied computing → E-learning;

## KEYWORDS

Learning analytics, MOOCs, Help-seeking, Learning design

## ACM Reference format:

Linda Corrin, Paula G. de Barba, and Aneesha Bakharia. 2016. Using learning analytics to explore help-seeking learner profiles in MOOCs. In *Proceedings of LAK '17, Vancouver, BC, Canada, March 13-17, 2017*, 5 pages. DOI: <http://dx.doi.org/10.1145/3027385.3027448>

## 1 INTRODUCTION

In online learning environments, learners are required to adopt a high level of autonomy to successfully complete their studies. In order to ensure that the design of online courses provides appropriate support for learners in such autonomous settings it is necessary to understand how and when learners seek help for their learning. The ability to seek help to resolve confusion and misunderstanding

is especially important in Massive Open Online Courses (MOOCs) where learners' access to teachers is limited due to the large number of learners and course activity. At the same time, many MOOCs provide opportunities for learners to seek help from their peers through tools such as discussion forums. What is still unclear is the extent to which this is taken up by learners for the purpose of getting help, and how this relates to other activities learners can undertake to get support for their learning. While learning analytics have been used extensively to explore learner behaviour patterns in MOOCs in relation to student performance and persistence, fewer studies have focused specifically on the activities learners undertake in relation to the self-regulatory learning action of help-seeking.

In this paper we present the initial results of an exploratory study that uses learning analytics techniques to explore the profiles of learners' help-seeking behaviour in MOOCs. This paper makes two important contributions to the learning analytics literature: firstly, by comparing the profiles of help-seeking behaviour across multiple MOOCs with different purposes (e.g., professional development vs. general interest) and designs; and secondly, by employing analysis techniques previously rarely used in learning analytics studies (e.g., t-sne and Affinity Propagation) for clustering and to enable better interpretation of profile groups. The outcomes of this study can help to identify ways that learners' help-seeking can be supported and inform future investigation of specific elements of help-seeking behaviours in MOOCs.

## 2 CONTEXT

Massive Open Online Courses (MOOCs) have become firmly cemented within the higher education landscape over the past eight years. The popularity of MOOCs, both in terms of learner demand for courses and the increasing involvement of universities in delivering them, provides opportunities for developing a greater understanding of how people learn in online environments. The scale of these courses allows unprecedented access to large sets of data on learner behaviour for researchers to explore educational issues including engagement [7], performance [6], and retention [1]. The application of learning analytics techniques to these large data sets offers the potential to uncover patterns and profiles of learner behaviour that can inform better learning support, feedback and course design.

In developing a better understanding of learning in online environments, researchers have used many different theoretical lens through which to explore this phenomenon. The approach adopted in this study relies on the lens of self-regulated learning. Self-regulated learners are active learners who plan, monitor, adapt and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
LAK '17, Vancouver, BC, Canada

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
978-1-4503-4870-6/17/03...\$15.00  
DOI: <http://dx.doi.org/10.1145/3027385.3027448>

reflect on their learning experience [16]. They regulate their emotions, cognition, context and behaviour to achieve specific learning goals. Learners who display high levels of self-regulation have been associated with higher academic performance and persistence [4][10]. The nature of online learning environments often requires greater levels of autonomy from learners than more traditional settings. As a result, learners need to be better at self-regulating their learning to succeed [2].

A key element of self-regulated learning is the help-seeking behaviours learners adopt to support their learning processes. Help-seeking is the ability to seek out assistance from peers and teachers when learners encounter a challenge they find too complex to solve [12]. Once they have identified the need for assistance, learners have to decide the best way to seek help in the given context.

This help-seeking process usually includes eight stages in which learners: determine whether there is a problem, determine whether help is needed, decide whether to seek help, decide on the type of help, decide on whom to ask, solicit help, obtain help, and process help received [12]. Learners can undertake information searches by consulting resources, such as a textbook, or have the option of help-seeking by consulting a human mediator, such as a peer. In online learning environments the concepts of information search and help-seeking overlap, as learners have the option to search for information that has been curated by their peers and teachers [17].

There are several ways that learners can seek help within a MOOC environment. The most common is via the discussion forum where learners can post problems they are encountering to which peers and/or teachers can reply. The active use of discussion forums in MOOCs has been correlated to higher levels of performance in MOOCs in some studies (e.g., [19]). In relation to help-seeking behaviours, a study by [9] found that clusters of different help-seeking behaviours had different course completion outcomes, with learners who engaged primarily with sub-forums dedicated exclusively to discussion of the final assignment more likely to fail the course. Other more passive forms of help-seeking behaviour in discussion forums include viewing and searching for posts on topics of interest. In addition, learners can seek help by pausing and searching for specific content in the instructional videos of the course.

Tools to support student help-seeking behaviour have been developed and implemented in MOOCs. For example, the Data, Analytics and Learning MOOC (DALMOOC) incorporated a "Quick Helper" tool designed to actively connect learners together to help to ensure learners who seek help through discussion forums receive a reply to their query [18].

Research on help-seeking in MOOCs has, to date, focused on students' self-report data to measure how they actively ask other people for help [11] [13]. Alternatively, the research study presented in this paper explores learners' help-seeking behaviours in MOOCs through the use of learning analytics techniques to create learner profiles. Such behaviours include not only asking others for help, but also indirect help-seeking behaviours, such as searching, subscribing and viewing forums or pausing and seeking specific locations in video. Recent research indicates the use of learning analytics (rather than self-report data) is a powerful method to investigate students' use of self-regulated learning skills [5]. Recent research also highlights the need for more studies to explore the instructional conditions, or learning designs, of online courses in

relation to learner usage profiles [15] [3]. Therefore this research was driven by the following research questions:

- (1) What learner profiles emerge from help-seeking behaviours displayed in MOOC environments?
- (2) What effect does the learning design have on different help-seeking behaviours in MOOCs in relation to learner performance?

### 3 METHOD

Four MOOCs were used in the analysis of learners' help-seeking behaviour in this study. These MOOCs were all developed by the University of Melbourne and delivered through the Coursera platform. Two of the MOOCs, *Discrete Optimization* and *Assessment and Teaching of 21st Century Skills*, represent MOOCs with a strong emphasis on professional development. Alternatively, the two other courses, *The French Revolution* and *Animal Behaviour*, were aimed more at a general interest audience. The cohorts of learners included in the study were taken from the 2015 offerings of each MOOC. The main design features of each MOOC are outlined in Table 1.

The sample included learners who achieved a grade of more than 40% for the course. Unlike other studies of engagement profiles in MOOCs that include all learners to be able to determine persistence, this study focused on the behaviour of those learners who participated in assessments throughout the course as a way to look at help-seeking behaviour. The sample size of learners for each MOOC used in the study was: *Discrete Optimization* (301), *Assessment and teaching of 21st century skills* (655), *The French Revolution* (825), and *Animal Behaviour* (658).

Data for each of the MOOCs was extracted from the Coursera course and clickstream databases. A feature matrix was then created which included 25 features which could represent help-seeking behaviours including active forum participation (i.e., the creation of forum posts, the creation of forum comments, forum thread subscription, forum post up voting, forum post down voting, forum tag creation and forum reputation points), passive forum participation (i.e., forum post views, forum comment views and thread views), assessment (i.e., assignment, quiz and survey attempts, and the number of times a quiz result screen was viewed), search (i.e., search queries in discussion forums and search queries for overall course), video interaction features (i.e., total videos accessed, number of times the play event was triggered, number of times the pause event was triggered, number of times the playback speed of a video was changed, number of times video seeking events were triggered and total video event interactions), total clickstream views, grade, whether the learner was enrolled in a signature track and notification subscription (i.e, email announcement subscription).

The feature matrix was scaled using min-max feature scaling prior to the performance of dimension reduction using t-distributed Stochastic Neighbor Embedding (t-SNE) [14]. The t-SNE dimension reduction technique provides a way to embed high-dimensional data (i.e., the learner feature matrix) into a lower dimensional space which can then be visualized in a scatter plot. The affinity propagation algorithm [8] was then used to find the number of clusters and simultaneously group similar learners within a cluster. Traditional clustering algorithms like k-means, require the number of clusters

**Table 1: Design features of MOOCs**

| MOOC   | Description   | Duration | Structure  | Discussion integration  |
|--|---|----------|------------|---|
| Animal Behaviour                               | An exploration of how scientists study animal behaviour and the evolutionary forces of natural and sexual selection | 8 weeks  | Sequential | No direct instruction to access the discussion forums, but forums established for each week |
| Assessment and teaching of 21st century skills | An introduction to how 21st century skills can be taught and assessed in education                                  | 6 weeks  | Sequential | Discussion prompts incorporated into videos   |
| The French Revolution                          | An examination of the origins, course and outcomes of the French Revolution   | 6 weeks  | Sequential | Weekly discussion prompts   |
| Discrete Optimization                          | An introduction to how to solve complex search problems with discrete optimization concepts and algorithms          | 8 weeks  | Open       | No direct instruction to access the discussion forums                                       |

to be specified as a parameter and additional techniques are required to help determine the ideal number of clusters that exist in a dataset. Affinity propagation was used within this study because its ability to automatically determine the number of clusters is useful in aiding exploratory data analysis and learner cluster interpretation. Affinity propagation also returns an exemplar for each cluster which in the context of this study represented the learner with the features that are most indicative of the cluster. The analysis of exemplar students led to the initial development of the learner profiles which were then validated by inspection of other learners in the cluster and feature statistics for the cluster. The resulting 2-D t-SNE scatter plot for the French Revolution MOOC is displayed in Figure 1, with each cluster labeled (and coloured).

Once the clusters had been generated, the learner exemplar and the cluster statistics (i.e., minimum, maximum and standard deviation for each feature) were analysed. Multiple clusters of MOOC participants were assigned to a learner profile based on forum activity, assessment performance, search and video interaction.

## 4 RESULTS AND DISCUSSION

Five learner profiles were identified using the clusters generated through the affinity propagation technique. The learner profiles and numbers of clusters associated with each profile across the four MOOCs are presented in Table 2.

The profile of *low engagement* included learners who made very little use of the learning resources and discussion forums in the course. These learners had low levels of performance on the assessments, but didn't seem to employ many help-seeking behaviours. In terms of differences across the four MOOCs, the percentage of students who fell into this category from *The French Revolution* MOOC was higher than the other three courses. This could be attributed to the fact that this MOOC is designed as a general interest course which may impact the motivation of students to strive for high levels of engagement and performance.

The *assessment-focused - low grades* profile group included learners who had multiple attempts at the assessment, but still received lower grades. These learners didn't make use of the discussion forum to seek help or watch many videos. Despite receiving feedback from multiple quizzes attempts, the performance of these students remained low. The exception in this profile group was the *Discrete Optimization* course. The grading and feedback for the assessment

in this course was automated around an algorithm and this was explained in detail to the students [20]. This awareness of the meaning of the feedback could account for the fact that learners were able to improve their grades on multiple attempts.

The *passive engagement* learner profile group represented the majority of learners across each of the courses. These learners made moderate use of resources and forums resulting in a moderate grade for the course. While their use of the forums was higher than the two previous learner profiles, there was still a low level of forum subscription and search, which represent more specific help-seeking behaviours in this context. An overwhelming majority of learners in *Discrete Optimization* fell within this learner profile. A large proportion of *21st Century Skills* learners also fell within this profile which could be attributed to the fact that use of the discussion forums was integrated within the instructional videos, prompting learners to access and engage with the forums.

The *active engagement* profile includes learners who have high grades with a moderate number of assessment attempts. These students have high levels of engagement with resources and videos, and participate actively in discussion forums. Like the *passive engagement* group they engage substantially in the discussion forums, but don't make use of the subscription and search functions.

The *assessment-focused - high grades* profile group had multiple attempts at assessments, resulting in higher grades. However, they did not engage with the discussion forums and made limited use of resources and videos. Interestingly, this profile group was found to exist in *Animal Behaviour* and *21st Century Skills*, but not in the other two MOOCs. These learners don't appear to engage in many help-seeking behaviours, yet still manage to perform very well overall. One possible explanation is that these learners have a high level of prior knowledge that allows them to perform well in assessments using, almost exclusively, feedback provided through quizzes to refine their knowledge in the context of the course.

This exploratory analysis has provided insights into the help-seeking behaviours of groups of students across different MOOCs. While differences in MOOC design can possibly explain some differences in the proportions of help-seeking profiles, further exploration is needed to understand how these design factors have impacted learner behaviour in more depth. The analysis also uncovered several interesting observations in relation to help-seeking

**Table 2: Learner profiles based on help-seeking behaviour**

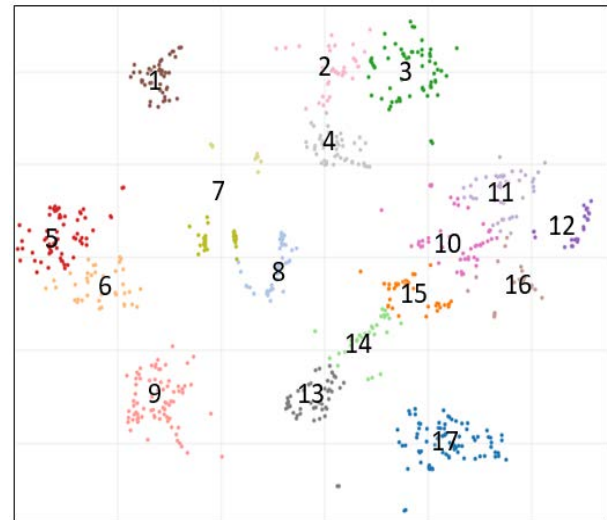
| Profile                          | Activity   | Forum Activity  | Animal Behaviour (n = 658) | 21st Century Skills (n = 655) | French Revolution (n = 825) | Discrete Optimization (n = 301) |
|----------------------------------|--|---|----------------------------|-------------------------------|-----------------------------|---------------------------------|
| Low engagement                   | Low grades, low pageviews, low video interaction, low, multiple grade attempts | No forum threads/post reads   | 19.3%                      | 12.0%                         | 34.6%                       | 14.6%                           |
| Assessment-focused - low grades  | Low grade, high assessment attempts, low pageviews, low video interaction      | No forum thread/post reads  | 15.8%                      | 23.3%                         | 26.6%                       | 0%                              |
| Passive engagement               | Medium grade, moderate pageviews   | Low forum thread subscription, low forum /course search, moderate forum thread/post reads | 28.3%                      | 34.6%                         | 24.8%                       | 67.1%                           |
| Active engagement                | High grades, moderate assessment attempts, high pageviews                      | High forum participation, low forum thread subscription, low forum/course search          | 10.7%                      | 10.8%                         | 14.0%                       | 18.3%                           |
| Assessment-focused - high grades | High grade, high assessment attempts, low pageviews, low video interaction     | No forum activity, low forum thread subscription, low forum/course search                 | 25.8%                      | 19.0%                         | 0%                          | 0%                              |

activities that could aid students, but are not commonly used, such as forum subscriptions and search features.

A benefit of using the affinity propagation technique over kmeans to identify the clusters of help-seeking learner behaviours is that it can generate learner exemplars. Further work on this data set will look at these exemplars using sequential and temporal analysis to explore these profiles in more detail. For example, examining the sequence of how learners access, pause, replay and seek specific locations in videos, which has been associated with achieving better grades. It is expected that focusing on these more specific actions can help to inform the development of more targeted support strategies and tools for learner help-seeking. The influence of learner demographics within each profile is another avenue of further research we plan to pursue.

## 5 CONCLUSION

This initial exploratory study has identified several interesting trends around the help-seeking behaviours of learners in MOOCs. Key help-seeking mechanisms within MOOC environments, such as discussion forums, still have low levels of engagement, with actions such as search and subscription remaining very low. However, it was clear that engagement with the discussion forums, whether active or passive, generally resulted in higher performance in the MOOCs overall. The fact that many students are either not or only passively engaging with discussion forums indicates that designers of MOOCs need to explore other ways that support can be provided to learners. It was also important to note the impact of different learning designs in the distribution of help-seeking profiles across the different MOOCs. For example, clearer explanations of assessment evaluation design (e.g., *Discrete Optimization*) and stronger integration of discussion tools into content delivery (*21st*



**Figure 1: Learner clusters found by the Affinity Propagation algorithm for the French Revolution MOOC.**

*Century Skills*) resulted in different patterns of behaviour in relation to activities that could be viewed as help-seeking.

Developing better strategies to assist learners with help-seeking in MOOCs is especially important in an environment where several MOOC providers are transitioning from cohort-based, timed delivery of courses to on-demand courses that learners can take when they want and work through at their own pace. This change to how and when learners commence and progress through the course has

implications for the ways that help can be sourced from peers. So if this contributes to the reduction in use or even disappearance of discussion forums as a key element of MOOCs, as predicted by some, what does this mean for help-seeking behaviour in MOOC environments?

These and many other questions about how help-seeking can be fostered and supported in MOOC environments remain. This short paper has presented the first steps in the context of a broader study of how learners seek support for their learning. The use of learning analytics offers potential to explore this phenomenon in new and different ways. The methods adopted in this study have helped to gain a high-level view of the patterns of help-seeking behaviour across a number of MOOC designs. Further research is needed to explore some of these specific behaviours within the MOOC context in more detail, including the use of sequential and temporal techniques. In working to build and refine these help-seeking learner profiles, it is also important to supplement the findings derived from learning analytics techniques, such as the ones used here, with reflections from learners about their help-seeking approaches. This will enable us to identify the strategies learners adopt to seek help that may fall between the gaps of tracked log events in MOOCs [21]. These insights can then be used to inform strategies for the development of tools and resources to provide greater assistance and support to learners, regardless of the scale of the online learning environment.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the support of the coordinators of the MOOCs included in this study: Professor Peter McPhee, Professor Raoul Mulder, Professor Patrick Griffin and Professor Pascal Van Hentenryck. A Special Research Initiative of the Australian Research Council supported this research: ARC-SRI Science of Learning Research Centre (project number SRI20300015).

## REFERENCES

- [1] Khaled M Alraimi, Hangjung Zo, and Andrew P Ciganek. 2015. Understanding the MOOCs continuance: The role of openness and reputation. *Computers & Education* 80 (2015), 28–38.
- [2] Anthony R Artino. 2007. Self-regulated learning in online education: A review of the empirical literature. *International Journal of Instructional Technology and Distance Learning* 4, 6 (2007), 3–18.
- [3] Aneesha Bakharia, Linda Corrin, Paula de Barba, Gregor Kennedy, Dragan Gašević, Raoul Mulder, David Williams, Shane Dawson, and Lori Lockyer. 2016. A conceptual framework linking learning design with learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. ACM, 329–338.
- [4] J Broadbent and WL Poon. 2015. Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education* 27 (2015), 1–13.
- [5] Moon-Heum Cho and Jin Soung Yoo. 2016. Exploring online students' self-regulated learning with self-reported surveys and log files: a data mining approach. *Interactive Learning Environments* (2016), 1–13.
- [6] Carleton Coffrin, Linda Corrin, Paula de Barba, and Gregor Kennedy. 2014. Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of the fourth international conference on learning analytics and knowledge*. ACM, 83–92.
- [7] Rebecca Ferguson and Doug Clow. 2015. Examining engagement: analysing learner subpopulations in massive open online courses (MOOCs). In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*. ACM, 51–58.
- [8] Brendan J Frey and Delbert Dueck. 2007. Clustering by passing messages between data points. *science* 315, 5814 (2007), 972–976.
- [9] Nabeel Gillani, Rebecca Eynon, Michael Osborne, Isis Hjorth, and Stephen Roberts. 2014. Communication communities in MOOCs. *arXiv preprint arXiv:1403.4640* (2014).
- [10] Carolyn Hart. 2012. Factors associated with student persistence in an online program of study: A review of the literature. *Journal of Interactive Online Learning* 11, 1 (2012), 19–42.
- [11] Nina Hood, Allison Littlejohn, and Colin Milligan. 2015. Context counts: How learners' contexts influence learning in a MOOC. *Computers & Education* 91 (2015), 83–91.
- [12] Stuart A Karabenick and Myron H Dembo. 2011. Understanding and facilitating self-regulated help seeking. *New Directions for Teaching and Learning* 2011, 126 (2011), 33–43.
- [13] René F Kizilcec, Mar Pérez-Sanagustín, and Jorge J Maldonado. 2017. Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education* 104 (2017), 18–33.
- [14] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, Nov (2008), 2579–2605.
- [15] Negin Mirriahi, Daniyal Liaqat, Shane Dawson, and Dragan Gašević. 2016. Uncovering student learning profiles with a video annotation tool: reflective learning with and without instructional norms. *Educational Technology Research and Development* 64, 6 (2016), 1083–1106.
- [16] Paul R Pintrich. 2000. *The role of goal orientation in self-regulated learning*. Academic Press.
- [17] Minna Puustinen and Jean-François Rouet. 2009. Learning with new technologies: Help seeking and information searching revisited. *Computers & Education* 53, 4 (2009), 1014–1019.
- [18] Carolyn Penstein Rosé, Oliver Fersckhe, Gaurav Tomar, Diyi Yang, Iris Howley, Vincent Alevén, George Siemens, Matthew Crosslin, Dragan Gasevic, and Ryan Baker. 2015. Challenges and opportunities of dual-layer MOOCs: Reflections from an edX deployment study. In *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning (CSCW 2015)*, Vol. 15. 848–851.
- [19] Shu-Fen Tseng, Yen-Wei Tsao, Liang-Chih Yu, Chien-Lung Chan, and K Robert Lai. 2016. Who will pass? Analyzing learner behaviors in MOOCs. *Research and Practice in Technology Enhanced Learning* 11, 1 (2016), 8.
- [20] Pascal Van Hentenryck and Carleton Coffrin. 2014. Teaching creative problem solving in a MOOC. In *Proceedings of the 45th ACM technical symposium on Computer science education*. ACM, 677–682.
- [21] George Veletsianos, Justin Reich, and Laura A Pasquini. 2016. The Life Between Big Data Log Events" Learners' Strategies to Overcome Challenges in MOOCs. *AERA Open* 2, 3 (2016), 1–10.