Attrition in Online Learning: Understanding Persistence and Dropout in Massive Open Online Courses

Omitted for blind review
Institution
Address
Email

ABSTRACT
Persistence and attrition in MOOCs are systematically analyzed using self-report and behavioral data collected from [N] online learners in [M] courses. Study 1 offers insights into reasons for disengaging from MOOCs and explores relationships with prior behavior and reported intentions. Study 2 is a case study to develop a deeper understanding of attrition in MOOCs by conducting a case study. Targeting online learners who were predicted likely dropouts in a particular course were invited to provide feedback via a survey.

Author Keywords
Online learning, persistence, attrition, dropout, disengagement, massive open online courses, MOOC, psychological factors

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; K.3.1. Computers and Education: Computer Uses in Education

INTRODUCTION
Educational environments has become increasingly diverse. Traditional schools and universities have a characteristically rigid structure, including instructor-defined—even nationally agreed—syllabi, fixed time schedules, entry requirements, and material costs to enter and exit. Novel institutional structures have been developed to overcome particular constraints. Community colleges, for instance, were created in an attempt to democratize education by offering instruction at a lower cost and by accomodating people with less flexibile schedules [9]. Distance learning programs intended to deliver education in remote parts of the world and for people who simply could not attend in-person classes. Course materials, including assessments, were delivered through mail ("correspondence education"), radio, television, and eventually the Internet, thereby addressing geographical and time related constraints of traditional instruction [18].

L@S'15, March 14–15, 2015, Vancouver, Canada.

DOI string from ACM form confirmation

The latest generation of online learning environments, characterized by massive open online courses (MOOCs), has pushed the boundary on the scale of education [27]. By design, MOOCs provide course materials to millions of people worldwide. This scale could be achieved by pre-recording lectures, designing assessments that can be graded automatically, and by leveraging the momentum of the number of poeple invovled (e.g., to facilitate peer learning or peer grading [16, 5]). Maybe by virtue of their large scale, their prominent instructors, or their adherence to contemporary interface designs, MOOCs rapidly became an online media phenomenon. People would sign up weeks in advance of the course launch date, many of whom would never even enter the course site. And among those who enroll and enter the site, a large proportion tends to only "sample" some content and leave again [14]. Many of the prototypical behaviors observed in MOOCs [14, 4] (other?) resemble those on online media platforms, such as YouTube or tumblr. This trend has also been reflected in the diversity of MOOC learners' motivations for enrolling [15].

Shortly after the first wave of courses had finished, extensive media coverage led to MOOCs becoming associated with high attrition rates [17, 21, 10]. Early MOOC research cautioned against dichotomizing learners into sucesses and failures based on course completion [14, 22]. Instead, more nuanced categorizations based on learner behavior [14, 6] (other?), motivations [15], or intentions [30] have been suggested. Ultimately, perspectives on persistence and attrition in MOOCs depend on how MOOCs have been conceptualized. Kizilcec and Schneider [15] proposed that MOOCs have bridged two different world: one world is governed by the user-centric norms of online media, where everyone is encouraged to be as active as they wish; the other world adheres to the "grammar of schooling", which presupposes instructor-defined goals that students strive to achieve [26]. Viewing MOOC participation as bridging these two worlds undoubtedly adds a layer of complexity to interpretations of attrition.

This paper presents a systematic investigation into attrition in MOOCs, based on self-report and behavioral data collected from [N] online learners in [M] courses. We begin by briefly reviewing the large literature on attrition in educational environments with a focus on important developments in understanding its causes. Building on this foundation of prior work, Study 1 offers insights into reasons for disengaging from MOOCs and explores relationships with prior behavior and reported intentions. In Study 2, we sought to develop a deeper understanding of attrition in MOOCs by conducting a case study. Targeting online learners who were predicted likely dropouts in a particular course were invited to provide feedback via a survey. TODO: complete paragraph based on what we actually do in Study 2.

RELATED WORK

Research and theorizing on attrition in education has a rich history. This review is intended to serve as a foundation to build on with the current research. The focus of this breif review is on how ways of thinking about attrition have developed over the last decades.

In-person Education

The majority of early work on attrition centered around theoretical models of students' decision to persist or dopout of a traditional higher education setting. An early model suggested that students' presistence is largely driven by their prior behavior, attitudes, adn norms [8]. The psychological processes involved in turning an intent to learn into the decision to persist were thought to be mediated by volition, i.e., the extent to which the student engages in goal-direct behaviors in the face of distraction [7]. Hence, motivation alone is necessary but not sufficient for persistence. Students may fail to sustain efforts in the absence of strong self-regulatory skills.

The next generation of psychological models, which were highly influential in the literature, emphasized the critical role of students' "fit" in the institution. Tinto's [25] student integration model posited that college students' decision to persist is a function of prior experiences and individual students' characteristics, and experiences during college. While piror experiences and characteristics are fixed, schools can influence the college experience, including the degree of social and academic integration. Tinto operationalized academic and social integration by GPA scores and the frequency of positive interactions with peers and instructors, respectively. This resonates with recent work highlighting the critical role of students' feelings of social belonging in achievement-oriented environments [29].

Tinto's work, which specifically targets traditional college students, prompted universities to be proactive in establishing environments that support student integration. Research on attrition in community college settings reiterates the importance of academic and social integration, but points out that non-persistence could indicate success depending on students' intent—students may leave after accomplishing their goal [3].

Distance Education and e-Learning

Dinstance education, in contrast to traditional in-person education, attracts a different student demographic and typically provides fewer opportunities for social integration. However, students in distance learning programs tend to lead social lives outside of school, maybe working part-time and living with their partner. Building on Tinto's model, Bean and Metzner [2] proposed a conceptual model of persistence that would be more applicable to nontraditional students. Persistence is thought to be a function of background characteristics (e.g., demographics), adademic and environmental variables (e.g., study habits, financial resources, work and family obligations), and academic and psychological outcomes (e.g., GPA, satisfaction). The significant change from Tinto's orginal model was the inclusion of environmental variables to account for the added complexity of nontraditional students' lives (see [12], for another adaptation of Tinto's model for distance education).

Rovai [24], combining Tinto's and Bean and Metzner's models with factors specific to online learning and pedagogical styles, proposed a composite persistence model specifically for students in online distance education programs. Among other novel factors, Rovai's model acknowledges the critical role of computer literacy in online learning. And yet, while the theoretical models become more developed, they became harder to apply to real-world settings, and hence, the empirical evidence to support them remained sparse. An exploratory study of reasons for attrition suggested the following eight constructs, based on over 1,000 online education students: academic and technical skills, learner motivation, time and support for studies, cost and access to the Internet, and technical problems [20]. An analysis of student behavior on an online education platform showed that 31% of variation in achievement could be accounted for by a small set of participation measures [19].

MOOCs

A survey of over one hundred learners who dropped out of MOOC showed that most learners indicated too little time due to work responsibilities, not enough social support inside and outside of the course, and insufficient academic and technical support from the course [11]. A qualitative analysis of public records, especially forum posts, from 42 MOOCs suggested the following reasons for attrition: lack of time, learner motivation, feelings of isolation, lack of interactivity, insufficient prior knowledge or skills, and hidden costs [13]. Based on just over one thousand learners who participated in the discussion forum of a particular MOOC, the likelihood of dropout was lower for those who actively participated in the first week of the course, those who served as an authority figure in the community on the forum, and those who did not engage in a particular subcommunity on the forum [23].

A number of machine learning approaches to predicting attrition have been tested, incluging hidden Markov models [1],

STUDY 1: REASONS FOR DISENGAGING

TODO: develop

RQ1 Why do people who drop out report they drop out?

RQ2 Who is most likely to drop out in terms of demographics, intentions/motivations, geo location, prior experience with online courses?

RQ3 What reasons do people from different subgroups (demographic, intention, etc.) report for dropping out?

RQ4 What proportion of people who dropped out were satisfied with what they got out of the course?

RQ5 What behaviors do people with different dropout reasons exhibit?

H1 prior experience – technical difficulties

H2 demographics (education) – course difficulty

Methods

TODO: Outline different courses and go through variabels in the course info table.

Table: course name; enrollment; pre-survey response rate; prop. demographics (age, gender, educ); who intended to do all; post-survey response rate; self-identified dropout rate

Results

Table: proportions for each dropout reason for each course

Table: correlations between reasons and one with cors between reasons and intentions

Plot: ecdf curves for proportion videos watched for each disengagement reason

Discussion

STUDY 2: UNDERSTANDING DISENGAGEMENT

To gain a better understanding of the attrition patterns identified in Study 1, we designed a smaller but more focused follow-up study. In Study 2, we reached out to learners who were likely to drop out of a course and asked them to provide feedback. Identifying learners who were likely to drop out was achieved with machine learning.

TODO: develop

H1: People who report 'not enough time' can be split into those who have external constraints and those who have low volition.

H2: people who report external committments (family, school), are more likely to specify the reason than people who have 'no time' because they have low volition

Methods

The particular MOOC under observation was an undergraduate level course on an advanced topic in computer science. It was offered in 2014 through Coursera. There were [N1] enrolled learners; [N2] watched more than one video, and [N3] attempted more than one assignment.

Predicting Disengagement

TODO: Details of the prediction algorithm.

Feedback Survey

Every learner who was predicted to disengage from the course was sent an email kindly requesting their help: "You are enrolled in [course name], but you've been less active recently. Could you help us understand why?" A very low response rate was expected, given that this subpopulation was defined by low engagement. 535 out of 6,050 learners started the survey (8.8% response rate), and 499 completed it (8.2% completion rate).

Learners were asked to report how satisfied they were with their progress in the course, and whether they were using the course materials more, less, or exactly as much as they would have liked. In addition, they reported intentions for engaging with each week's course content. Then, they were asked to openly report "what challenges, inside or outside of the course, [they] experienced while [they were] taking this course, if any?" The instructions encouraged them to list all challenges they could think of. This question was deliberately asked prior to any survey questions that could suggest particular reasons for disengagement.

Two research assistants independently developed a codebook for the resulting 435 non-empty open responses. Their codebooks were consolidated and applied on Mechanical Turk, where 4 'classification experts' independently coded all 200 randomly selected responses. Specific updates to the initial codebook were informed by category frequency, category correlations, and inter-coder agreement. The updated codebook was then applied by 4 Mechanical Turk 'classification experts' to the 235 remaining open responses.

Following the open resposne question, learners reported the difficulty of the course material and the extent to which their progress was hindred by a number of obstacles (see Table [XX]). The survey also included a measure of learners' sense of social and academic fit [28] (17 items, $M = 4.67, SD = 0.72, \alpha = 0.87$); mindset (4 items on the nature of intelligence and talent[1]; $M = 2.45, SD = 0.90, \alpha = 0.78$), and goal striving (5 items on motivation, perceived importance, committment, confidence, and distractions; $M = 3.08, SD = 0.83, \alpha = 0.76$).

TODO: update numbers once finalized, explain use of imputation

Results

TODO: Add results of prediction. How well did the prediction work? And who took the survey; how sure was the model that those who took the survey would actually drop out, relative to those who didn't take the survey?

TODO: Write up results of diagnostic survey, keeping it simple for now.

Discussion

GENERAL DISCUSSION

CONCLUSION

ACKNOWLEDGMENTS

Omitted for blind review.

REFERENCES

1. Balakrishnan, G., and Coetzee, D. Predicting student retention in massive open online courses using hidden markov models, 2013.

[1]Items were adpated from the mindset questionnaire available at `http://mindsetonline.com/testyourmindset/`.

2. Bean, J. P., and Metzner, B. S. A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research 55*, 4 (1985), 485–540.
3. Bers, T. H., and Smith, K. E. Persistence of community college students: The influence of student intent and academic and social integration. *Research in Higher Education 32*, 5 (1991), 539–556.
4. Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., and Seaton, D. Studying learning in the worldwide classroom: Research into edxs first mooc. *Research & Practice in Assessment 8* (2013), 13–25.
5. Cambre, J., Kulkarni, C., Bernstein, M. S., and Klemmer, S. R. Talkabout: small-group discussions in massive global classes. In *Proceedings of the first ACM conference on Learning@ scale conference*, ACM (2014), 161–162.
6. Clow, D. Moocs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, ACM (2013), 185–189.
7. Corno, L., and Kanfer, R. The role of volition in learning and performance. *Review of research in education* (1993), 301–341.
8. Fishbein, M., and Ajzen, I. *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley Publishing Company, 1975.
9. Goldrick-Rab, S. Challenges and opportunities for improving community college student success. *Review of Educational Research 80*, 3 (2010), 437–469.
10. Guthrie, D. Moocs are toast or at least should be. *Forbes* (2013, July).
11. Gütl, C., Rizzardini, R. H., Chang, V., and Morales, M. Attrition in mooc: Lessons learned from drop-out students. In *Learning Technology for Education in Cloud. MOOC and Big Data*. Springer, 2014, 37–48.
12. Kember, D. A longitudinal-process model of drop-out from distance education. *The Journal of Higher Education* (1989), 278–301.
13. Khalil, H., and Ebner, M. Moocs completion rates and possible methods to improve retention-a literature review. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, vol. 2014 (2014), 1305–1313.
14. Kizilcec, R. F., Piech, C., and Schneider, E. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge*, ACM (2013), 170–179.
15. Kizilcec, R. F., and Schneider, E. Motivation as a lens for understanding online learners: Towards data-driven design with the olei scale. *ACM Transactions on Computer-Human Interaction (TOCHI)* (in press), 24.
16. Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., and Klemmer, S. R. Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI) 20*, 6 (2013), 33.
17. Lewin, T. After setbacks, online courses are rethought. *New York Times* (2013, December).
18. Moore, M. G., and Kearsley, G. *Distance education: A systems view*. Wadsworth Publishing Company, 1966.
19. Morris, L. V., Finnegan, C., and Wu, S.-S. Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education 8*, 3 (2005), 221–231.
20. Muilenburg, L. Y., and Berge, Z. L. Student barriers to online learning: A factor analytic study. *Distance Education 26*, 1 (2005), 29–48.
21. Parr, C. Mooc completion rates 'below 7%'. *Times Higher Education* (2013, May).
22. Rivard, R. Measuring the mooc dropout rate. *Inside Higher Ed 8* (2013).
23. Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., and Sherer, J. Social factors that contribute to attrition in moocs. In *Proceedings of the first ACM conference on Learning@ scale conference*, ACM (2014), 197–198.
24. Rovai, A. P. In search of higher persistence rates in distance education online programs. *The Internet and Higher Education 6*, 1 (2003), 1–16.
25. Tinto, V. Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research* (1975), 89–125.
26. Tyack, D., and Tobin, W. The "grammar" of schooling: Why has it been so hard to change? *American Educational Research Journal 31*, 3 (1994), 453–479.
27. Waldrop, M. Online learning: Campus 2.0. *Nature 495*, 7440 (2013), 160–163.
28. Walton, G. M., and Cohen, G. L. A question of belonging: race, social fit, and achievement. *Journal of personality and social psychology 92*, 1 (2007), 82.
29. Walton, G. M., and Cohen, G. L. A brief social-belonging intervention improves academic and health outcomes of minority students. *Science 331*, 6023 (2011), 1447–1451.
30. Wilkowski, J., Deutsch, A., and Russell, D. M. Student skill and goal achievement in the mapping with google mooc. In *Proceedings of the first ACM conference on Learning@ scale conference*, ACM (2014), 3–10.