Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses

René F. Kizilcec*
Dept. of Communication
Stanford University
Stanford, CA-94305
kizilcec@stanford.edu

Chris Piech*
Dept. of Computer Science
Stanford University
Stanford, CA-94305
piech@cs.stanford.edu

Emily Schneider*
School of Education
Stanford University
Stanford, CA-94305
elfs@cs.stanford.edu

ABSTRACT

As MOOCs grow in popularity, the relatively low completion rates of learners has been a central criticism. This focus on completion rates, however, reflects a monolithic view of disengagement that does not allow MOOC designers to target interventions or develop adaptive course features for particular subpopulations of learners. To address this, we present a simple, scalable, and informative classification method that identifies a small number of longitudinal engagement trajectories in MOOCs. Learners are classified based on their patterns of interaction with video lectures and assessments, the primary features of most MOOCs to date.

In an analysis of three computer science MOOCs, the classifier consistently identifies four prototypical trajectories of engagement. The most notable of these is the learners who stay engaged through the course without taking assessments. These trajectories are also a useful framework for the comparison of learner engagement between different course structures or instructional approaches. We compare learners in each trajectory and course across demographics, forum participation, video access, and reports of overall experience. These results inform a discussion of future interventions, research, and design directions for MOOCs. Potential improvements to the classification mechanism are also discussed, including the introduction of more fine-grained analytics.

Categories and Subject Descriptors

K.3.1 [**Computers and Education**]: Computer Uses in Education; K.3.1 [**Computers and Education**]: Distance learning—*Massive Open Online Course, Learner Engagement Pattern*

*equal contribution

General Terms

Algorithms, Measurement

Keywords

Learning Analytics, Learner Engagement Patterns, MOOCs

1. INTRODUCTION

Massive open online courses (MOOCs) are the most recent and highly publicized entrant to a rapidly expanding universe of open educational resources. As of late 2012, the majority of MOOCs are virtual, distributed classrooms that exist for six to ten weeks at a time. These MOOCs are structured learning environments that emphasize instructional videos and regular assessments, centralizing activities on a single platform. This is a distinct model from the set of learner-directed, open-ended courses that are now known as "cMOOCs" because of their grounding in connectivist theories of learning [25, 20, 8].

The relatively low completion rates of MOOC participants has been a central criticism in the popular discourse. This narrative implies a binary categorization of learners: those who pass the class by adhering to the instructor's expectations throughout the course–and everyone else. This monolithic view of so-called "noncompleters" obscures the many reasons that a learner might disengage from a MOOC. It also makes no allowances for learners who choose to participate in some aspects of the MOOC but not others, staying engaged with the course but not earning a statement of accomplishment. In contrast, one could emphasize the importance of individual differences and consider all learners to be unique in their interactions with the platform. But whereas the monolithic view overgeneralizes, this individualist perspective overcomplicates. In this paper, we seek to strike a balance by identifying a small yet meaningful set of patterns of engagement and disengagement. MOOC designers can apply this simple and scalable categorization to target interventions and develop adaptive course features [5].

With no cost to entry or exit, MOOCs attract learners with a wide range of backgrounds and intentions, as well as personal or technical constraints to participation. Given the heterogeneity of the population, we would be remiss to make *a priori* assumptions about the appropriate characteristics or behaviors around which to categorize learners, or which pathways and outcomes are more or less valuable for their learning. Analogous challenges can be found in research on

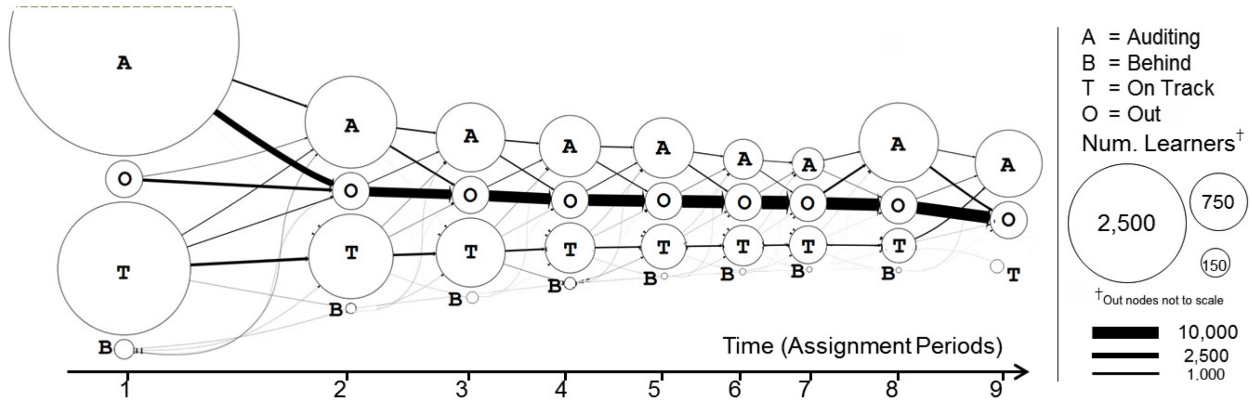

Figure 1: Labels for learners (GS-level)–arcs show movement of students from one assignment period to next.

community colleges–the closest brick-and-mortar analogue to MOOCs in terms of the diversity of educational objectives among their students [14]–and on unstructured virtual inquiry environments, where there is not a clear notion of "correct" pathways through the available resources. Using unsupervised clustering techniques, community college researchers have developed meaningful typologies of students based on longitudinal enrollment patterns [2] and survey measures of engagement [23]. Likewise, cluster-based analyses for inquiry environments have distinguished meaningful patterns in learner engagement with content [1].

In this paper we employ a methodology for characterizing learner engagement with MOOCs that builds on methods used in this previous literature. We define learner trajectories as longitudinal patterns of engagement with the two primary features of the course–video lectures and assessments. We uncover four prototypical categories of engagement consistently across three MOOCs by clustering on engagement patterns. We focus on interactions with course content, because learning is a process of individual knowledge construction that emerges in a dynamic process of interactions among learners, resources, and instructors [4, 25]. In MOOCs, these interactions are shaped by the design of instruction, content, assessment, and platform features. To inform effective design changes and interventions along these dimensions that would target the needs of learners on a particular trajectory, we compare clusters based on learner characteristics and behaviors.

2. COURSE DEMOGRAPHICS

Our analysis of learner trajectories is based on three computer science courses that vary in their level of sophistication: "Computer Science 101" covers high school level content (HS-level), "Algorithms: Design and Analysis" covers undergraduate level content (UG-level), and "Probabilistic Graphical Models" is a graduate level course (GS-level). Table 1 provides basic demographic information and summarizes how many learners were active on the course website at any point in time (as opposed to simply enrolling and never participating). In all three courses, the vast majority of active learners are employed full-time, followed by graduate and undergraduate students. Moreover, most learners in the UG-level and GS-level courses come from technology-related industries. The majority of learners in the UG-level course report to hold a Master's or a Bachelor's degree. Geographically, most learners are located in the United States, followed by India and Russia.

Table 1 also reports the distribution of active learners over the quantiles of the 2011 Human Development Index (HDI)–a composite measure of life expectancy, education, and income indices [29]. The distribution in the GS- and HS-level courses is very similar, with over two-thirds of active learners from very high-HDI countries. The distribution in the UG-level course is less skewed between very high-, high-, and medium-HDI countries, though low-HDI countries account for a similarly low 3% of learners.

Table 1: Course Demographics

	HS	UG	GS
Active Learners	46096	26887	21108
Gender (M/F)	64%/36%	88%/12%	88%/12%
Age	33 (14)‡	31 (11)‡	36 (12)‡
HDI			
Very High	69%	54%	70%
High	13%	17%	14%
Medium	15%	26%	15%
Low	3%	3%	1%

‡ Mean (Std. Dev.)

3. CLUSTERING

Our learning analytics methodology is designed to identify a small number of canonical ways in which students interact with MOOCs. In our analysis we first compute a description for each student of the way in which the student was "engaged" throughout the duration of a course and then apply clustering techniques to find subpopulations in these engagement descriptions. Running this methodology over the courses in our study uncovers four prototypical engagement patterns for learners' interactions with the contemporary instantiation of MOOCs.

The first step in our methodology is to generate a rough description of each student's individual engagement in a course. For each assessment period, all participants are labeled either "on track" (did the assessment on time), "behind" (turned in the assessment late), "auditing" (didn't do the assessment but engaged by watching a video or doing a quiz), or "out" (didn't participate in the course at all). These labels were chosen because they could be easily collected, and would make sense in any MOOC that is based on videos and assessments, regardless of content area or the pedagogical strategies of the course. Figure 1 visualizes the longitudinal distribution of learners assigned to each label for the GS-level course. For each assessment period, nodes represent the number of learners in each category; between assessment periods, an arc represents the number of learners who retain the same label or move between labels. Due to space constraints the "Out" nodes are not to scale. The complete list of labels that a participant is assigned for each assessment periods is called her "engagement description" As a concrete example of an engagement description: imagine a learner in the GS-level course had completed the first five assignments on time, finished the sixth assignment late and then continued to watch videos without bothering with the last three assignments. Using the notation in Figure 1, that particular student's engagement description would have been, [T, T, T, T, T, B, A, A, A].

Once we had engagement descriptions for each learner in a course, we applied the k-means clustering algorithm–the standard centroid-based clustering algorithm–to identify prototypical engagement patterns. To calculate the similarity between engagement descriptions for two students, a computation which is needed for clustering, we assigned a numerical value to each label (on track = 3, behind = 2, auditing = 1, out = 0) and computed the L1 norm of the list of numbers. Since we wanted to account for the random properties of k-means we repeated our clustering one hundred times and selected the solution with the highest likelihood. Though clustering was performed separately on all three courses, the process extracted the same four high-level, prototypical engagement trajectories (Table 2 shows their distribution in the three classes):

1. 'Completing': learners who completed the majority of the assessments offered in the class. Though these participants varied in how well they performed on the assessment, they all at least attempted the assignments. This engagement pattern is most similar to a student in a traditional class.
2. 'Auditing': learners who did assessments infrequently if at all and engaged instead by watching video lectures. Students in this cluster followed the course for the majority of its duration. No students in this cluster obtained course credit.
3. 'Disengaging': learners who did assessments at the beginning of the course but then have a marked decrease in engagement (their engagement patterns look like Completing at the beginning of the course but then the student either disappears from the course entirely or sparsely watches video lectures). The moments at which the learners disengage differ, but it is generally in the first third of the class.
4. 'Sampling': learners who watched video lectures for only one or two assessment periods (generally learners in this category watch just a single video). Though many learners "sample" at the beginning of the course, there are many others that briefly explore the material when the class is already fully under way.

To evaluate the clusters produced by this methodology we tested that (1) the trends derived were robust to perturbations in the methodology, (2) the clusters that we arrived at had a healthy "goodness of fit" for the data, and (3) that the trends made sense from an educational perspective. The results below lend support that the clusters extracted are meaningful and useful.

Table 2: Cluster Breakdown

Course	Auditing	Completing	Disengaging	Sampling
HS	6%	27%	28%	39%
UG	6%	8%	12%	74%
GS	9%	5%	6%	80%

(1) Though we had extracted trends, it was necessary to test whether they reflected meaningful patterns in learning, or if they were a manifestation of the parameters that we used to explore engagement. We hoped to show that the patterns we identified were so strong that even if we had made a few minor changes in our methodology, the same trends of engagement would hold. First we tested whether the patterns in the class were robust enough that the clusters did not change substantially when we experimented with different feature sets. Including "assignment pass" and removing "behind" from the set of labels we assigned to learners in the Algorithms course produced highly analogous centroids and similar labeling, 95% overlap in cluster labels and 94% overlap respectively. In addition, we tried running our clustering with a different choice for k (number of clusters) and found that increasing k divided the four high level patterns into sub-clusters. For example using $k = 5$ and clustering on the UG level course split the Sampling cluster into learners who sampled a video at the beginning of the course and learners who sampled a video in one of the later assessment periods.

(2) It was also necessary to show that the four high-level clusters of students provided an accurate generalization of the data. To verify the "goodness of fit" of our clustering we ran the Silhouette cluster validation test [22]. A positive silhouette score reflects that, on average, a given engagement description is more similar to other descriptions in its cluster than to descriptions in the other clusters (which in turn suggests that the clusters reflect true subgroups of the original population). The maximum silhouette score of 1.0 means that all learners in a cluster are exactly the same. Though our clustering classified some students that were halfway between two of the categories, the overwhelming majority of learners fit cleanly into one of the trajectories (98% positive silhouette, average silhouette score = 0.8).

(3) The final evaluation of our clustering methodology was that the algorithm returned trends that make sense from an educational point of view. The trends of engagement pass a common sense test: it is plausible to imagine *a posteriori* that students would interact in an educational platform in these high level ways. This is important because it provides a framework which enables research that can hypothesize other properties of students in these clusters. Since our labels were drawn from a small discrete set of engagement labels, we extracted meaningful patterns of engagement (Completing, Auditing, etc). In contrast, using assignment grades or lecture counts as features produced clusters that were mostly defined by student scores in the first week (e.g. 'learners who got a high grade in assignment one and then dropped out', 'learners who received a medium grade in assignment one and then dropped out', etc.). These clusters are less informative of learning processes and potential pedagogical improvements.

4. CLUSTER ANALYSIS

The plurality of engagement trajectories calls for an equally diverse set of tools and interventions to support these subpopulations of learners. We compare the clusters along behavioral features routinely recorded in the MOOC database, as well as self-report features collected through optional surveys. The goal is to provide educators, instructional designers, and platform developers with insights for designing effective, and potentially adaptive, learning environments that best meet the needs of MOOC participants. In this section we first describe and motivate the set of features to compare trajectories on, and then present the results of our cross-cluster analyses. In the following section we offer interpretations of these findings, suggest design changes for future MOOCs, and highlight research opportunities.

4.1 Features

Understanding who learners are, why they enroll in the course, and other activities in the course is a first step towards illuminating potential influences on the self-selection of learners into these engagement patterns. Differences in the distribution of particular features across clusters may indicate that these demographic variables or learning processes affect learners' engagement decisions. In all courses, learners received a survey at the end of the course. In the UG-level course, an additional pre-course survey was administered. Table 3 contains survey response rates by engagement group for each course. Note the high response rates in the UG-level course.

Table 3: Survey Response Rates

	HS	UG (pre)	UG (post)	GS
Auditing	13%	23%	14%	23%
Completing	43%	31%	45%	65%
Disengaging	4%	25%	3%	29%
Sampling	3%	20%	1%	5%

Survey Demographics: The demographic section of the optional surveys included age, gender, employment status, highest level of education achieved, and years of work experience.

Geographical Location: Learners' IP addresses were recorded and looked up on a country level using MaxMind's GeoLite database. The country labels were then merged with the 2011 Human Development Index data [29]. Are MOOCs meeting their promise of global access? How do learners in different parts of the world interact with these courses?

Intentions: At the start of the course, learners reported their reasons for enrolling by choosing applicable options from a set of predefined reasons. (E.g. "Enhance my resume for career or college advancement" or "It's free") We computed the probability of indicating each reason given the learner's engagement trajectory. MOOCs attract a variety of learners with particular sets of objectives and motivations. Understanding learners' goals is a precondition to effective designs that provide affordances for the varied needs of learners.

Overall Experience: In post-course surveys, learners rated their "overall experience with the course" on a 7-point Likert scale from 'Poor' to 'Excellent'. This measure provides insight into learners' satisfaction with the course experience.

Forum Activity: A rich history of research in computer-supported collaborative learning, as well as classroom and informal settings, shows that learning is enhanced through collaboration and discourse with a community [27]. The discussion forum provides the opportunity for this type of social learning in MOOCs. We measure learners' active participation on the forum by counting the number of posts and comments each learner created during the course.

Streaming Index (SI): This measure serves as a proxy for learners' access to in-video assessments, which are only available when streaming videos off the course website. Access to in-video assessments is pedagogically important because formative assessment that gives leaners instant feedback has been associated with positive learning outcomes: Opportunities for frequent, formative testing enable learners to reflect on their knowledge state [3] and actively retrieve information in a way that facilitates learning [21]. Although the clustering of engagement patterns is partly based on video consumption, video access (streaming versus downloading) is independent of clustering. SI is defined as the proportion of overall lecture consumption that occurs online on the platform, as opposed to offline (downloaded).

$$\text{Streaming Index (SI)} = \frac{\text{online lecture consumption}}{\text{total lecture consumption}}$$

4.2 Results

Learner clusters are compared along the feature dimensions introduced above using formal statistical tests. A one-way analysis of variance (ANOVA) is performed on each dimension (Table 4) and Tukey Honest Significant Differences (HSD) adjustments (p_{HSD}) are used for post hoc pair-wise cluster comparisons (Table 6) [11]. The tables report the statistical and practical significances of the comparisons. The latter is reported in terms of effect size: partial eta-squared (*partial* η^2) for multiple clusters and Cohen's d for two clusters [6]. By convention, *partial* $\eta^2 > .14$ is considered a large effect, and *partial* $\eta^2 > .06$ medium; $d > .8$ is considered a large effect, and $d > .5$ medium. Absolute effect sizes can be extracted from group averages in Table 4. In the case of contrasting intentions to enroll, the statistical tests are based on 10,000 bootstrapped permutations of engagement group labels. To test for significance we evaluated the likelihood of observing the reasons that learners reported given their actual engagement group.

4.2.1 Survey Demographics

Note that the following demographic comparisons between engagement groups are only valid under the assumption that responding to the survey is independent of the demographic indicators (e.g. males and females are equally likely to respond to the survey).

Gender: All three courses enrolled more male than female learners, though this trend was much more prominent for courses with more sophisticated content. There were around seven times more men than women in the UG- and GS-level courses (odds ratio of 7.4 and 7.5, respectively). The gender gap was much less prominent in the HS-level course, with only about twice as many men than women (odds ratio of 1.8). A log linear model of gender on cluster membership yields log odds for each engagement trajectory with confidence intervals for each course (Figure 2). Within each course, the gender ratios across the four engagement trajectories are not significantly different from each other

Table 4: Comparisons between Engagement Trajectories (One-Way ANOVAs)

Indicator	Average: Auditing	Completing	Disengaging	Sampling	F	*p*	Partial η^2
HS							
Overall Experience†	.894	.912	.830	.796	109	<.001*	.047
Streaming Index	.869	.880	.900	.855	61.8	<.001*	.004
Forum Activity	.242	.788	.189	.017	1536	<.001*	.091
UG							
Overall Experience†	.731	.874	.716	n.a.	84.1	<.001*	.153*
Streaming Index	.643	.664	.723	.743	48.0	<.001*	.006
Forum Activity	.251	1.71	.238	.024	1315	<.001*	.128
GS							
Overall Experience†	.771	.794	.657	.687	44.9	<.001*	.056
Streaming Index	.519	.667	.655	.661	64.8	<.001*	.009
Forum Activity	.536	7.18	1.98	.090	2692	<.001*	.277*

n.a. = not available due to low survey response rate
* Significant at $p < .05$ or $d > .8$ † Self-report measure (scaled to unit interval)

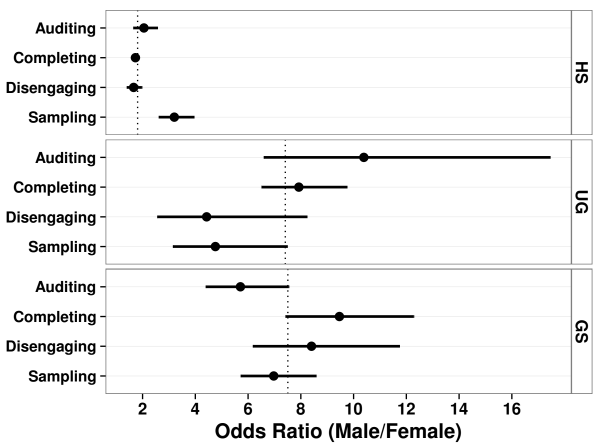

Figure 2: Odds ratio between number of males and females with 95% C.I. and overall gender odds ratio in course (dotted line)

(except for Sampling learners in the HS-level course), suggesting that gender is not associated with engagement trajectories. However, the ratio for Completing learners lies significantly below the the course-wide average (dotted lines in Figure 2) in the HS-level course (p=.05), but just significantly above in the GS-level course (p=.06). This may indicate a trend where females are relatively less frequently Completing learners in higher-level courses.

Age: Learner engagement groups are approximately equally distributed within age brackets, except in the GS-level course, where there were fewer elderly (65+) Completing and Auditing learners, and none under the age of 18.

Employment status: In all courses, learners on different engagement trajectories are approximately equally distributed within the three most represented employment statuses: working full-time, graduate and undergraduate student.

4.2.2 Geographical Location

To extend the analysis of how active learners are distributed over countries with different HDI levels, Table 5 shows the distribution over engagement trajectories within each HDI tier. As HDI increases, the proportion of Completing and Disengaging learners increases, while the proportion of Sampling learners decreases. However, the distribution for low-HDI countries might not be representative, given that learners from low-HDI countries account for only 1% of all active learners. To circumvent this issue, we analyzed the distribution of engagement patterns for the four most represented countries (US, India, Russia, and the UK) which happen to span over three HDI levels: the US and UK rank 'very high', Russia ranks 'high', and India ranks 'medium'. The analysis confirms the pattern observed for medium-HDI countries: in all three courses, learners from India participate considerably more as Sampling (ca. 14% points above other three countries), than as Completing and Disengaging learners (ca. 9% and 7% points below).

Table 5: HDI Level Breakdown (GS-level)

	Very High	High	Medium	Low
Auditing	13%	8%	11%	14%
Completing	8%	6%	4%	2%
Disengaging	10%	9%	5%	4%
Sampling	69%	77%	80%	80%
All Learners	70%	14%	15%	1%

4.2.3 Intentions

For all three courses the two most frequently chosen reasons for enrolling are, because they find it fun and challenging, and they are interested in the topic. Moreover, the probability of enrolling to enhance their resume is particularly high for Completing learners (15% in the HS-, 33% in the UG-, and 20% in GS-level course). In ALGO, Completing learners were the most likely to say they were in the class because they thought it was fun and challenging (61%, p<.001), followed by Auditing (58%, p<.05), Disengaging (55%) and Sampling learners (52%, p<.001).

4.2.4 Overall Experience

Ratings of overall experience (Figure 3) are highly significantly different between engagement groups in all three courses (p<.001 in all courses; *partial* η^2=.153 in the UG-level, and *partial* η^2=.056 in the GS-level course). In the HS-

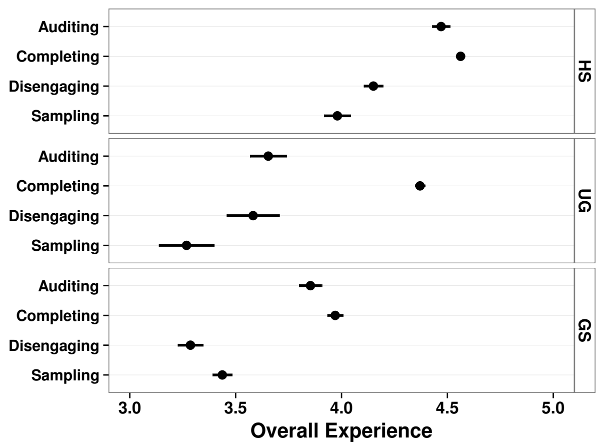

Figure 3: Overall experience with ±1 standard error bars

level and GS-level courses, the overall experience of Completing and Auditing learners is not significantly different from each other, but significantly above Disengaging (d=5.66 in the HS-level and d=.648 in the GS-level course) and Sampling learners (d=.785 in the HS-level and d=.465 in the GS-level course). The UG-level course exhibits a different pattern, with Completing learners having a significantly better overall experience than the other engagement groups.

4.2.5 Forum Activity

Forum activity (Figure 4) varies significantly between engagement trajectories with medium to large effect sizes, with Completing learners participating at significantly higher rates than learners in other engagement trajectories ($p<.001$). For example, in the GS-level course, Completing learners exhibit significantly higher levels of activity on the discussion board compared to Auditing (d=.721, mean=.536), Disengaging (d=.480, mean=1.98), and Sampling learners (d=1.97, mean =.09). The significance of these differences is preserved when controlling for the different durations of these learners' participation in the course. On average, Completing learners write 1.71 posts and comments in the UG-level, .788 in the HS-level, and 7.18 in GS-level course.

4.2.6 Streaming Index

A consistent pattern in all three courses is an average Streaming Index (SI) above 0.5 for each engagement trajectory, which indicates that streaming is the dominant form of access to video lectures (Figure 5). In the HS-level course, the SI is consistently higher than the other courses across all engagement patterns: streaming accounts for around 88% of video consumption, compared 70% in the UG-level and 63% in GS-level courses. Surprisingly, within each course, there are significant differences in SI between most engagement trajectories, though effect sizes are only marginal. The most notable difference is that of Auditing learners in the GS-level course, who watch about half (SI=.519) of the lectures offline, compared to Completing, Disengaging, and Sampling learners (with SIs between .655 and .667). In the UG-level course, the SI of Completing and Auditing, and Disengaging and Sampling learners are not significantly different ($p=.405$ and $p=.068$, respectively), while all other pair-wise compar-

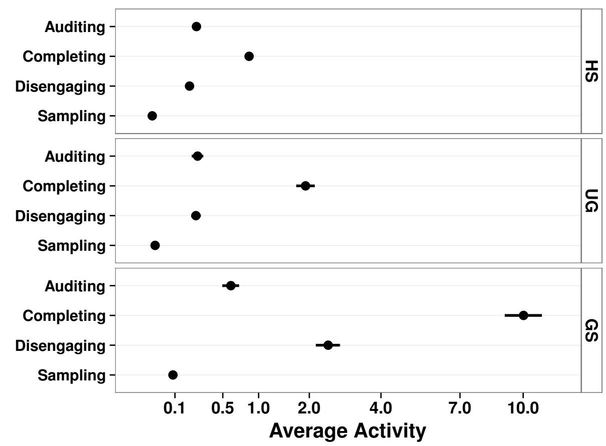

Figure 4: Forum activity with ±1 standard error bars (square-root scale)

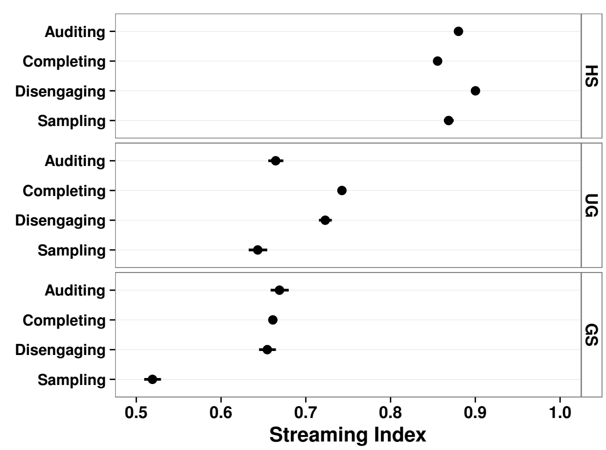

Figure 5: Streaming Index with ±1 standard error bars

isons are statistically significant ($p<.001$). This indicates that Completing learners in the UG-level course tend to watch lectures offline less than Disengaging learners (10% points difference in SI).

5. DISCUSSION

First, we reflect on our classification methodology and propose ways to extend and adapt it for future MOOC research. Then, we discuss the results of the cluster analysis in terms of future research questions and design directions for MOOCs. We conclude with cross-course comparisons and a broad set of remarks on the scope of MOOC design, research, and global accessibility.

5.1 Extension of Analytics Methodology

The process of extracting patterns of engagement has given us new insights into potential next steps for analytics research applied to the MOOC context. Given the unprecedented volume of data collected through massive open access courses, this environment provides an exciting opportunity for the Learning Analytics community. In this paper we

Table 6: Post Hoc Pair-Wise Comparison between Engagement Trajectories

Indicator	Comp.-Audi. p_{HSD}	Comp.-Audi. d	Comp.-Dise. p_{HSD}	Comp.-Dise. d	Comp.-Samp. p_{HSD}	Comp.-Samp. d	Audi.-Dise. p_{HSD}	Audi.-Dise. d	Audi.-Samp. p_{HSD}	Audi.-Samp. d	Dise.-Samp. p_{HSD}	Dise.-Samp. d
HS												
Overall Experience†	.132	.133	<.001*	.566	<.001*	.785*	<.001*	.336	<.001*	.461	.005*	.149
Streaming Index	.225	.043	<.001*	.078	<.001*	.082	<.001*	.123	.121	.041	<.001*	.152
Forum Activity	<.001*	.205	<.001*	.282	<.001*	.413	.023*	.056	<.001*	.546	<.001*	.278
UG												
Overall Experience†	<.001*	.846*	<.001*	1.03*	n.a.	n.a.	.804	.059	n.a.	n.a.	n.a.	n.a.
Streaming Index	.405	.053	<.001*	.150	<.001*	.195	<.001*	.200	<.001*	.246	.068	.049
Forum Activity	<.001*	.318	<.001*	.367	<.001*	.917*	.963	.007	<.001*	.407	<.001*	.259
GS												
Overall Experience†	.354	.114	<.001*	.648	<.001*	.465	<.001*	.500	<.001*	.342	.145	.121
Streaming Index	<.001*	.390	.832	.041	.932	.019	<.001*	.352	<.001*	.336	.945	.016
Forum Activity	<.001*	.721	<.001*	.480	<.001*	1.97*	<.001*	.284	<.001*	.356	<.001*	.982*

n.a. = not available due to low survey response rate
* Significant at $p < .05$ or $d > .8$ † Self-report measure (scaled to unit interval)

outlined our first foray into the new dataset. While our algorithm was useful for identifying high level patterns in how students are approaching the contemporary instances of MOOCs, there are several improvements that we would recommend for future analytics research.

The strategy behind our clustering technique was to create a single variable for engagement, and to look for trends in how that variable changed over time. Our coarse feature set was useful for consistently identifying very high level patterns of engagement across different courses. We are interested to see what details in learning patterns can be expressed through a more nuanced measure of engagement, particularly one that is built from finer-grained time slices and the incorporation of more user information. In our study we used an assessment period (approximately one week) as the smallest granule of time for assigning labels of engagements to learners, a modelling simplification which was the result of the data which was immediately available on the first MOOC classes. Since all user interactions with the learning system are time stamped, we could construct a model of engagement with a granularity on the order of hours (if not smaller). A finer view of time could allow our understanding of students to delve into the details of user work sessions. Moreover, in conjunction with a more precise time frame we could also incorporate more types of learner data in our clustering–for example, the timing of learners' participation in the forum or the resources they turn to while in the process of completing assessments. Scores received on quizzes and assignments would add the dimension of achievement levels to the engagement trajectory model. A richer set of features, one that included a smaller time granularity and more user data, would allow a clustering algorithm to uncover more subtle patterns. However, the cost of using complex feature sets is that the patterns extracted may miss the big picture, which we have sought to provide in this paper.

Another adjacent space in this line of analytics research is exploration into the different applications of learner engagement classification. The clustering detailed in this paper provides a quick and easy way to compare different course instances in the MOOC context. Being able to contrast student engagement patterns could be used to explore both the impacts of different pedagogies and how students themselves change over time. Since MOOCs are relatively new to most learners, it is reasonable to hypothesize that users are going to adapt over time to better take advantage of free material. As a result we predict that learner patterns of engagement will also change–a trend which could be explored through clustering engagement over present and future offerings of the same course.

In general, for those studying MOOCs in the future, we recommend that they incorporate an understanding of the high level ways in which students engage. This lens, we believe, is much more insightful than a raw report of the number of students who enrolled or the number of students who obtained a certificate.

5.2 Interpretation of Results

The clusters reveal a plurality of trajectories through a course that are not currently acknowledged in the design and discourse around MOOCs. Auditing appears to be an alternative engagement pathway for meeting learner needs, with Auditing learners reporting similarly high levels of overall experience to Completing learners in two of three courses. This implies different underlying preferences or constraints for Auditing and Completing learners, and points to an opportunity to design features to actively support these engagement trajectories. Auditors could be identified via self-report or based on a predictive model that should be developed for the early detection of engagement patterns. For example, Auditing learners could be encouraged to focus on video-watching and not be shown potentially frustrating reminders about assessment completion. Moreover, instructors could downplay the importance of assessments when outlining expectations for the course, in order to avoid discouraging learners from following this alternative engagement path. Another design strategy could be removing assessments altogether for Auditing learners. However, evidence from cognitive psychology suggests that testing not only assess learning but facilitates it [18], which implies that even though assessments do not fit with the engagement profile of Auditing learners, MOOC designers should not deprive them of the option to engage in assessment activities that could serve to enhance their learning.

Compared to Auditing and Completing learners, Disengaging and Sampling learners almost universally report lower levels of overall experience. In the context of our conception of learning as a process of interactions with the learning environment, we can think of these prototypical patterns of engagement as reflecting a trajectory of interactions that led, at some point, to the learner disengaging from the course. From the surveys, the most prominent reasons that learners across the three courses selected on a list of reasons for disengaging were: personal commitment(s), work conflict, and course workload. While the personal constraints reflected in the first two reasons may be unassailable, the three together can be interpreted to mean that some of these learners may have been better served by a course that was offered at a slower pace or even entirely self-paced. Investigating points of disengagement is a ripe area for future work. More qualitative or survey-based data should be gathered on why learners choose to leave these courses. This data should be combined with more fine-grained analytics to develop a predictive model for when learners are likely to disengage in the future and what category of disengagement their choice falls into.

Cross-cluster comparisons in survey responses and learning processes allow us to develop hypotheses about the mechanisms of how or why a learner may have stayed on a particular trajectory. These hypotheses can be designed around and tested in future work. For example, forum activity is a behavioral measure excluded from the clustering algorithm that differentiates the trajectories and deepens our understanding of the activities of highly engaged learners. Completing learners exhibit the highest level of activity on the forum; notably, this rate is much higher than that of Disengaging learners, who are initially assessment-oriented and then disengage from the course. While to some extent this is a reflection of Completing learners' high level of engagement with the course overall, we may hypothesize that participation on the forum creates a positive feedback loop for some learners, as they are provided with social and informational inputs that help them stay on their trajectory towards completion. This hypothesis can be tested using encouragement designs, such as reputation systems, or by leveraging social influence by displaying participation levels or contributions of other learners [17]. Platform designers should also consider building other community-oriented features to promote pro-social behavior, such as text or video chat, small-group projects, or facilitated discussions. Linking these community features more strongly to the content in the course–for example, "situated" discussions linked to a point in a video or other resource–may further promote learning. In addition to being theory-driven, these designs and interventions should be based on future research that delves more deeply into the mechanisms of MOOC learners' engagement on the forum–including those learners who read the forum but do not contribute to it–and how these interactions relate to their decisions in the course overall. Future research should examine the structure of the community in terms of the social networks that develop, as well as the incentives to contribute and build trust among members. Another strand of research could explore how discourse on MOOC discussion boards facilitates the construction of knowledge [7].

5.3 Cross-Course Comparisons

The clusters also act as a standard set of outcomes for comparing the three courses. While each course adheres to a standard MOOC format, differences across courses in the distribution of learners in the clusters can bring into relief the content and instructional strategies of each course. For example, the HS-level course stands out from the UG- and GS-level course with over half of the participants being Completing learners or disengaging after being on that trajectory initially. This speaks to the wider accessibility of the entry-level course content, especially considering that the HS-level course has a far higher proportion of women enrolling, as well as double the number of active learners as the other two courses. Notably, the HS-level course also has a 26% higher average Streaming Index (SI) than the UG- and 40% higher than GS-level courses. This variation in SI may be partially due to the relative levels of difficulty of the courses. But another likely influence is that in-video exercises are only available to those who stream the videos, and whereas the videos in the UG- and GS-level courses primarily feature multiple choice questions, the in-video exercises in the HS-level course tend to be short programming challenges. These programming challenges are likely to be fun and rewarding to participants, and additionally enhance learning by requiring learners to actively demonstrate their knowledge [4]. MOOC designers and instructors should be prompted by this observation to continue to develop performance-based approaches to assessment. A future experiment could test the relative importance of these types of assessments for learning.

Another trend illuminated by comparing the HS-level course to the other courses is the result that females are relatively less frequently Completing learners in the GS-level course. This finding is consistent with research on stereotype threat, which shows that women tend to perform worse than equally skilled men on more challenging or frustrating quantitative task [28]. Among other explanations, it is theorized that this is because the feeling of challenge is likely to evoke anxiety that failing will confirm negative stereotypes about their group (that women are not good at quantitative tasks). Moreover, this effect is more likely to occur for individuals who care highly about the domain–as is the case with women who are enrolled in the GS-level course. Interventions demonstrated to counteract stereotype threat among women taking math tests include presenting the test as one where there are no gender differences associated with results–one where "everyone can achieve"–or a test that is described as designed to help you learn, not one that is diagnostic of your current skills [26]. Both of these devices could be used to frame assessments to counteract instances of stereotype threat in MOOCs.

Two trends in the characteristics of participants in the three MOOCs are particularly salient given the dominant themes in popular discourse about MOOCs. The first is why people choose to participate in MOOCs. Much commentary has focused on the role that MOOCs can play in credentialing and opportunities for job (re)training. While acquiring new skills, along with the certification of those skills, is certainly important to many participants, there are far more who are driven by the intellectual stimulation offered by the courses. MOOCs are evidently playing an important role in providing opportunities for engaging in lifelong learning outside of the confines of an institution, and can potentially serve as a powerful means of harnessing the "cognitive surplus" [24] that has emerged in a post-industrial age. Analogous to the case of learners who audit, designers and instructors should be aware of the needs and goals of learners who are enrolling for personal enrichment, and consider how content or instruction could be adapted to better satisfy them. Future research should explore these populations more thoroughly, turning to surveys, interviews or case studies as a source of contextually rich data about their needs and experiences.

The second trend concerns the promise that MOOCs hold for global access to education. Though there are many exceptions, it is notable that the learners in all three courses tend to be well-educated professionals from high-HDI countries. Moreover, the majority are male. These facts are partially an artefact of the technical nature of these courses. The awareness of MOOCs is also likely much higher among learners from the US, which dominates the enrollment of the three courses under analysis. But broadband access is likely to be a factor as well, as many learners in low- and medium-HDI countries are faced by intermittent, slow, or metered bandwidth that would make it a challenge to fully engage with video-heavy courses. MOOC designers should consider decreasing videos or offering only the audio version of the lecture, two strategies that would also have implications for pedagogy and learning. The skew in geographical distribution is a clear call to action for those in the MOOC space who are focused on issues of access and equity, and explanations for this phenomenon should be pursued in order to develop more culturally sensitive and accommodating MOOCs.

6. CONCLUSION

Learners in MOOCs who do not adhere to traditional expectations, centered around regular assessment and culminating in a certificate of completion, count towards the high attrition rates that receive outsized media attention. Through our analysis we present a different framework for the conversation about MOOC engagement, which accounts for multiple types of student engagement and disengagement. We started out with the assumption that there are a small number of alternative patterns of interactions with MOOC content. Through our research we were able to extract, across all three classes studied, four prototypical learner trajectories; three of which would have been considered "noncompleting" under a monolithic view of course completion. Using these patterns as a lens to more closely analyze learner behavior and backgrounds across the different trajectories, we were able to suggest research and design directions for future courses.

This work is one of the first applications of analytics techniques into the new wealth of learner data that is generated by MOOCs–datasets that we believe present exciting opportunities for the learning analytics community. Though we were able to find high-level patterns, the vast amounts of information available should allow for the discovery of more subtle and deeper trends. A particularly rich area for future research is combining more fine-grained analytics with data on the noncognitive factors that inevitably influence the choices they make when moving through a MOOC. Motivation, self-regulation, tenacity, attitudes towards the processes of learning, and feelings of confidence and acceptance are but some of many psychological factors that affect academic performance [12, 10]. Along with other unobserved latent variables, these internal states are likely associated with choices that learners make about particular activities as well as with overall patterns of engagement with the course. Those factors that are found to be influential could inspire the design of tools, features, or interventions that are either broadly applicable or adapted to the needs of particular types of learners. Interventions can also be developed to directly target these factors, such as the promotion of micro-steps to simplify the learning process and increase learners' ability to succeed [13], or interventions designed to promote a growth mindset among learners [9].

The large scale and virtual nature of MOOCs creates a fertile ground for experiments based on the hypotheses and design directions suggested by this paper. Modifications to subsequent instances of the same course would yield interesting insights, as would the continued comparison of multiple courses with different structures or approaches to instruction. Other innovative designs of MOOC instruction, content, or platform features–based on principles of the learning sciences or human-computer interaction–should likewise be subject to experimentation and evaluation. One potential design area is the development of simple "cognitive tools" [15], such as an integrated note-taking or concept-mapping system that would allow learners to actively interpret the course content, or task lists and calendar features for staying organized. Another is addressing learners' prior knowledge in the field, which is widely understood to mediate learners' encounters with new information and subsequent academic performance. Calibrating prior knowledge could aid in providing adaptive content to learners, such as a finely-tuned hinting structure as part of assessment procedures [16], or a set of open educational resources linked to from within instruction on a particular topic. A third challenging design problem opens up in light of the increasing ubiquity of media multitasking [19], especially in an environment where learners' attention can be quickly compromised by attending to their social networking needs in the next browser tab.

A powerful promise of MOOCs is the unprecedented level of global access to a vast set of educational opportunities. We have the chance to design these new learning environments both for learners who want a standard assessment-centric course and learners who have less structured motivations. Using a standard set of outcomes will allow researchers and designers across the MOOC space to develop a collective awareness of optimal approaches for meeting the needs of MOOC learners. The engagement trajectory model is one viable option for a high-level characterization of the effect of refinements and interventions in MOOCs.

7. ACKNOWLEDGMENTS

We would like to thank Stanford's Office of the Vice Provost for Online Learning, Daphne Koller, Clifford Nass, and Roy Pea for their support and guidance in the preparation of this paper. We are also indebted to our anonymous reviewers, the LAK program chairs, and the members of the Stanford Lytics Lab for their invaluable feedback.

8. REFERENCES

[1] S. Amershi and C. Conati. Automatic recognition of learner types in exploratory learning environments. *Handbook of Educational Data Mining*, page 213, 2010.

[2] P. Bahr. The bird's eye view of community colleges: A behavioral typology of first-time students based on cluster analytic classification. *Research in Higher Education*, 51(8):724–749, 2010.

[3] P. Black and D. Wiliam. Assessment and classroom learning. *Assessment in education*, 5(1):7–74, 1998.

[4] J. Bransford, A. Brown, and R. Cocking. *How people learn: Brain, mind, experience, and school*. National Academies Press, 2000.

[5] P. Brusilovsky and E. Millán. User models for adaptive hypermedia and adaptive educational systems. *The adaptive web*, pages 3–53, 2007.

[6] J. Cohen. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum, 1988.

[7] B. De Wever, T. Schellens, M. Valcke, and H. Van Keer. Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, 46(1):6–28, 2006.

[8] S. Downes. New technology supporting informal learning. *Journal of Emerging Technologies in Web Intelligence*, 2(1):27–33, 2010.

[9] C. Dweck. *Mindset: The new psychology of success*. Ballantine Books, 2007.

[10] C. Dweck, G. Walton, and G. Cohen. Academic tenacity: Mindset and skills that promote long-term learning. *Gates Foundation. Seattle, WA: Bill & Melinda Gates Foundation*, 2011.

[11] B. Everitt and T. Hothorn. *A handbook of statistical analyses using R*. Chapman & Hall/CRC, 2009.

[12] C. Farrington, M. Roderick, E. Allensworth, J. Nagaoka, T. Keyes, D. Johnson, and N. Beechum. Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance: A critical literature review. *Chicago: University of Chicago Consortium on Chicago School Research*, 2012.

[13] B. Fogg. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*, page 40. ACM, 2009.

[14] S. Goldrick-Rab. Challenges and opportunities for improving community college student success. *Review of Educational Research*, 80(3):437–469, 2010.

[15] B. Kim and T. Reeves. Reframing research on learning with technology: in search of the meaning of cognitive tools. *Instructional Science*, 35(3):207–256, 2007.

[16] K. Koedinger, A. Corbett, et al. Cognitive tutors: Technology bringing learning science to the classroom. *The Cambridge handbook of the learning sciences*, pages 61–78, 2006.

[17] R. Kraut, P. Resnick, S. Kiesler, Y. Ren, Y. Chen, M. Burke, N. Kittur, J. Riedl, and J. Konstan. *Building successful online communities: Evidence-based social design*. The MIT Press, 2012.

[18] M. McDaniel, H. Roediger, and K. McDermott. Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, 14(2):200–206, 2007.

[19] E. Ophir, C. Nass, and A. Wagner. Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences*, 106(37):15583–15587, 2009.

[20] C. Rodriguez. Moocs and the ai-stanford like courses: Two successful and distinct course formats for massive open online courses. *Learning*, 2012.

[21] H. Roediger III and J. Karpicke. Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science*, 17(3):249–255, 2006.

[22] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(0):53 – 65, 1987.

[23] V. Saenz, D. Hatch, B. Bukoski, S. Kim, K. Lee, and P. Valdez. Community college student engagement patterns a typology revealed through exploratory cluster analysis. *Community College Review*, 39(3):235–267, 2011.

[24] C. Shirky. *Cognitive surplus: Creativity and generosity in a connected age*. ePenguin, 2010.

[25] G. Siemens. Connectivism: A learning theory for the digital age. 2004.

[26] S. Spencer, C. Steele, and D. Quinn. Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1):4–28, 1999.

[27] G. Stahl, T. Koschmann, and D. Suthers. Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer, editor, *Cambridge handbook of the learning sciences*, pages 409–426. Cambridge, UK: Cambridge University Press, 2006.

[28] C. Steele, S. Spencer, and J. Aronson. Contending with group image: The psychology of stereotype and social identity threat. *Advances in experimental social psychology*, 34:379–440, 2002.

[29] United Nations Development Programme. 2011 human development report. Retrieved from `http://hdr.undp.org/en/media/HDR_2011_Statistical_Tables.xls`. (Accessed 18-Jan-2013).