Linguistic Reflections of Student Engagement in Massive Open Online Courses

Miaomiao Wen, Diyi Yang and Carolyn Penstein Rosé
Language Technology Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{mwen,diyiy,cprose}@cs.cmu.edu

Abstract

While data from Massive Open Online Courses (MOOCs) offers the potential to gain new insights into the ways in which online communities can contribute to student learning, much of the richness of the data trace is still yet to be mined. In particular, very little work has attempted fine-grained content analyses of the student interactions in MOOCs. Survey research indicates the importance of student goals and intentions in keeping them involved in a MOOC over time. Automated fine-grained content analyses offer the potential to detect and monitor evidence of student engagement and how it relates to other aspects of their behavior. Ultimately these indicators reflect their commitment to remaining in the course. As a methodological contribution, in this paper we investigate using computational linguistic models to measure learner motivation and cognitive engagement from the text of forum posts. We validate our techniques using survival models that evaluate the predictive validity of these variables in connection with attrition over time. We conduct this evaluation in three MOOCs focusing on very different types of learning materials. Prior work demonstrates that participation in the discussion forums at all is a strong indicator of student commitment. Our methodology allows us to differentiate better among these students, and to identify danger signs that a struggling student is in need of support within a population whose interaction with the course offers the opportunity for effective support to be administered. Theoretical and practical implications will be discussed.

1 Introduction

The recent development of Massive Open Online Course (MOOC) websites such as Coursera[1], edX[2] and Udacity[3], demonstrates the potential of distance learning and lifelong learning to reach the masses. However, one disappointment has been that only one in every 20 students who enroll in such courses actually finish(Koller et al. 2013). In order to understand the attrition problem and work towards solutions, especially given the varied backgrounds and motivations of students who choose to enroll in a MOOC (DeBoer et al. 2013), we need to highlight and understand the value sought and obtained by the participants of MOOCs including that reflected in their discussion posts, especially the "non-completing" population (Koller et al. 2013).

In this paper, we propose to gauge a student's engagement using linguistic analysis applied to the student's forum posts within the MOOC course. Based on the learning sciences literature, we quantify a student's level of engagement in a MOOC from two different angles: (1) displayed level of motivation to continue with the course and (2) the level of cognitive engagement with the learning material. Student motivation to continue is important- without it, it is impossible for a student to regulate him or her effort to move forward productively in the course. Nevertheless, for learning it is necessary for the student to process the course content in a meaningful way. In other words, cognitive engagement is critical. Ultimately it is this grappling with the course content over time that will be the vehicle through which the student achieves the desired learning outcomes.

Conversation in the course forum is replete with terms that imply learner motivation. These terms may include those suggested by the literature on learner motivation or simply from our everyday language. E.g. "I tried very hard to follow the course schedule" and "I couldn't even finish the second lecture." In this paper, we attempt to automatically measure learner motivation based on such markers found in posts on the course discussion forums. Our analysis offers new insights into the relation between language use and underlying learner motivation in a MOOC context.

Besides student motivational state, the level of cognitive engagement is another important aspect of student participation(Carini et al. 2006). For example, "This week's video lecture is interesting, the boy in the middle seemed tired, yawning and so on." and "The video shows a classroom culture where the kids clearly understand the rules of conversation and acknowledge each others contribution." These two posts comment on the same video lecture, but the first post is more descriptive at a surface level while the second one is more interpretive, and displays more reflection. We measure this difference in cognitive engagement with an estimated level of language abstraction. We find that users whose posts show a higher level of cognitive engagement are more likely to continue participating in the forum discussion.

The distant nature and the sheer size of MOOCs require new approaches for providing student feedback and guid-

[1]https://www.coursera.org/
[2]https://www.edx.org/
[3]https://www.udacity.com/

ing instructor intervention (Ramesh et al. 2013). One big challenge is that MOOCs are far from uniform. In this paper, we test the generality of our measures in three Coursera MOOCs focusing on distinct subjects. We demonstrate that our measures of engagement are consistently predicative of student dropout from the course forum across the three courses. With this validation, our hope is that in the long run, our automatic engagement measures can help instructors target their attention to those who show serious intention of finishing the course, but nevertheless struggle through due to dips in learner motivation. Our linguistic analysis provides further indicators that some students are going through the motions in a course but may need support in order to fully engage with the material. Again such monitoring might aid instructors in using their limited human resources to the best advantage.

In the remainder of the paper we begin by describing our dataset and discussing related work. Next, we explain how we automatically measure student engagement from a user's forum posts from the two perspectives we highlight above. We then continue with a survival analysis that estimates the influence of our two measures of engagement on MOOC dropout rate. Finally, we conclude with a summary and possible future work.

2 Coursera dataset

In preparation for a partnership with an instructor team for a Coursera MOOC that was launched in Fall of 2013, we were given permission by Coursera to crawl and study a small number of other courses. Our dataset consists of three courses: one social science course, "Accountable Talk: Conversation that works[4]", offered in October 2013, which has 1,146 active users (active users refer to those who post at least one post in a course forum) and 5,107 forum posts; one literature course, "Fantasy and Science Fiction: the human mind, our modern world[5]", offered in June 2013, which has 771 active users who have posted 6,520 posts in the course forum; one programming course, "Learn to Program: The Fundamentals[6]", offered in August 2013, which has 3,590 active users and 24,963 forum posts. All three courses are officially seven weeks long. Each course has seven week specific subforums and a separate general subforum for more general discussion about the course. Our analysis is limited to behavior within the discussion forums.

3 Related Work

3.1 Learner Motivation

Most of the recent research on learner motivation in MOOCs is based on surveys and relatively small samples of hand-coded user-stated goals or reasons for dropout (e.g. Cheng et al. 2013; Christensen et al. 2013; DeBoer et al. 2013; Poellhuber et al. 2013). Poellhuber et al. (2013) find that user goals specified in the pre-course survey were the strongest predictors of later learning behaviors. Motivation is identified as an important determinant of engagement in MOOCs in the Milligan et al.(2013) study. However, different courses design different enrollment motivational questionnaire items, which makes it difficult to generalize the conclusions from course to course. Another drawback is that learner motivation is volatile. In particular, distant learners can lose interest very fast even if they had been progressing well in the past (Keller & Suzuki, 2004). It is important to monitor learner motivation and how it varies along the course weeks. We propose to automatically measure learner motivation based on linguistic cues in the forum posts.

3.2 Cognitive engagement

Research has consistently found that the cognitive processes involved in higher-order thinking lead to better knowledge acquisition (e.g. Chi & Bassock, 1989; Graham & Golan, 1991; Chi, 2000). Previous work has investigated students' cognitive engagement in both face-to-face (Corno & Mandinach, 1983; Newman et al. 1996) and computer mediated communication (CMC) environments (Garrison et al. 1999; Zhu 2006). In this paper, we try to measure the cognitive engagement of a MOOC user based on how much personal interpretation are contained in the posts.

3.3 MOOC Analysis

The MOOC literature so far has focused on a summative view of user participation and dropout - trying to assess the rate at which different groups of users complete the course (e.g. Kizilcec et al. 2013; Brinton et al. 2013; Ramesh et al. 2013). There are mainly three types of information available about the participation patterns of MOOC users: the survey information, the clickstream behavioral data, and the forum posts. From the survey information, we can partly understand the initial motivation at the time of enrollment of a subset of users who filled in the survey. Unfortunately, even for those users, this information does not help us understand the dynamics of motivational change. From the clickstream data of a user's online activities, we can see if people are working hard or not, but we cannot tell the reasons why their level of activity changes from time to time. The discussion forums provide the students with social learning opportunities while at the same time providing a portal into their minds. The activeness of a course's online forum negatively correlates with the volume of students that drop out of the course (Brinton et al. 2013). In our work, we utilize linguistic features in the forum posts that correlate with dropout to get insights into the user experience that would be invisible from the survey information or clickstream data.

Given the rich recent work on MOOC user dropout analysis, very little has attempted finer-grained content analysis of the course discussion forums. One exception is Ramesh et al. (2013), which uses sentiment and subjectivity of user posts to predict engagement/disengagement. However, neither sentiment nor subjectivity ended up being predictive of engagement in that work. One explanation is that engaged learners also post content with negative sentiment on the course, such as complaints about peer-grading. Thus, the problem is more complex than the operationalization used in that work.

[4]https://www.coursera.org/course/accountabletalk

[5]https://www.coursera.org/course/fantasysf

[6]https://www.coursera.org/course/programming1

In our work, we use survival models to understand how attrition happens along the way as students participate in a course. This approach has been applied to online medical support communities to quantify the impact of receipt of emotional and informational support on user commitment to the community (Wang et al. 2012). Yang et al. (2013) and Rose et al. (2014) have also used survival models to measure the influence of social positioning factors on drop out of a MOOC. Our research contributes to this body of work.

4 Methods

4.1 Predicting Learner Motivation

The level of a student's motivation strongly influences the intensity of the student's participation in the course. Previous research has shown that it is possible to categorize learner motivation based on a students' description of planned learning actions (Ng & Bereiter, 1991; Dowson & McInerney, 2003). The identified motivation categorization has a substantial relationship to both learning behavior and learning outcomes. But the lab-based experimental techniques used in this prior work are impractical for the ever-growing size of MOOCs. It is difficult for instructors to personally identify students who lack motivation based on their own personal inspection in MOOCs given the high student to instructor ratio. To overcome these challenges, we build machine learning models to automatically identify level of learner motivation based on posts to the course forum. We validate our measure in a domain general way by not only testing on data from the same course, but also by training on one course and then testing one other in order to uncover course independent motivation cues. The linguistic features that are predicative of learner motivation provide insights into what motivates the learners.

4.1.1 Creating the Human-Coded Dataset: MTurk We used Amazon's Mechanical Turk (MTurk) to make it practical to construct a reliable annotated corpus for developing our automated measure of student motivation. Amazon's Mechanical Turk is an online marketplace for crowdsourcing. It allows requesters to post jobs and workers to choose jobs they would like to complete. Jobs are defined and paid in units of so-called Human Intelligence Tasks (HITs). Snow et al. (2008) has shown that the combined judgments of a small number (about 5) of naive annotators on MTurk leads to ratings of texts that are very similar to those of experts. This applies for content such as the emotions expressed, the relative timing of events referred to in the text, word similarity, word sense disambiguation, and linguistic entailment or implication. As we show below, MTurk workers' judgments of learner motivation are also similar to coders who are familiar with the course content.

We randomly sampled 514 posts from the Accountable Talk course forums and 534 posts from the Fantasy and Science Fiction course forums. The non-English posts were manually filtered out. In order to construct a hand-coded dataset for training machine learning models later, we employed MTurk workers to rate each message with the level of learner motivation towards the course the corresponding post had. We provided them with explicit definitions to use in making their judgment. For each post, the annotator had to indicate how motivated she perceived the post author to be towards the course by a 1-7 Likert scale ranging from "Extremely unmotivated" to "Extremely motivated". Each request was labeled by six different annotators. We paid $0.06 for rating each post. To encourage workers to take the numeric rating task seriously, we also asked them to highlight words and phrases in the post that provided evidence for their ratings. To further control the annotation quality, we required that all workers have a United States location and have 98% or more of their previous submissions accepted. We monitored the annotation job and manually filtered out annotators who submitted uniform or seemingly random annotations.

We define the motivation score of a post as the average of the six scores assigned by the annotators. The distributions of resulting motivation scores are shown in Figure 1. The following two examples from our final hand-coded dataset of the Accountable Talk class illustrate the scale. One shows high motivation, and the other demonstrates low motivation. The example posts shown in this paper are lightly disguised and shortened to protect user privacy.

- Learner Motivation = 7.0 (Extremely motivated)
  Referring to the last video on IRE impacts in our learning environments, I have to confess that I have been a victim of IRE and I can recall the silence followed by an exact and final received from a bright student.... Many ESL classes are like the cemetery of optional responses let alone engineering discussions. The Singing Man class is like a dream for many ESL teachers or even students if they have a chance to see the video! ...Lets practice this in our classrooms to share the feedbacks later!
- Learner Motivation = 1.0 (Extremely unmotivated)
  I have taken several coursera courses, and while I am willing to give every course a chance, I was underwhelmed by the presentation. I would strongly suggest you looking at other courses and ramping up the lectures. I'm sure the content is worthy, I am just not motivated to endure a bland presentation to get to it. All the best, XX.

4.1.2 Inter-annotator Agreement To evaluate the reliability of the annotations we calculate the intra-class correlation coefficient for the motivation annotation. Intra-class correlation(Koch, 1982) is appropriate to assess the consistency of quantitative measurements when all objects are not rated by the same judges. The intra-class correlation coefficient for learner motivation is 0.74 for the Accountable Talk class and 0.72 for the Fantasy and Science Fiction course.

To assess the validity of their ratings, we also had the workers code 30 Accountable Talk forum posts which had been previously coded by experts. The correlation of MTurkers' average ratings and the experts' average ratings was moderate ($r = .74$) for level of learner motivation.

We acknowledge that the perception of motivation is highly subjective and annotators may have inconsistent scales. In an attempt to mitigate this risk, instead of using the raw motivational score from MTurk, for each course, we break the set of annotated posts into two balanced groups

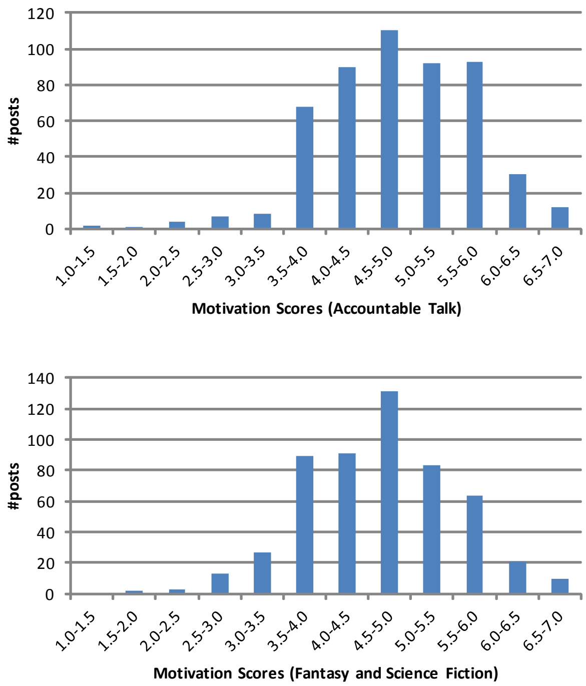

Figure 1: Annotated motivation score distribution.

based on the motivation scores: "motivated" posts and "unmotivated" posts.

4.1.3 Linguistic Markers of Learner Motivation In this section, we work to find domain-independent motivation cues so that a machine learning model is able to capture motivation expressed in posts reliably across different MOOCs. Building on the literature of learner motivation, we design five linguistic features and describe them below. The features are binary indicators of whether certain words appeared in the post or not. Table 1 describes the distribution of the motivational markers in our Accountable Talk annotated dataset. We do not include the Fantasy and Science Fiction dataset in this analysis, because they will serve as a test domain dataset for our prediction task in the next section.

Apply words (Table 1, line 1): previous research on E-learning has found that motivation to learn can be expressed as the attention and effort required to complete a learning task and then apply the new material to the work site or life (Esque & McCausland, 1997). Actively relating learning to potential application is a sign of a motivated learner (Moshinskie, 2001). So we hypothesize that words that indicate application of new knowledge can be cues of learner motivation.

The Apply lexicon we use consists of words that are synonyms of "apply" or "use": "apply", "try", "utilize", "employ", "practice", "use", "help", "exploit" and "implement".
Need words (Table 1, line 2) show the participant's need, plan and goals: "hope", "want", "need", "will", "would like", "plan", "aim" and "goal". Previous research has shown that learners could be encouraged to identify and articulate clear aims and goals for the course to increase motivation (Locke & Latham, 2002; Milligan et al. 2013).
LIWC-cognitive words (Table 1, line 3): The cognitive mechanism dictionary in LIWC (Pennebaker & King, 1999) includes such terms as "thinking", "realized", "understand", "insight" and "comprehend".
First person pronouns (Table 1, line 4): using more first person pronouns may indicate the user can relate the discussion to self effectively.
Positive words (Table 1, line 5) from the sentiment lexicon (Liu et al. 2005) are also indicators of learner motivation. Learners with positive attitudes have been demonstrated to be more motivated in E-learning settings (Moshinskie, 2001). Note that negative words are not necessarily indicative of unmotivated posts, because an engaged learner may also post negative comments. This has also been reported in earlier work by Ramesh et al. (2013).

The features we use here are mostly indicators of high user motivation. The features that are indicative of low user motivation do not appear as frequently as we expected from the literature. This may be partly due to the fact that students who post in the forum have higher learner motivation in general.

Feature	In Motivated post set	In Unmotivated post set
Apply**	57%	42%
Need**	54%	37%
LIWC -cognitive**	56%	38%
1st Person***	98%	86%
Positive***	91%	77%

Table 1: Features for predicting learner motivation. A binomial test is used to measure the feature distribution difference between the motivated and unmotivated post sets(**: p < 0.01, ***: p < 0.001).

4.1.4 Experimental Setup To evaluate the robustness and domain-independence of the analysis from the previous section, we set up our motivation prediction experiments on the two courses. We treat Accountable Talk as a "development domain" since we use it for developing and identifying linguistic features. Fantasy and Science Fiction is thus our "test domain" since it was not used for identifying the features.

For each post, we classify it as "motivated" or "unmotivated". The amount of data from the two courses is balanced within each category. In particular, each category contains 257 posts from the Accountable Talk course and 267 posts for the Fantasy and Science Fiction course.

We compare three different feature sets: a unigram feature representation as a baseline feature set, a linguistic classifier (Ling.) using only the linguistic features described in the previous section, and a combined feature set (Unigram+Ling.). We use logistic regression for our binary classification task. We employ liblinear (Fan et al. 2008) in Weka (Witten & Frank, 2005) to build the linear models. In order

	In-domain		**Cross-domain**	
Train	Accountable	Fantasy	Accountable	Fantasy
Test	Accountable	Fantasy	Fantasy	Accountable
Unigram	71.1%	64.0%	61.0%	61.3%
Ling.	65.2%	60.1%	61.4%	60.8%
Unigram+Ling.	72.3%	66.7%	63.3%	63.7%

Table 2: Accuracies of our three classifiers for the Accountable Talk course (Accountable) and the Fantasy and Science Fiction course (Fantasy), for in-domain and cross-domain settings. The random baseline performance is 50%.

to prevent overfitting we use Ridge (L2) regularization.

4.1.5 Motivation Prediction We now show how our feature based analysis can be used in a machine learning model for automatically classifying forum posts according to learner motivation.

To ensure we capture the course-independent learner motivation markers, we evaluate the classifiers both in an in-domain setting, with a 10-fold cross validation, and in a cross-domain setting, where we train on one course's data and test on the other (Table 2). For both our development (Accountable Talk) and our test (Fantasy and Science Fiction) domains, and in both the in-domain and cross-domain settings, the linguistic features give 1-3% absolute improvement over the unigram model.

The experiments in this section confirm that our theory-inspired features are indeed effective in practice, and generalize well to new domains. The bag-of-words model is hard to be applied to different course posts due to the different content of the courses. For example, many motivational posts in the Accountable Talk course discuss about teaching strategies. So words such as "student" and "classroom" have high feature weight in the model. This is not necessarily true for the other courses whose content has nothing to do with teaching.

In this section, we examine learner motivation where it can be perceived by a human. However, it is naive to assume that every forum post of a user can be regarded as a motivational statement. Many posts do not contain markers of learner motivation. In the next section, we measure the cognitive engagement level of a student based on her posts, which may be detectable more broadly.

4.2 Level of Cognitive Engagement

Level of cognitive engagement captures the attention and effort in interpreting, analyzing and reasoning about the course material that is visible in discussion posts (Stoney & Oliver, 1999). Previous work uses manual content analysis to examine studentscognitive engagement in computer-mediated communication (CMC)(Fahy et al. 2001; Zhu 2006). In the MOOC forums, some of the posts are more descriptive of a particular scenario. Some of the posts contain more higher-order thinking, such as deeper interpretations of the course material. Whether the post is more descriptive or interpretive may reflect different levels of cognitive engagement of the post author. Recent work shows that level of language abstraction reflects level of cognitive inferences (Beukeboom, 2014). In this section, we measure the level of cognitive engagement of a MOOC user with the level of language abstraction of her forum posts.

4.2.1 Measuring Level of Language Abstraction Concrete words refer to things, events, and properties that we can perceive directly with our senses, such as "trees", "walking", and "red". Abstract words refer to ideas and concepts that are distant from immediate perception, such as "sense", "analysis", and "disputable"(Turney et al. 2011).

Previous work measures level of language abstraction with Linguistic Inquiry and Word Count (LIWC) word categories (Gill & Oberlander, 2002; Pennebaker & King, 1999; Yarkoni, 2010; Beukeboom, 2013). For a broader word coverage, we use the automatically generated abstractness dictionary from Turney et al. (2011) which is publicly available. This dictionary contains 114,501 words. They automatically calculate a numerical rating of the degree of abstractness of a word on a scale from 0 (highly concrete) to 1 (highly abstract) based on generated feature vectors from the contexts the word has been found in.

The mean level of abstraction was computed for each post by adding the abstractness score of each word in the post and dividing that by the total number of words. The following are two example posts from the Accountable Talk course Week 2 subforum, one with high level of abstraction and one with low level of abstraction. Based on the abstraction dictionary, abstract words are in italic and concrete words are underlined.

- Level of abstraction = 0.85 (top 10%)
  I *agree*. Probably *what* you just have to keep in mind is that you are there to help them *learn* by giving them opportunities to *REASON* out. In that case, you will not just *accept* the student's *answer* but let them *explain* how they arrived towards it. Keep in mind to appreciate and *challenge* their *answers*.
- Level of abstraction = 0.13 (bottom 10%)
  I teach <u>science</u> to gifted <u>middle school</u> students. The students learned to have conversations with me as a class and with the expert her wrote Chapter <u>8</u> of a text published in <u>2000</u>. They are trying to design erosion control features for the building of a <u>basketball</u> court at the bottom of a <u>hill</u> in <u>rainy</u> <u>Oregon</u>.

We believe that level of language abstraction reflects the understanding that goes into using those abstract words when creating the post. In the Learn to Program course forums, many discussion threads are solving actual programming problems, which is very different from the other two courses where more subjective reflections of the course contents are shared. Higher level of language abstraction reflects the understanding of a broader problem. More concrete words are used when describing a particular bug a student encounters. Below are two examples.

- Level of abstraction = 0.65 (top 10%)
  I have the same problems here. Make *sure* that your *variable* names match exactly. *Remember* that under-bars connect words together. I *know* something to do with the read_board(board_file) function, but still need someone to *explain* more clearly.
- Level of abstraction = 0.30 (bottom 10%)
  >>> print(python, is)('python', 'is') >>> print('like', 'the', 'instructors', 'python') It leaves the 'quotes' and commas, when the instructor does the same type of print in the example she gets not parenthesis, quotes, or commas. Does anyone know why?

5 Validation Experiments

We use survival analysis to validate that participants with higher measured level of engagement will stay active in the forums longer, controlling for other forum behaviors such as how many posts the user contributes. We apply our linguistic measures described in Section 4 to quantify student engagement. We use the in-domain learner motivation classifiers with both linguistic and unigram features (Section 4.1.5) for the Accountable Talk class and the Fantasy and Science Fiction class. We use the classifier trained on the Accountable Talk dataset to assign motivated/unmotivated labels to the posts in the Learn to Program course.

5.1 Survival Model Design

Survival models can be regarded as a type of regression model, which captures influences on time-related outcomes, such as whether or when an event occurs. In our case, we are investigating our engagement measures' influence on when a course participant drops out of the course forum. More specifically, our goal is to understand whether our automatic measures of student engagement can predict her length of participation in the course forum. Survival analysis is known to provide less biased estimates than simpler techniques (e.g., standard least squares linear regression) that do not take into account the potentially truncated nature of time-to-event data (e.g., users who had not yet left the community at the time of the analysis but might at some point subsequently). From a more technical perspective, a survival model is a form of proportional odds logistic regression, where a prediction about the likelihood of a failure occurring is made at each time point based on the presence of some set of predictors. The estimated weights on the predictors are referred to as hazard ratios. The hazard ratio of a predictor indicates how the relative likelihood of the failure occurring increases or decreases with an increase or decrease in the associated predictor. We use the statistical software package Stata (Stata, 2001). We assume a Weibull distribution of survival times, which is generally appropriate for modeling survival.

For each of our three courses, we include all the active students, i.e. who contributed one or more posts to the course forums. We define the time intervals as student participation weeks. We considered the timestamp of the first post by each student as the starting date for that student's participation in the course discussion forums and the date of the last post as the end of participation unless it is the last course week.

Dependent Variable:

In our model, the dependent measure is **Dropout**, which is 1 on a student's last week of active participation unless it is the last course week (i.e. the seventh course week), and 0 on other weeks.

Control Variables:

Cohort1: This is a binary indicator that describes whether a user had ever posted in the first course week (1) or not (0). Members who join the course in earlier weeks are more likely than others to continue participating in discussion forums(Yang et al. 2013).

PostCountByUser: This is the number of messages a member posts in the forums in a week, which is a basic effort measure of engagement of a student.

CommentCount: This is the number of comments a user's posts receive in the forums in a week. Since this variable is highly correlated with PostCountByUser ($r > .70$ for all three courses). In order to avoid multicollinearity problems, we only include PostCountByUser in the final models.

Independent variables:

AvgMotivation is the percentage of an individual's posts in that week that are predicted as "motivated" using our model with both unigram and linguistic features (Section 4.1.4).

AvgCogEngagement: This variable measures the average abstractness score per post each week.

We note that AvgMotivation and AvgCogEngagement are not correlated with PostCountByUser ($r < .20$ for all three courses). So they are orthogonal to the simpler measure of student engagement. AvgMotivation is not correlated with AvgAbstractness ($r < .10$ for all three courses). Thus, it is acceptable to include these variables together in the same model.

5.2 Survival Model Results

Table 3 reports the estimates from the survival models for the control and independent variables entered into the survival regression.

Effects are reported in terms of the hazard ratio (HR), which is the effect of an explanatory variable on the risk or probability of participants drop out from the course forum. Because all the explanatory variables except Cohort1 have been standardized, the hazard rate here is the predicted change in the probability of dropout from the course forum for a unit increase in the predictor variable(i.e., Cohort1 changing from 0 to 1 or the continuous variable increasing by a standard deviation when all the other variables are at their mean levels).

Our variables show similar effects across our three courses (Table 3). Here we explain the results on the Accountable Talk course. The hazard ratio value for Cohort1

	Accountable Talk		Fantasy and Science Fiction		Learn to Program	
Control/Indep. Variable	**HR**	**Std. Err.**	**HR**	**Std. Err.**	**HR**	**Std. Err.**
Cohort1	.68***	.05	.82*	.08	.81*	.04
PostCountByUser	.86***	.02	.90***	.02	.76***	.04
AvgMotivation	.58*	.13	.82*	.08	.84***	.04
AvgCogEngagement	.94*	.02	.92**	.03	.53**	.13

Table 3: Results of the survival analysis(*: p<0.05, **: p<0.01, ***: p<0.001).

means that members survival in the group is 32% [7] higher for those who have posted in the first course week. Similarly, the hazard ratio for PostCountByUser indicates that survival rates are 14% [8] higher for those who posted a standard deviation more posts than average.

Controlling for when the participants started to post in the forum and the total number of posts published each week, both learner motivation and average level of abstraction significantly influenced the dropout rates in the same direction. Those whose posts expressed an average of one standard deviation more learner motivation (AvgMotivation) are 42% [9] more likely to continue posting in the course forum. Those whose posts have an average of one standard deviation higher cognitive engagement level (AvgCogEngagement) are 6% [10] more likely to continue posting in the course forum. AvgMotivation is relatively more predicative of user dropout than AvgCogEngagement for the Accountable Talk course and the Fantasy and Science Fiction course, while AvgCogEngagement is more predicative of user dropout in the Learn to Program course. This may be due to that in the Learn to Program course more technical problems are discussed and less posts contain motivation markers.

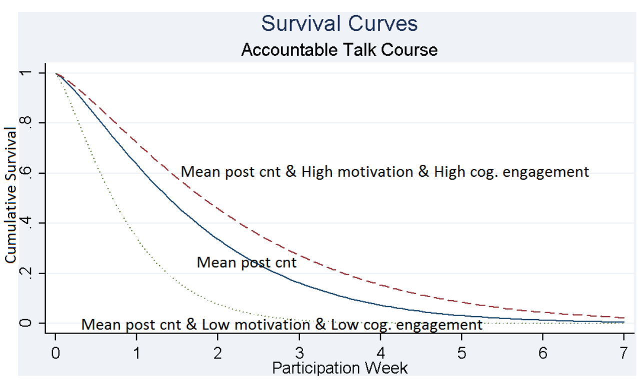

Figure 2: Survival curves for students with different levels of engagement in the Accountable Talk course.

Figure 2-4 illustrate these results graphically, showing three survival curves for each course. The solid curve shows survival with the number of posts, motivation, and cognitive engagement at their mean level. The top curve shows survival when the number of posts is at its mean level, and average learner motivation and level of cognitive engagement in the posts are both one standard deviation above the mean, and the bottom curve shows survival when the number of posts is at its mean, and the average expressed learner motivation and level of cognitive engagement in the posts are one standard deviation below the average.

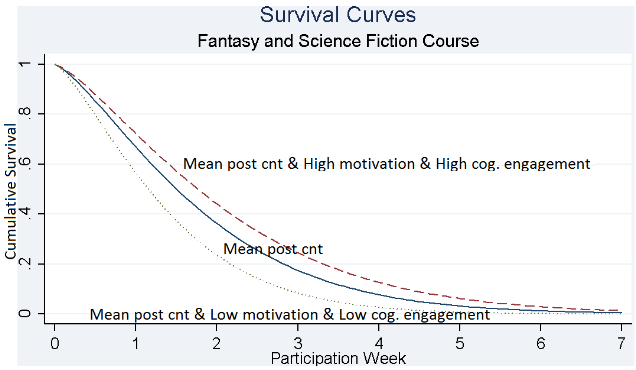

Figure 3: Survival curves for students with different levels of engagement in the Fantasy and Science Fiction course.

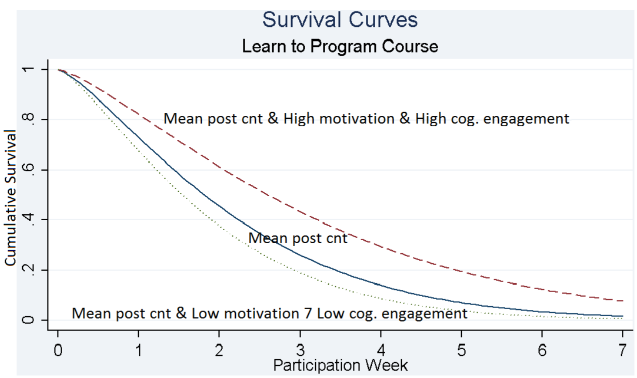

Figure 4: Survival curves for students with different levels of engagement in the Learn to Program course.

5.3 Implications

In contrast to regular courses where students engage with class materials in a structured and monitored way, and instructors directly observe student behavior and provide feedback, in MOOCs, it is important to target the limited instructor's attention to students who need it most(Ramesh et al.

[7] 32% = 100% - (100% * 0.86)
[8] 14% = 100% - (100% * 0.86)
[9] 42% = 100% - (100% * 0.58)
[10] 6% = 100% - (100% * 0.94)

2013). The automated linguistic models designed in this paper can help monitor MOOC user engagement from forum posts. By identifying students who are likely to end up not completing the class before it is too late, we can perform targeted interventions (e.g., sending encouraging emails, posting reminders, allocating limited tutoring resources, etc.) to try to improve the engagement of these students. For example, our motivation prediction model could be used to improve targeting of limited instructor's attention to users who are motivated in general but are experiencing a temporary lack of motivation that might threaten their continued participation, in particular, those who have shown serious intention of finishing the course by joining the discussion forums. One possible intervention that can be based on this type of analysis might suggest instructors reply to those with recent motivation level lower than it has been in the past. This may help students get past a difficult part of the course. We can also recommend reading the highly motivated posts to the other users, which may serve as an inspiration.

Based on the predictive engagement markers, we see it is important for the students to be able to apply new knowledge and engage in deeper thinking. Discussion facilitation can influence levels of cognitive engagement (Corno and Mandinach, 1983). The instructor can encourage learners to reflect on what and how learning addressed needs. Work on automated facilitation from the Computer Supported Collaborative Learning (CSCL) literature might be able to be adapted to the MOOC context to make this feasible (Adamson et al. in press).

6 Conclusion

We present a study on how to measure MOOC student engagement based on linguistic analysis on forum posts. We identify two new measures that quantify engagement and validate the measures on three Coursera courses with diverse content. We automatically identify the extent to which posts in course forums express learner motivation and cognitive engagement. The survival analysis results validate that the more motivation the learner expresses, the lower the risk of dropout. Similarly, the more personal interpretation a participant shows in her posts, the lower the rate of student dropout from the course forums.

6.1 Limitations and future work

An important limitation of this study is that, even though the activeness in a course's online forum closely correlates with the student's drop out of the course, exactly when (and why) the students drop out of a course entirely is not publicly accessible information. Gillanni(2013) shows that those that engage explicitly in the discussion forums are often higher-performing than their counterparts in the course. For the "invisible" users who have never interacted with other learners/staff on the discussion forums, we can only rely on click-stream data to understand their behavior. Another limitation is that even though we use longitudinal data, our findings are correlational. Student motivation is generally categorized as intrinsic and extrinsic in previous work(Ryan and Deci 2000). In our work, we did not distinguish between the two motivation types. Because in MOOC forums, we observe that there is a limited number of posts that demonstrate extrinsic motivations(E.g. taking the course for high grades or the certification). In future work, with annotated motivation types, it will be interesting to study how students with different observed types of motivation behave differently in MOOCs. We also hope to utilize social interactions in forums, such as who talk to whom information, to better understand social learning in MOOCs (Sun et al. 2011). We also plan to design and test a targeted intervention making use of the predicted engagement level, which will allow us to measure the practical impact of our findings as well as to evaluate whether the correlational evaluation we present here holds up to an experimental test of causality.

Acknowledgments

We want to thank Ryan Carlson, David Adamson and Tanmay Sinha, who helped provide the data for this project. The research reported here was supported by National Science Foundation grant 0835426.

References

Adamson, D., Dyke, G., Jang, H. J., Rosé, C. P. 2013. Towards an Agile Approach to Adapting Dynamic Collaboration Support to Student Needs. *International Journal of AI in Education 24(1), pp91-121.*.

Belanger, Y. and Thornton, J 2013. Bioelectricity: A Quantitative Approach Duke University's First MOOC. *Duke University*.

Beukeboom, C. J., Tanis, M., and Vermeulen, I. E. 2013. The Language of Extraversion Extraverted People Talk More Abstractly, Introverts Are More Concrete. *Journal of Language and Social Psychology, 32(2), 191-201*.

Beukeboom, C. J. 2014. Mechanims of linguistic bias: How words reflect and maintain stereotypic expectancies. . *Social Cognition and Communication, 31, 313-330*.

Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z. and Wong, F. M. F. 2013. Learning about social learning in MOOCs: From statistical analysis to generative model. *arXiv preprint arXiv:1312.2159*.

Carini, R. M., Kuh, G. D. and Klein, S. P. 2006. Student engagement and student learning: Testing the linkages. *Research in Higher Education, 47(1), 1-32*.

Cheng, J., Kulkarni, C. and Klemmer, S. 2013. Tools for predicting drop-off in large online classes. *In Proceedings of the 2013 conference on Computer supported cooperative work companion. ACM*.

Chi, M.T.H. 2000. Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. . *In R. Glaser (Ed), Advances in instructional psychology: Educational design and cognitive science (pp. 161-238)*.

Chi, M.T.H., and Bassock, M. 1989. Learning from examples via self-explanations. *Knowing, learning, and instruction: Essays in honor of Robert Glaser (1989): 251-282.*.

Christensen, G., A. Steinmetz, B. Alcorn, A. Bennet, D. Woods, and EJ Emmanuel 2013. The MOOC Phenomenom: Who Takes Massive Open Online Courses and Why?. *University of Pennsylvania, n.d. Web. 6 Dec. 2013* .

Clow, Doug. 2013. MOOCs and the funnel of participation. *Third International Conference on Learning Analytics and Knowledge. ACM* .

Corno, Lyn, and Ellen B. Mandinach. 1983. The role of cognitive engagement in classroom learning and motivation. *Educational psychologist 18(2): 88-108.* .

DeBoer, Jennifer, G. S. Stump, D. Seaton, and Lori Breslow. 2013. Diversity in MOOC students' backgrounds and behaviors in relationship to performance in 6.002 x. *In Proceedings of the Sixth Learning International Networks Consortium Conference*.

Dowson, M., and McInerney, D. M. 2003. What do students say about their motivational goals?: Towards a more complex and dynamic perspective on student motivation. *Contemporary Educational Psychology, 28(1), 91-113*.

Esque, T. and McCausland, J. 1997. Taking ownership for transfer: A management development study case. *Performance Improvement Quarterly, 10 (2), 116-133*.

Fahy, P.J., Crawford, G. and Ally, M. 2001. Patterns of interaction in a computer conference transcript. *International Review of Open and Distance Learning, 2(1)*.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R. and Lin, C.-J. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research (9)*.

Fuchs, L.S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K., and Dutka, S. 1997. Enhancing students' helping behavior during peer-medicated instruction with conceptual mathematical explanations. *Elementary School Journal, 97, 223-249*.

Garrison, D. R., Anderson, T. and Archer, W. 1999. Critical inquiry in a text-based environment: Computer conferencing in higher education. *The internet and higher education, 2(2), 87-105.*.

Gillani, Nabeel. 2013. Learner communications in massively open online course. *OxCHEPS Occasional Paper 53* .

Girish, Balakrishnan 2013. Predicting Student Retention in Massive Open Online Courses using Hidden Markov Models. *EECS Department, University of California, Berkeley*.

Graham, S., & Golan, S. 1991. Motivational influences on cognition: Task involvement, ego involvement, and depth of information processing. *Journal of Educational Psychology 83:187-194*.

Huang, Jonathan, Chris Piech, Andy Nguyen, Leonidas Guibas. 2013. Syntactic and Functional Variability of a Million Code Submissions in a Machine Learning MOOC. *In Proceedings of the 1st Workshop on Massive Open Online Courses at the 16th Annual Conference on Artificial Intelligence in Education*.

Keller, J., and Suzuki, K. 2004. Learner motivation and e-learning design: A multinationally validated process. *Journal of Educational Media, 29(3), 229-239*.

Kizilcec, Ren F., Chris Piech, and Emily Schneider 2013. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. *In Proceedings of the Third International Conference on Learning Analytics and Knowledge. ACM*.

Koch, Gary G 1982. Intraclass correlation coefficient. *In S. Kotz and N. L. Johnson (Eds), Encyclopedia of statistical sciences (pp.213-217).New York:John Wiley and Sons* .

Koller, Daphne, Andrew Ng, Chuong Do and Zhenghao Chen. 2013. Retention and Intention in Massive Open Online Courses. *In Depth. Educause. N.p., 3.* .

Liu, Bing., Minqing Hu, and Junsheng Cheng. 2005. Opinion Observer: analyzing and comparing opinions on the Web. *In Proceedings of WWW, pages 342-351.*.

Locke, E. A., and Latham, G. P. 2002. Building a practically useful theory of goal setting and task motivation. *American Psychologist, 57(9), 705-717*.

Milligan, C., Littlejohn, A., and Margaryan, A. 2013. Patterns of Engagement in Connectivist MOOCs. *Journal of Online Learning and Teaching, 9(2)*.

Moshinskie, J. 2001. How to keep e-learners from e-scaping. *Performance Improvement, 40(6), 30-37*.

Ng, Evelyn, and Carl Bereiter 1991. Three levels of goal orientation in learning. *In Journal of the Learning Sciences 1.3-4*.

Pennebaker, J. W. and King, L. A. 1999. Linguistic Styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77, 1296-1312*.

Poellhuber, Bruno., Normand Roy, Ibtihel Bouchoucha, Jacques Raynauld, Jean Talbot and Terry Anderso. 2013. The Relations between MOOC's Participants Motivational Profiles, Engagement Profile and Persistence. *In MRI Conference, Arlington*.

Ramesh, Arti, Dan Goldwasser, Bert Huang, Hal Daum III, and Lise Getoor 2013. Modeling Learner Engagement in MOOCs using Probabilistic Soft Logic. *In workshop of NIPS*.

Rose, C. Carlson, D. Yang, M. Wen, L. Resnick, P. Goldman, and J. Sheerer. 2014. Social factors that contribute to attrition in moocs. *In ACM Learning at Scale*.

Roscoe, R. D., and Chi, M. T. H. 2008. Tutor learning: The role of explaining and responding to questions. *Instructional Science, 36. pp. 321-350*.

Ryan, Richard M., and Edward L. Deci. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *In Contemporary educational psychology 25.1 (2000): 54-67* .

Snow, R., O'Connor, B., Jurafsky, D. and Ng, A. Y. 2008. Cheap and fast — but is it good? Evaluating non-expert annotations for natural language tasks. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 254-263)*.

Stoney, C. and Oliver, R. 1999. Can higher order thinking and cognitive engagement be enhanced with multimedia? *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning.* .

Stata Corporation 2011. Stata Statistical Software Release 7.0: Programming. *Stata Corporation* .

Sun, T, W. Chen, Z. Liu, Y. Wang, X. Sun, M. Zhang, and C.-Y. Lin. 2011. Participation maximization based on social influence in online discussion forums. *In Proceedings of ICWSM, 2011.*

Turney, P. D., Neuman, Y., Assaf, D. and Cohen, Y. 2011. Literal and metaphorical sense identification through concrete and abstract context. *In Proceedings of the 2011 Conference on the Empirical Methods in Natural Language Processing (pp. 680-690)* .

Vedder, P. 1985. Cooperative learning: A study on processes and effects of cooperation between primary school children. *Westerhaven, Groningen: Rijkuniversiteit Groningen.*.

Wang, Yi-Chia, Robert Kraut, and John M. Levine 2012. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. *In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. ACM*.

Witten, I. H. and Frank, E. 2005. Data mining: Practical machine learning tools and techniques (2nd ed.) *Elsevier*.

Wittrock, M. C. 1990. Generative processes of comprehension. *Educational Psychologist, 24, pp. 345-376.* .

Yang, D., Sinha, T., Adamson, D. and Rosé, C. P. 2013. “Turn on, Tune in, Drop out”: Anticipating student dropouts in Massive Open Online Courses. *In workshop of NIPS*.

Zhu, E. 2006. Interaction and cognitive engagement: An analysis of four asynchronous online discussions. *Instructional Science, 34(6), 451-480.*