Exploring the Effect of Student Confusion in Massive Open Online Courses

Diyi Yang
Language Technologies Institute,
Carnegie Mellon University
diyiy@cs.cmu.edu

Robert E. Kraut
Language Technologies Institute,
Carnegie Mellon University
kraut@cs.cmu.edu

Carolyn P. Rosé
Language Technologies Institute,
Carnegie Mellon University
cprose@cs.cmu.edu

Although thousands of students enroll in Massive Open Online Courses (MOOCs) for learning and self-improvement, many get confused, harming learning and increasing dropout rates. In this paper, we quantify these effects in two large MOOCs. We first describe how we automatically estimate students' confusion by looking at their clicking behavior on course content and participation in the course discussion forums. We then apply survival analysis to quantify the impact of confusion on students' dropout. The results demonstrate that the more confusion students express themselves and the more they are exposed to other students' confusion, the sooner they drop out of the course. We also explore the effects of confusion expressed in different contexts and related to different aspects of courses. We conclude with implications for the design of interventions to improve student retention in MOOCs.

1. Introduction

Massive Open Online Courses (MOOCs) enable thousands upon thousands of students to register for courses to learn at their convenience and with no monetary cost. Despite tremendous potential, MOOCs suffer from extremely high rates of attrition[1], with over 85% of students registering for a class leaving before they complete it (Jordan, 2014). Although there are multiple reasons for this high dropout, an important one is learners' confusion and frustration, expressed either explicitly or implicitly (Conole, 2013). Here, we define *Confusion* as a state in which a student hits an impasse and is uncertain of how to move forward. Previous explorations of student affect in educational contexts have demonstrated a strong connection between affect, engagement, and learning (Arroyo et al., 2009). However, little systematic research has examined how confusion affects students' participation in online courses, where dropping out is much easier than in traditional classrooms.

Confusion has a complex influence on learning and engagement. Confusion experienced during learning can be associated with positive as well as with negative outcomes. In certain

[1] http://www.katyjordan.com/MOOCproject.html

circumstances, a prompt response to confusion (e.g., support from an instructor or help from other sources), or simply students' experience of overcoming confusion through their own efforts, may lead to a beneficial effect (DMello et al., 2014). On the other hand, students who experience confusion may struggle to stay involved in a course and ultimately drop out, especially if they are at risk of dropping out for other reasons. The focus of our work is to better understand student participation as they struggle, express confusion, and ultimately stop participating in a MOOC.

Existing studies on confusion in MOOC context mainly target at students who actively participate in course discussion forums (Agrawal et al., 2015; Yang et al., 2015). However, forum students are only a small portion of course participants and have different levels of course activities compared to students that are not active in the forums. Behavior traces can help identify periods of confusion and both its causes and consequences (Lee et al., 2011; Baker et al., 2012; Gupta and Rosé, 2010). For example, students may become confused when watching lecture videos or working on assignments. They might express their confusion via the discussion forum with detailed descriptions of the problem (e.g. "I'm stuck on this question" and "I'm very lost on this concept"), or through traceable interactions with course features, for example, re-watching a video or slowing the video speed. Without receiving help from other students or teaching staff, students might remain confused. Persistent confusion might promote low self-efficacy (Caprara et al., 2008) and negative attitudes toward the course and ultimately lead to dropout. On the other hand, instructors' ability to monitor confusion from behavioral traces could be leveraged to offer just-in-time support for confused students, and this support might increase the overall course success.

In this work, we explore ways to measure students' confusion in MOOCs and the impact of confusion on retention for two different student populations: (1) students who are active in discussion forums and (2) all course participants, regardless of their discussion activity. Our investigation of confusion is conducted in two-stages. The first stage is the development of an automated measure of confusion, which we can apply to all sessions in which students in a MOOC participate. The approach starts with a forum-level confusion classifier, a machine learning model that predicts confusion trained on a human-annotated sample of forum post. The predictive features include both students' language behavior in posts and their video watching behavior. We applied this model, based on hand-annotated posts, to estimate confusion in all posts and then used the predicted confusion students revealed explicitly in their posts to train a second machine learning model to predict the confusion they revealed implicitly in their non-posting behavior during a MOOC session. We call this the session-level confusion classifier. The forum-level and session-level estimates of confusion correlate moderately well. This mapping provides a mechanism to acquire confusion labels for students without posts. We present both quantitative and qualitative analysis to validate the prediction performance of both the forum-level confusion classifier and session-level one. The second stage of our research uses survival analysis to examine the effects of confusion on students' continued participation.

The contribution of our work is three-fold. We developed a bootstrapped procedure that uses a confusion measure based on the language students revealed explicitly in posts to learn how the vast majority of students, who never post, reveal confusion in non-language behavior. Empirically, we identified several important connections between student confusion and their dropout from MOOC courses. Practically, our findings provide guidance for the development of interventions that could promote retention by providing just-in-time support for confused students.

2. Related Work

Analysis of Attrition in MOOCs

The growth of MOOC platforms such as Coursera and edX has raised hopes about the potential of distance and lifelong learning, hopes that undercut by the extremely high attribution rates (Koller et al., 2013). A cottage industry of research on MOOC dropout rates reflects the strong concern about attrition (Anderson et al., 2014; Yang et al., 2013). For example, Rose et al. (2014) investigate the influence of social positioning and sub-community participation on student commitment to MOOC courses, and Yang et al. (2014) demonstrate that students who share similar behavior patterns influence other students' commitment to the courses. To investigate whether models trained on previous courses would perform well on a new course, Boyer et al. (2015) used transfer learning techniques to predict dropout and transfer knowledge across courses. These analyses focus on social connection and social positioning, but do not deeply explore student affect. Work from Ramesh (2013) incorporates content analysis to investigate the correlation between sentiment and subjectivity of users' posts and their engagement. In a finer-grained content analysis work, Wen et al., (2014b) used sentiment from forum posts to monitor students' opinions towards the course, and found a correlation between sentiment and student attrition. Additionally, Wen et al. (2014a) have used computational linguistic models to measure learner motivation and cognitive engagement from forum posts and their association with student attrition. Here we see an expected pattern of motivation and cognitive engagement consistently associated with commitment, but weak or inconsistent effects associated with simple measures of sentiment. Although this prior work provided a multi-faceted understanding of student engagement through examination of their forum posts, little research has dealt specifically with understanding students' experience of confusion while participating in MOOC courses.

Confusion and Learning

Students experience confusion when they are confronted with an anomaly, contradiction, or an impasse, and are uncertain about how to proceed. Confusion has a mixed influence on students' learning. On the positive side, confusion, as one of the most frequently studied forms of student affect (Lehman et al., 2008), has been found to correlate with learning (Lehman et al., 2012), particularly with learning at deeper levels of comprehension. Confusion causes students to stop, reflect, and begin active problem solving to resolve their confusion. Struggling with confusion as a cognitive activity may enable learners to acquire a deeper understanding of complex topics (Lehman et al., 2012; DMello et al., 2014). Confusion can have positive effects when students can effectively regulate their confusion or if the learning environment provides sufficient scaffolding to help them do so (DMello et al., 2014; Lehman et al., 2012; Pardos et al., 2013).

On the other hand, prolonged confusion reduces student achievement (Lee et al., 2011). (DMello et al., 2014) also emphasizes the importance of rapidly resolving confusion when it arises by quickly providing explanations and other scaffolds. Confusion during non-remediation problem solving is also negatively correlated with student achievement (Pardos et al., 2013). Lee et al. (2011) explored the relationship between novice programmers' confusion and achievement based on logs from their programming courses, and concluded that prolonged confusion is associated with lower student achievement. Furthermore, confusion associated with failure to resolve an impasse (DMello and Graesser, 2012) might transition into frustration, boredom, and ultimate disengagement from the learning process (Larson and Richards, 1991).

In the MOOC context, the distant nature and the size of MOOCs introduces limitations on opportunities for students to interact with others as effectively as in traditional classroom learning or intelligent tutor situations (VanLehn et al., 2003; D'Mello et al., 2010). Lacking immediate feedback, interactive communication, or timely support increases the likelihood of members leaving such communities (Wang et al., 2012), especially when members get confused in the learning process. To be more successful, MOOCs require new interventions to provide confused students with feedback, interaction and even instructor interventions (Ramesh et al., 2013). For example, Agrawal et al. (2015) introduces an instructional aid that tries to detect and address confusion in forum posts. Such intervention depends on an ability to monitor behavior traces to identify the experience of confusion as well as an understanding of how confusion plays out in the context (Howley et al., 2015).

Modeling Confusion and its Relationship to Dropout

In order to regulate student confusion and ultimately increase their engagement, first we must be able to identify confusion in learner experiences (Lehman et al., 2012; Lehman and Graesser, 2015) and use these measures to determine the connection between confusion and dropout. Considerable research has studied automatic identification of confusion in the context of educational software such as intelligent tutoring systems (Baker et al., 2012). Researchers have investigated student affect changes, including confusion through approaches such as classification or knowledge engineering (Pekrun et al., 2002). Most successful affect detectors are built through physiological sensors, vocal patterns identification (Calvo and D'Mello, 2010) or through log files (Baker et al., 2012; Ocumpaugh et al., 2014), and conversation cues (DMello et al., 2008). In addition, there are some studies (D'mello and Graesser, 2007; Bosch et al., 2014; Liu et al., 2005) that investigate the automatic detection of a learner's affective states from posture patterns, facial expressions, dialogue features, or fusion of several cues (Nicolaou et al., 2011), obtained from the interaction with intelligent tutoring systems. For instance, Hussain et al. (2012) presented a model to detect learners' affective states using multimodal cues, i.e. using video and physiological signals. Similarly, Conati et al. (2009), developed a detector based on a combination of questionnaire and log data to predict self-reported student affect. Their model was better at identifying focused and curious students, but less successful at identifying students who were confused. However, approaches relying upon sensors or self-reported affect (Wixon et al., 2014) can only be used on datasets for which such signals are present, and also have limitations in terms of large-scale deployment. We address these limitations in our work in the MOOCs by relying only on information readily available in that context.

Generalization and Scalability

Recently, techniques involving clickstream level and language level analysis are popular and are widely used in the context of educational data mining field (Werner et al., 2013), such as understanding students' language choices via dialogue acts (Ezen-Can and Boyer, 2015) and exploring log file level analysis to investigate students' course performance (Saarela and Kärkkäinen, 2015). But the success of these approaches depends upon having sufficient data from a representative sample of students. While this constraint can be met in classroom studies, where students are compelled to participate in all learning activities, it is more difficult to meet in studies of MOOCs. Although watching videos is the most popular activity in MOOCs, many students fail to participate in the more active learning activities they offer, including the forums, quizzes, and assignments they offer (Koedinger et al., 2015; Jordan, 2014) or fail to complete surveys. Several research studies attempt to predict or understand student dropout or develop interventions by relying on data from these types of source(Wen et al., 2014a; Yang et al., 2013; Ramesh et al., 2013; Alario-Hoyos et al., 2013). For example, Kizilecec and colleagues (2015) developed a 13-item questionnaire derived from open-ended responses to capture learners' motivations for attending courses, but this analysis is limited to students who volunteered to complete the survey. Chaturvedi et al. (2014) targeted students who have forum posts and predicted instructor interventions in MOOC forums. There are some works targeting students who watch videos, which is a very large population in the MOOC context. For example, Kim et al. (2014) reports a large-scale analysis of in-video dropout and peaks in viewership and student activity. We acknowledge this trade-off between data quantity and model accuracy, which sets an explicit challenge for research that tries to make predictions in the presence of limited data.

Our research goal is to develop measures of student confusion even for students who are less active in MOOCs. We use both clickstream level mining techniques and automated language analysis to detect confusion among the most active students and then generalize these models to apply to the less active ones. Although, as we will show, our ability to accurately measure confusion depends upon the richness of students' behavior traces, with accuracy is highest for students who are involved with the widest range of activities, we are nevertheless able to make useful predictions for the less active students as well. Trade-offs between accuracy and completeness are discussed later in this work.

3. Research Questions

As we have argued, confusion is an important factor in the experiences of students in MOOCs. Active students can demonstrate their confusion in the discussion forum by making explicit posts to state their difficulties in the learning process (Yang et al., 2015), but as previously indicated the vast majority of students don't participate in course forums. This motivates us to devote our efforts to not only the most committed students with forum activities, but also to a broader group of students who have little or limited forum activities.

Our research questions in this work can be summarized as follows:

1. Can we design a mechanism to annotate the confusion of all course participants, regardless of whether students have ever made a post (or viewed the forums?) or not?

2. Can we automatically build machine learning models to identify student confusion based on evidence from their course participation histories?

3. How do different types of confusion influence students' continued participation, both overall and in relation to an array of specific course contexts?

It is relatively straightforward to measure the explicit confusion conveyed in students' posts. We first annotated a set of explicit posts in discussion forums to construct an automated measure of confusion. We then build a forum confusion classifier to predict the degree of confusion associated with student posts from students' language and their interaction with content in the hour surrounding their posts.

However, it is much more difficult to assess the confusion of students who have never made posts. To solve this problem, we used automated estimates of confusion in forum posts as ground truth. That is, we applied the forum classier model to estimate the confusion expressed in all forum posts, and then map students confusion expressed explicitly in their posts to their confusion expressed implicitly during a time session. Using this estimate as the ground truth, we build a session confusion classifier. The session confusion classifier used 17 categories of clickstream data but no language cues. We then applied the session confusion classifier to all sessions that a student participated in.

In the final section of the paper, we present both quantitative and qualitative analysis to validate the prediction performance of both the forum confusion classifier and the session confusion classifier. Our investigation demonstrates that expression of post confusion and being exposed to others post confusion influences students continued participation, and students' session confusion and reading others' posts expressing confusion also affect student retention.

Specifically, Research Question 1 will be answered in Section 5.2.1. and Section 5.3.1.; Research Question 2 will be answered in Section 5.2.3. and Section 5.3.2.. Research Question 3 will be answered in Section 6..

4. Data Preparation

In our investigation, we partnered with faculty at a well-known West Coast state university, which provided the data from two Coursera MOOCs. Our dataset for this paper consists of two Coursera courses: one mathematics course, "Algebra" and one economics course "Microeconomics". Algebra had 2,126 active users (active users refers to those who post at least once in a course forum) and 7,994 forum posts; Microeconomics had 2,155 active users and 4,440 forum posts. We excluded posts that lie in the Study Group subforums. A flag that indicates whether the issue was resolved or not was also provided for each thread. The duration of each of the two courses was 12 weeks. Besides forum records, each student's interaction (student clicks) with the course materials was also recorded via clickstream. Algebra has 8,686,230 student clicks, and Microeconomics has 2,709,053 clicks. This clickstream data provides us with the opportunity to investigate the relationship between click patterns and student confusion. Around 3.4% students in Algebra course and 8.3% students in the Microeconomics have posted at least once in the forum. In this work, we classify the confusion of not only students who have posting behavior, such as views or active posts, but also students who have no forum viewing records.

5. Confusion Prediction

Students expressed their confusion explicitly, in their posts, or implicitly, as they read others' posts and interacted with the course's video, quizzes, homework and other material. In order to offer the opportunity to identify students who are experiencing confusion, we built machine learning models to automatically identify the level of confusion expressed in students' posts and their implicit session confusion experienced in their learning processes. Such models use statistical procedures to map a set of input features to a set of output categories. In our work, we extract the input features both from students' click behaviors and their posting behaviors, including clicking patterns in courses, presence of domain-specific content words, and high-level linguistic features. The output is a numerical value indicating the degree of confusion expressed in each post or each time session. In these following sections, we first discuss how we created a human annotated corpus to develop an automated measurement of student confusion. Next, by mapping direct confusion measurements to other course behaviors, we were then able to build implicit and explicit classifiers of confusion, which we evaluated for accuracy.

5.1. Overview of Confusion Prediction

This section presents the overview of our confusion prediction task. In detail, it will be conducted in several steps: Forum Confusion Labeling, Forum Confusion Classifier, Generalization to All Posts, Session Confusion Labeling, Session Confusion Classifier and All Sessions with Labels, as shown in Figure 1. Forum confusion labeling (Section 5.2.1.) describes how to annotate the confusion contained in students' posts. We sampled a subset of forum posts and asked annotators to judge the confusion expressed in those messages. Based on this annotated dataset, Forum Confusion Classifier component (Section 5.2.2. and 5.2.3.) constructs sets of features and trains a classifier to predict the confusion contained in each post. If those classifiers achieve high accuracy, we can then apply them to estimate the confusion scores of all posts in the forum. Session Confusion Labeling in Section 5.3.1. describes how we map the confusion of all posts in a session into the confusion score of that session. Based on those sessions that have confusion labels, we train a Session Confusion Classifier (Section 5.3.2.) with constructed click pattern features. With acceptable accuracy of session confusion prediction, we can apply the session classifier to all sessions of all participants in a course.

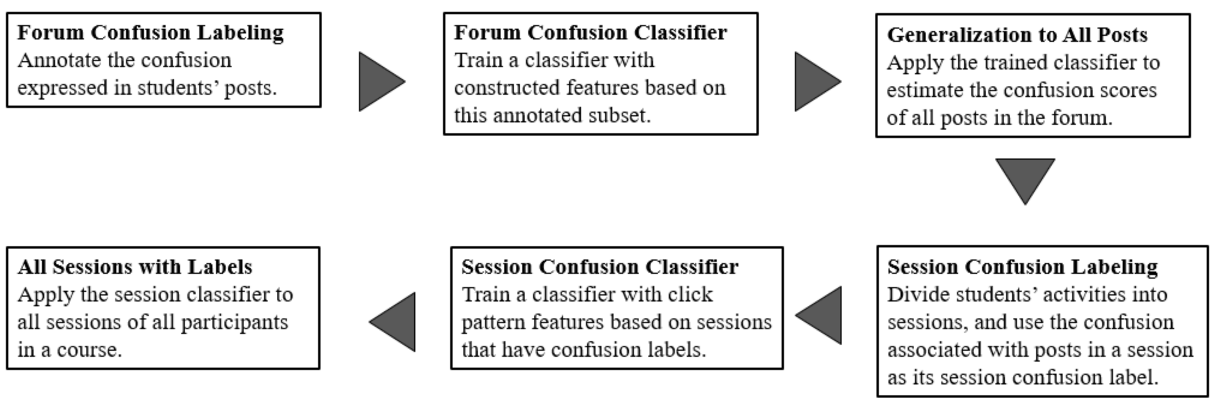


Figure 1: Overview of Forum and Session Confusion Classification

5.2. Predicting Forum Confusion

This section presents how we create the human annotated corpus for measuring student forum confusion and how different sets of features are designed to represent confusion, as well as qualitative and qualitative evaluations of confusion prediction in the discussion forums.

5.2.1. Forum Confusion Labeling

We used Amazon's Mechanical Turk (MTurk) to construct a reliable, hand-coded dataset to serve as the ground truth for a forum classifier that automatically measured the confusion students expressed in their posts. Amazon Mechanical Turk[2] is a Internet crowdsourcing market, which allows requesters advertise tasks, known as HITs (Human Intelligence Tasks), and workers (Turkers) to select tasks and complete them for a monetary payment set by the requester. Buhrmester et al. (Buhrmester et al., 2011) demonstrated that MTurk data quality matched or surpassed psychometric standards of traditional studies, for simple tasks. In addition, Snow et al. (Snow et al., 2008) showed that the combined ratings of five to seven Turkers yielded judgments of textual content, such as expressions of emotion, event timing, similarity of words, disambiguation of word meaning, and language-based entailment/implication, comparable to judgments made by experts.

To increase the annotation quality, we required Turkers to have a United States location and a 98% approval rate for their previous work on MTurk. We randomly sampled 522 posts from the Algebra course forum, and 584 posts form the Microeconomics course forum. Non-English posts were filtered out, and we replaced personal identifiers with anonymous unique identifiers. We obscured personal details of student experiences in order to preserve the privacy of the students in the course. For each post, Turkers judged the level of confusion contained in the message on a 4-point Likert scale with the labels “No Confusion”, “Slightly Confused”, “Moderately Confused” and “Seriously Confused”. We provided them with explicit definitions and examples to use in helping their judgments. Each post was labeled by five different workers. To encourage workers to take the rating task seriously, we asked Turkers to highlight the portion of the post’s text that supported their judgments. Turkers received $0.06 for rating each post.

We aggregated the 5 workers’ responses for each post by averaging their ratings. Thus, each post has a 1-4 scale numerical value indicating the level of confusion contained in the message. Following are two examples from our final hand-coded datasets, with one example illustrating serious confusion and another indicating slight confusion.

- Student Confusion = 4.0 (Seriously Confused)

  I am completely lost on quiz 3. (I’ve never been awesome at math of any kind) Quiz 1 &2 were easy breezy, but 3 is throwing me. I feel like the video barely touched on what this is asking for. Or maybe I am just *that* bad at this...? Anyone want to help me understand better?

- Student Confusion = 2.0 (Slightly Confused)

  Hi, I am having problems running coursera on Chrome from my laptop. I do not have this problem using Explorer or Chrome from another PC. Have anyone else has the same issues?

The intra-class correlation coefficients, evaluating the reliability of the annotations, 0.745 for Algebra, and 0.801 for Microeconomics, indicate good agreement among judges. Moreover the average of the Turkers’ judgments was strongly correlated with the average rating for 50 posts made by three experts (r=0.86 and 0.80 on the Algebra and Macroeconomics course respectively). To provide positive and negative training instances for classification, we divided the data into“confused” posts and “unconfused” posts based on a median split.

[2] https://www.mturk.com/mturk/welcome

5.2.2. Classifier Construction: Feature Space Design

This part describes the different types of features associated with forum confusion. The features we used to predict the human annotations of posts included the language students used in their posts and the type of materials students clicked on in the three hours before their post: video lectures, quizzes, forums and other course materials.

Linguistic Features

The Linguistic Inquiry and Word Count (LIWC), designed by Pennebaker et al. (2001) calculates the degree to which people use different categories of words across a wide array of texts, including emails, speeches, poems, or transcribed daily speech. LIWC dictionaries were selected based on their relevance to confusion affect. For example, "I, my" reflects that the student is describing something related to himself/herself, whereas negative terms such as "depress, struggle" express a negative affect related to learning. Negation words or phrases such as "not", "shouldn't", or "did not" might also act as a proxies for potential confusion. Similarly, insight words such as consider or think reflect students' active orientation towards learning. To summarize, we incorporated 15 dictionaries from LIWC into our machine learning models as listed below, in the format of the dictionary name and its representative words: *first person singular pronouns* ('I, my, me'), *first person plural pronouns* ('we, our, us'), *second person pronouns* ('you, you'll'), *third person pronouns* ('she, their, them'), *positive emotions* ('happy, pretty, good'), *anxiety or fear* ('nervous, afraid, tense'), *anger* ('hate, kill, pissed'), *sadness or depression* ('cry, sad'), *insight* ('think, know, consider'), *discrepancy* ('should, would, could'), *negations* ('no, never, not'), *assents* ('yes, ok, mm, hmm'), *adverbs* ('very, much'), *certainty* ('always, never'). *nonfluencies* ('uh, rr, hm').

Question Features

Confusion is often conveyed by asking questions (Wilson, 1989). Thus, we counted the number of question marks in the sentences as a reliable proxy for identifying questions. In addition, because not all questions are asked directly and end with a question mark, we took advantage of a heuristic rule to detect question sentences (i.e., whether the sentences start with a modal verb or a question word (Wang et al., 2012)). For example, sentences beginning with "is, was, had, does, can, may, which, why", etc., might indicate a question and the student's confusion. Additionally, we calculated whether sentences begin with a confusion related fixed expression such as " I am confused", "I was stuck" or "I am struggling with".

Click Patterns

Clickstream data in MOOCs describes students' complete interaction with each course component, and can be regarded as a shallow but ever present indicator of students' experiences. Although those traces of log data can help reveal students' learning patterns, researchers are only just beginning to investigate how click patterns in MOOCs correlate with student engagement in a meaningful way. (Sinha et al., 2014) investigates students' information processing behaviors while interacting with video lectures. Effective analysis of such click patterns helps to better identify students' confusion when combined with their language choices in discussion forum posts, and also may eventually provide the means for us to analyze confusion in students enrolled in MOOCs even without their participating in discussion boards. For example, imagine a student becomes confused about a concept when watching class videos. Even though he/she did not make an explicit post about the issue, his/her click behaviors, which may include activities such as re-watching and referring to course lectures, may suggest potential confusion.

The clickstream data allow us to distinguish the type of material students with which students were interaction before making a forum post: taking quizzes (quiz), watching lectures (lecture), participating in forums (forum), and viewing other course materials (course). We examined sequences of these behavior in the three hour window before the student has made a post, retaining the most frequent 20 sequences of of up to four behaviors. We set the maximum sequence-length to be four to avoid sparseness. For example, one student had a click pattern of "quiz-quiz-forum" before making a post "*I try to submit the quiz w1: [...] but it says wrong. Could someone tell me what is wrong here?? Thank you*". In this case, accessing the forum after taking the quiz suggest that the student encountered problems during the quiz and turned to the forum to seek help from peers. Similarly, another student showed the pattern "quiz-lecture-quiz-lecture," which suggests this student referred to the lectures for clarification and help after working on a quiz.

5.2.3. Forum Confusion Classification Method

We began by comparing the accuracy of several classifiers' performance in identifying the degree of confusion expressed in each discussion forum post. We trained several Logistic Regression classifiers to identity whether a forum post expressed confusion or not. We evaluated their performance in terms of Accuracy and Cohen's Kappa using a 10-fold cross validation: a unigram[3] (bag of words[4]) feature representation as a baseline feature set (**Unig**), classifiers using click patterns **Click**, linguistic features **Ling**, and question features **Question** individually, a classifier **CLQ** combining Click, Ling, and Question feature sets, and a classifier **CLQ+Unig** taking all the combination of previous feature sets. Our classifiers are trained on one course at a time. We also conducted feature selection to reduce the high dimensionality. The reduced model is denoted as **Reduced(CLQ+Unig)**. Because the labeled data for both Algebra and Microeconomics courses had been divided into two balanced sets, baseline accuracy of a random model would be 50% and Kappa would be 0.

5.2.4. Result Discussion: Quantitative Discussion

Performance results are shown in Table 1 and Table 2. **CLQ+Unig** achieves better performance than models comprising individual types of features Unig, Click, Ling, Quest and CLQ. In contrast, the performance of the reduced CLQ+Unig model is superior to that of the full CLQ+Unig feature set. Table 3 shows some important features identified by the **CLQ+Unig** model on the two different datasets. This is consistent with our initial analysis that confusion is closely related to viewing quizzes and exams.

Given that the Reduced(CLQ+Uing) model achieves adequate levels of accuracy of 80.3% on the Algebra data set and 71.2% on the Microeconomics data, we then applied the model to the task of predicting the confusion degree of all remaining posts in the two courses.

[3]This is the bag of words representation, i.e. use each word as a feature and count the number of occurrences of each word in a message. The dimension of unigram features is the size of word vocabulary.

[4]https://en.wikipedia.org/wiki/Bag-of-words_model

Classifiers	Accuracy	Kappa	#Features
Unig	0.791	0.581	2967
Click	0.587	0.176	20
Ling	0.647	0.293	15
Question	0.683	0.370	4
CLQ	0.723	0.448	39
CLQ + Unig	0.796	0.582	3006
Reduced(CLQ + Unig)	**0.803**	**0.606**	632

Table 1: **Performance Comparison of Different Forum Confusion Classifiers on Algebra Course**. Here, Unig denotes unigram features; Click means Click Patterns; Ling refers to linguistic features; Question denotes Question Features. The best performance is highlighted in bold.

Classifier	Accuracy	Kappa	#Features
Unig	0.698	0.375	4921
Click	0.567	0.049	20
Ling	0.594	0.056	15
Question	0.652	0.333	4
CLQ	0.678	0.309	39
CLQ + Unig	0.706	0.388	4960
Reduced(CLQ + Unig)	**0.712**	**0.403**	650

Table 2: **Performance Comparison of Different Forum Confusion Classifiers on Microeconomics Course**. Here, Unig denotes unigram features; Click means Click Patterns; Ling refers to linguistic features; Question denotes Question Features. The best performance is highlighted in bold.

Courses	Algebra	Microeconomics
Most Important Features (Feature Weight)	question marker count(1.16) 1st pers singular (1.31) question word count(0.52) click pattern (0.38) impersonal pronouns (-0.15) certainty (-0.17) negation (-0.19) adverbs (-0.20)	question marker count(1.30) start with modal words (1.09) 1st pers singular(0.73) question word count(0.17) adverbs (-0.17) affect (-0.18) click pattern(-0.19) negation (-0.20) insight (-0.28)

Table 3: Features with highest weights associated with forum confusion classifier. Here, *certainty, negation, adverbs, insight, affect* are linguistic categories in LIWC (Pennebaker et al., 2001). *start with modal words* refers to whether a sentence begins with modal verb. *question word count* represents the number of question words in a sentence (Wang et al., 2012).

5.2.5. Performance Discussion: Qualitative Findings

Course Differences: Considering that Algebra and Microeconomics can be considered technical and less technical courses respectively, we compared the average confusion students expressed in the two course discussion forums in Figure 2. We found that students expressed more confusion in Algebra and less confusion in Microeconomics. For example, 28% posts in the Algebra course have an average confusion score larger than 0.90, while that score is 6% in Microeconomics. Besides, 54.6% posts in the Algebra course have expressed confusion (having a confusion score larger than 0.5) and that of Microeconomics is 35.1. A possible explanation might be that technical courses focus more on problem solving, where confusion is natural to discuss as students work through the problems. Nontechnical courses require less problem solving, and thus confusion might be more indirectly expressed via discussion rather than through direct questions.

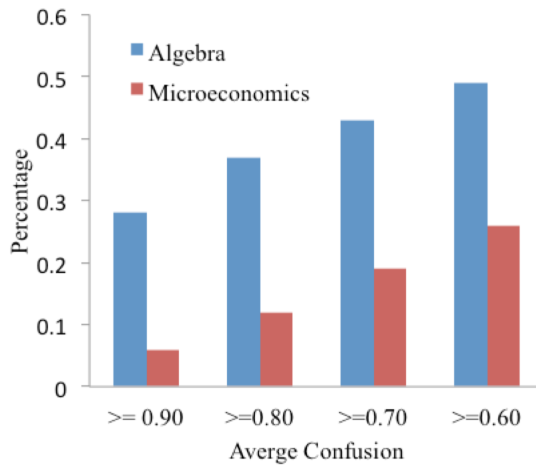
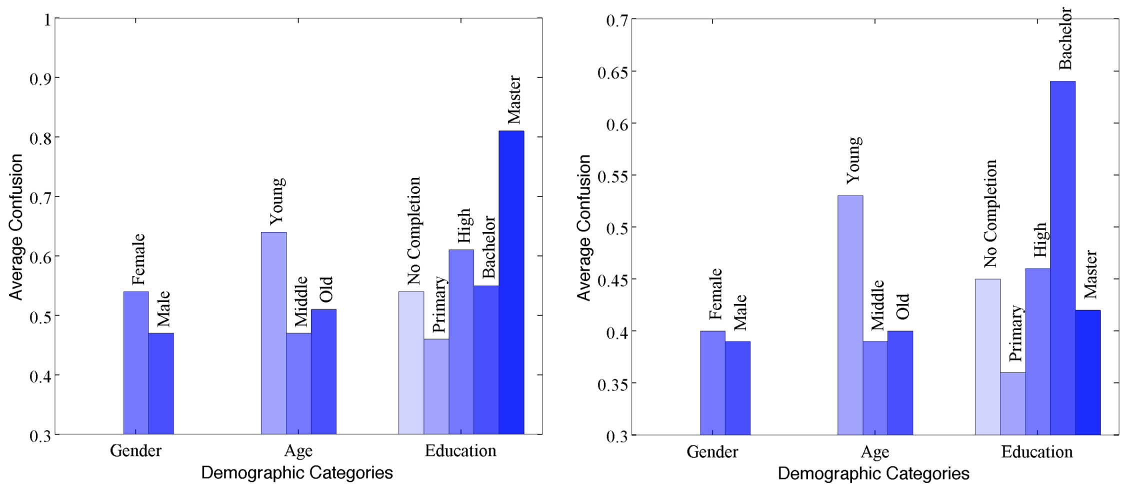


Figure 2: **Qualitative Forum Confusion Comparison over Two Courses**. It presents what percentage of forum posts have an average confusion score over 0.90, 0.80, 0.70 and 0.60.

Demographic Differences: We also examined how confusion varies across different demographic groups, in terms of *Gender*, *Age* and *Highest Education Level*. The average confusion degree comparisons over different demographics are presented in Figure 3(a) and Figure 3(b). We found that females tend to express more confusion than males; young people have the highest expressed confusion; and middle-aged people are the least likely to express confusion. In addition, students who achieved higher education degrees shared a relatively low level of confusion compared to students whose highest education level was primary school or no completion.

5.3. Predicting Session Confusion

Based on the forum confusion labels, in this section, we present how we acquired the labels the amount of confusion students expressed in a interaction session in the course by bootstrapping the forum session results, the features used to predict session confusion, and the performance and validity of the resulting session confusion classifier.



(a) Algebra Course (b) Microeconomics Course

Figure 3: Qualitative Forum Confusion Comparison over Gender, Age and Education. Y axis represents the average confusion score, ranging from 0 to 1. X axis represents sub-categories of gender, age and education, where labels are obtained from students' profile data.

5.3.1. Session Confusion Labeling

Clickstream data shows the complete history of students' interactions with different course materials. Figure 4 is an anonymized record from the clickstream data. The *username* field shows who made a click, *timestamp* shows when, *page url* shows the page the student clicked, *from* shows the previous web page, *key* shows the type of action. The *key* field identifies four action types: actions applied to *Video*, *Quiz*, *Forum*, and *Pageview*, while the *value* structure provides more details about the action applied to an object. For example, in the data record in Figure 4 shows that the student was watching the lecture 53 video and her last click was also in lecture 53.

We processed this clickstream data to identify relatively contiguous sessions session, with the goal of identifying the amount of confusion students' experience during a session. **Session** is defined as a time interval where a student's activities happen. Figure 5 illustrates how we divide a student's click activities into a series of sessions. A session is defined as a set of user clicks separated from preceding or following sessions by a gap time of at least time t when no clicks occurred. When students post within a session, we can estimate a **Session Confusion** score by averaging the confusion of posts that are posted by that student.

5.3.2. Classifier Construction: Feature Space Design

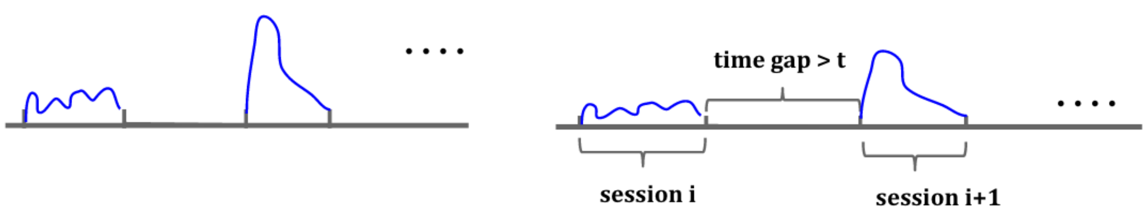
The Session Confusion score can then provide the ground truth for a machine learning classifier to predict students' confusion within a session from their activities during the session. We used the following activity features as input to a Logistic Regression model to predict session confusion scores.

1. *Video Count*: how many times a student clicks the lecture video in a given session.

2. *Quiz Count*: the number of quizzes a student has worked on during that session.

```
{
    "key": "user.video.lecture.action",
    "value": "{
            \"currentTime\":29.84013557434082,
            \"playbackRate\":1,
            \"paused\":false,
            \"error\":null,
            \"networkState\":2,
            \"readyState\":4,
            \"eventTimestamp\":1360680004982,
            \"initTimestamp\":1360679962105,
            \"type\":\"play\",
            \"prevTime\":29.84013557434082}",
    "username": "b88bf6c220d5177cc7609627c4505d49a2b097a0",
    "timestamp": 1360680006185,
    "page_url":"https://class.coursera.org/algebra-001/lecture/view?lecture_id=53",
    "client": "spark",
    "session": "3146941305-1360679731844",
    "language": "en-ca",
    "from":"https://class.coursera.org/algebra-001/lecture/view?lecture_id=53",
    "user_ip": "---.---.---.---",
    "user_agent": "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.0; Trident/5.0)"
}
```


Figure 4: **An Anonymized Example Click Record**. This presents the format of click-stream data. Specifically, *username* field shows who (anonymized) made a click, *timestamp* shows when, *page url* shows the page the student clicked, *from* shows the previous web page, *key* shows the type of action. *key* field identifies four action types: actions applied to *Video*, *Quiz*, *Forum*, and *Pageview*, while the *value* structure provides more details about the action applied to an object.



(a) A student's click activities over time

(b) Divide a student click activities into sessions

Figure 5: **Division of Student Click Activities into Sessions**. A new session begins when the time interval between two successive inter-transaction clicks adds up to time t. Here, we define t as three hours.The visited contents and activities during each session are considered to be part of that session.

3. *Quiz Attempt Count*: the times a student has attempted to solve quizzes during that session.
4. *Quiz Submission Count*: the quiz submission times during that session.
5. *Forum Count*: how many times a student has browsed the forum in a session.
6. *Pageview Count*: the number of times a student browses course pages in a session.
7. *Total Click Count*: the total number of click actions occurring in that session.

8. *Video/Quiz/Forum/Pageview Duration*: how long a student has watched the lecture video/worked on quizzes/viewed forum threads/browsed course pages on average.

9. *Total Duration*: the total time a student has devoted to four actions during a session.

10. *Average Duration*: the average duration of different actions in a session.

11. *Maximum/Minimum Duration*: the longest/shortest time a student has spent on a single action.

12. *Click Patterns*: the combination of all click patterns consisting of action quiz, forum, video and pageview.

13. *Play Count*: the number of times a student plays the video in a session.

14. *Pause Count*: the number of times a student pauses the video in that session.

15. *Seeked Count*: the number of times a student seeks the video in a session.

16. *Ratechange Fast*: the times of changing the video play rate to fast in a session. We compare the play rate of the click event that occurs just before the click and students' current play rate, to determine whether the student has sped up/slowed down the playing speed.

17. *Ratechange Slow*: the times of changing the video play rate to slow in a session.

5.3.3. Session Confusion Classification: Method and Result

To predict the session confusion using our constructed features, we used Logistic Regression with 10-fold cross validation. The training data consist of 3,540 sessions with confusion scores for the Algebra course and 2,005 labeled sessions for the Microeconomics course. To make the classifier easier to interpret, we converted the continuous Session Confusion Score (a real number in the range of 0 to 1) to a binary value (i.e. confused and unconfused session), based on a median split within course.

Performances in terms of Accuracy and Kappa are presented in Table 4. Note that none of these predictors in our session classifier include the text of students' forum posts. Thus the resulting model can be applied to all of students' sessions whether or not they include posts. The accuracy of the session confusion classifier is low, but substantially greater than chance. Despite its modest accuracy, this classifier provides confusion estimates for the 92% of students who never participated in course forums. Then we apply this classifier to predict the session confusion scores of all unlabeled sessions of course participants, which enables us to track each student's confusion during his/her participation in a course.

Courses	Accuracy	Kappa	Features
Algebra	0.657	0.261	46
Microeconomics	0.585	0.103	46

Table 4: Session Confusion Classification Performances on Two MOOCs

5.3.4. Result Discussion

In this section, we aim at validating whether estimates of students' confusion based on the predicted Session Confusion score make sense and help explain students' commitment to and performance in the course. We first compared students with and without posting behaviors by investigating the two groups' click patterns. Next, we checked the correlation between our identified session confusion and students' course performance. Lastly, we tested the validity of our hypothesis that confusion is an individual process.

Click Patterns Associated With Confusion

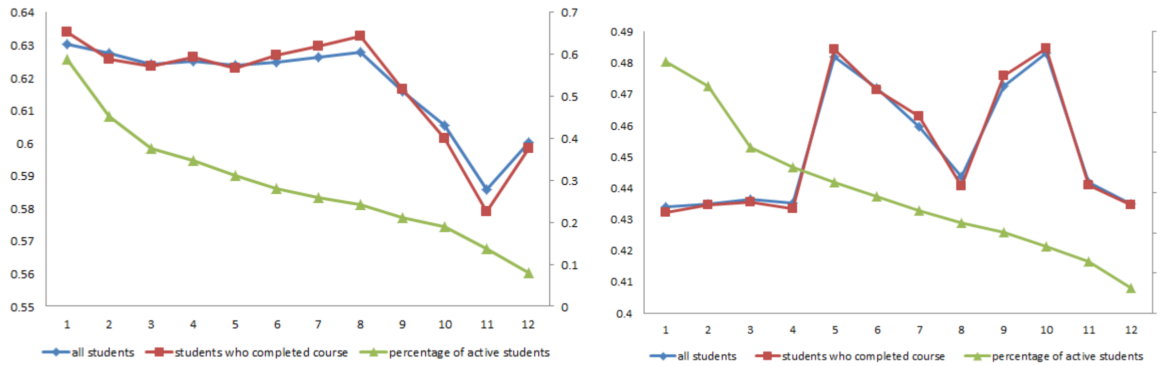
In this section, we showed the top ten ranked click patterns associated with confusion, and validated that learners who post in the forums have similar click patterns when they express confusion, compared to learners who do not post. For this purpose, we use Exact Pattern Mining method (Han et al., 2007) to mine the click patterns associated with confusion in two different student groups. The length of pattern is limited to 3. "A $\rightarrow$ B" denotes a student did B directly after doing A and it results in high confusion. The mined patterns associated with two groups are compared in Table 5. Based on the results, students who post in the discussion forums exhibit confusion patterns that are consistent with those of students that do not actively post. For example, the top ranked pattern *quiz* $\rightarrow$ *pageview* of students with posts (SwP) group ranks third among students with no posts (SwNP). The 2nd pattern is the same for both SwP and SwNP. Other highly ranked patterns in SwP are also found in similar ranking positions of SwNP. This demonstrates that it is reasonable to apply the trained classifier on SwP to SwNP; the two groups of students express their confusion through their clicking patterns in similar ways.

Pattern Ranking	Students with Posts	Students without Posts
1	quiz → pageview	pageview → lecture
2	pageview → quiz	pageview → quiz
3	lecture → pageview	quiz → pageview
4	pageview → lecture → pageview	lecture → pageview
5	pageview → quiz → pageview	pageview → quiz → pageview
6	lecture → quiz	pageview → lecture → pageview
7	quiz → pageview → lecture	quiz → pageview → quiz
8	quiz → pageview → quiz	lecture → quiz
9	lecture → pageview → quiz	quiz → pageview → lecture
10	lecture → pageview → lecture	lecture → pageview → quiz

Table 5: Most frequent click patterns associated with confusion among students who have posted and students who have not posted.

Confusion Over Course Weeks

After using the above trained session confusion classifier to predict the unlabeled sessions, we can monitor and track the confusion each student has expressed implicitly or explicitly in his/her learning process. To validate whether this predicted session confusion provides an understanding of students' activities, we visualized how students' confusion changes over time in two courses



(a) Algebra Course (b) Microeconomics Course

Figure 6: **Students' Confusion Changes over Course Weeks in Two Courses**. A week's confusion score is defined as the average confusion scores of all sessions in that week. Since two courses have different level of confusion as shown in Figure 2, we use different scales here. Each data point represents the average confusion of all students in a given course week, and the standard deviation across those measurements ranges between 0.04 and 0.10.

in Figure 6, based on the available confusion scores on each session for all students and also for students who have completed the course. We found that, at the very beginning of each course, students show a high level of confusion; as time goes on, such confusion decreases. This is consistent across all students and also students who completed the course. One possible explanation is that as students stay longer and get more involved in the learning process, they might have a better comprehension of the course materials. Around the 12th week in the Algebra course, preparing for the course final causes correlations with an increase in student confusion. The two peaks in the Microeconomics are associated with its Midterm and Final respectively. This is consistent with our expectation that confusion is often provoked by working on problems like quizzes, and people are dropping out not because of their personal confusion but because of specific units of the course being confusing in general.

Is Confusion An Individual Process?

In this part, we want to explore whether students' confusion in courses is related to course-wide issues, or is individual in nature. Some course-wide issues might include poor course materials, unsatisfying instructor styles or difficult quiz problems. Here, we represent a student's confusion towards a video that he/she struggles with in a session as that session's predicted confusion score. Then we compute each student's confusion towards each lecture. In this way, we measure students' average confusion towards each lecture video and have a ranking of which lectures are associated with the most confusion. The top 10 most confused lectures among all students are denoted as *Confused Lecture Set*. For each student, we have a list of which lectures in the Confused Lecture Set he/she has watched, and what is his/her ranking of such lectures based on his/her confusion scores. Therefore, each student is associated with his/her own ranking list of lectures in the Confused Lecture Set. We then compute the Spearman's rank correlation coefficient[5] between any two students' ranking lists.

[5] https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

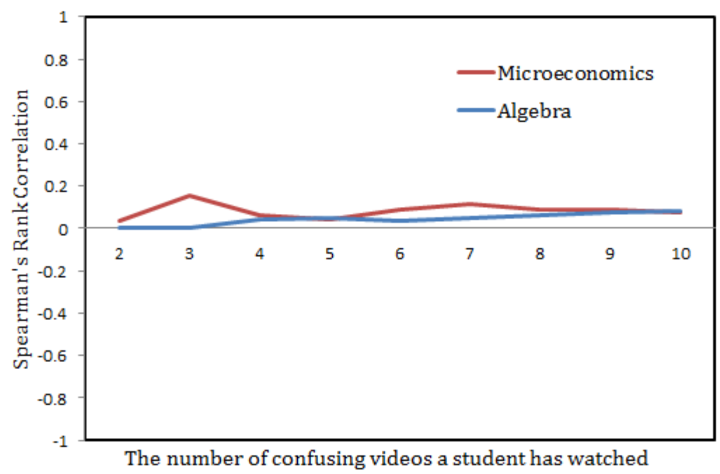


Figure 7: **Spearman's Rank Correlation on Students' Watched Videos in Two MOOCs**. It describes the average rank correlation between any two students who have watched at least certain number of confusing videos.

We visualized this coefficient across two courses in Figure 7, to present the average rank correlation between students who have watched at least certain number of videos in Confused Lecture Set. A rank coefficient of 1 implies the two rankings are the same; -1 represents that one ranking is the reverse of another; and 0 means the two rankings are completely different. As we can see from Figure 7, the coefficient in Algebra is in the range of 0 to 0.2, which indicates that a student's confusion is almost independent of another student's confusion. This finding is consistent in the Microeconomics course. Therefore, we conclude that confusion is a type of behavior associated with individuals.

6. Effects of Confusion on Commitment

Our hypothesis in this section is that different kinds of student confusion may lead to different effects on students' survival in MOOC courses. To test this hypothesis, we first investigated how different types of confusion experienced during students' learning processes influences students' continued participation and how such effects might be at least partly mitigated. Next, we explore the source of students' confusion and examine the influences of confusion expressed within different course contexts. We use survival analysis to explore these quantitatively, controlling for other forum behaviors and clicking behaviors. For example, the number of posts students contributes within a time period is a strong indicator of a prior level of commitment and thus makes an effective control variable. The constructed models are used to quantify the impact of confusion.

6.1. Method

6.1.1. Survival Analysis and Hazard Ratio

Survival analysis comprises a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest, which in our case is a student to dropping out of the course after experiencing confusion. Compared to simple linear regression models, survival analysis reduces biases from censoring, when the event of interest didn't occur during the observation period.

Mathematically, a survival function is defined as

$$S(t) = P(T \geq t)$$

where t is time, and T is a random variable indicating the time of dropping out. Survival model could be regarded as a type of regression model, which captures the changes of probability of survival over time. The hazard rate, is the instantaneous rate at which events occur, given no previous events:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}$$

The ratio of hazard rates, i.e. Hazard Ratio, corresponds to the conditions described by two levels of an explanatory variable. For example, for the explanatory variable *total number of posts*, the hazard ratio compares students who contribute one standard deviation more posts and students who have average level of posts.

6.1.2. Survival Analysis Design

Here we used the Stata statistical software package (Stata, 2011) to conduct parametric regression survival analysis with time-varying independent variables and assumed a Weibull distribution of survival times. Survival model design in previous studies associates one data point with each user and predicts how long a student survives based on persistent characteristics or variables that summarize his/her whole participation.

Different from this setting, we used a hierarchical mixed model where a data point is a time period, and we have multiple observations per student, in order to measure the impact of time varying features of the student's participation on the timing of dropout. That is, our unit of analysis is the student-week. We conducted two separate analyses, focusing on the influence of forum confusion and session confusion respectively.

Dependent Variables

- **Forum dropout** for forum confusion: In investigating the influence of post confusion, we consider a student dropping out from participation in the course community if that student has no activities in the course forum (even if they may engage in other forms of engagement with the materials). This dependent variable is a binary indicator, with 1 on a student's last week of active participation unless it is the last course week, and 0 otherwise.

- **Course dropout** for session confusion: When it comes to session confusion context, we consider a student dropping out from participation in the course community if that student has no activities in this course. This binary indicator is 1 on a student's last week of active participation unless it is the last course week, and 0 otherwise.

Control Variables

- **Total Posts**: This is the number of posts a student contributed to the forums in the current week. It could be regarded as a basic measurement of student engagement (Wen et al., 2014a).

- **Thread Starters**: This is the number of discussion threads a student initiated in the current week. Because students who start a thread are often asking help from others, they may have a higher probability of being confused compared to those who participate in discussions initiated by others.
- **Cohort**: This binary indicator measures whether a student has activities in the first week, which can be a proxy of student's motivation towards that course.
- **Posted**: This variable indicates whether a student has ever made an explicit post in the discussion forum.
- **Viewed**: This variable describes whether a student without posting behaviors has ever read a forum thread.

Independent Variables

- **User's Forum Confusion** (ExprConfusion):
  This measures the average confusion per post a student has expressed in a week. It was calculated by averaging the forum confusion scores of an individual's posts in that week.
- **User's Exposure to Forum Confusion** (UserExpoConfusion):
  This measures the forum confusion a student was exposed to during a week, and was calculated by averaging the forum confusion scores of posts in the threads that a student initiated during a week.
- **User's Exposure to Others Forum Confusion** (OthersExpoConfusion):
  This measures the average confusion a student was exposed to by averaging the measured confusion of posts in all the threads he/she participated in those he/she initiated.
- **Confusion Resolved** (Resolved):
  This variable indicates how many threads were initiated by a student and later resolved. Each thread has a flag, indicating whether the question is resolved or not. This is provided in the datasets. Students sometimes express confusion through initiating threads with questions. Others providing satisfactory help to such threads might relieve the confusion of those students.
- **Received Replies** (Reply):
  This variable indicates how many threads a student initiated that have received a response from others. Student communication in the discussion forums is a vital component in MOOCs where personalized interaction is limited.
- **User's Session Confusion** (Session Confusion):
  This measures the average confusion a student has expressed via his/her clickstream record in a week. It was calculated by averaging confusion scores of an individual's complete sessions in that week.
- **Session Read Confusion** (Read Confusion):
  This variable indicates whether a student has encountered confusion based on reading forum posts from other students. Information about which threads a student has read is

provided in the clickstream data. This is calculated by averaging the confusion of all the threads that a user has viewed in a week.

Courses	Algebra				Microeconomics			
	Mean	Std. Dev	Min	Max	Mean	Std. Dev	Min	Max
TotalPost	8.41	17.18	1	232	6.01	9.56	1	172
Starter	0.26	0.57	0	5	0.16	0.46	0	5
ExprConfusion	0.29	0.34	0	1	0.23	0.28	0	1
UserExpoConfusion	0.10	0.20	0	1	0.05	0.14	0	1
OtherExpoConfusion	0.21	0.23	0	1	0.15	0.17	0	1
Resolved	0.36	0.46	0	1	0.25	0.42	0	1
Reply	0.40	0.47	0	1	0.26	0.42	0	1
Cohort	0.69	0.46	0	1	0.73	0.44	0	1
Viewing	0.37	0.48	0	1	0.47	0.50	0	1
Posting	0.09	0.29	0	1	0.08	0.28	0	1
Session Confusion	0.34	0.42	0	1	0.38	0.43	0	1
Read Confusion	0.37	0.46	0	1	0.16	0.35	0	1

Table 6: Descriptive Statistics for the Variables Entered into Survival Regression Analysis.

All variables have been standarized, with a mean of zero and standard deviation of one. Of these variable, only students who posted in the course forums had values for expressed confusion, user exposed confusion, others exposed confusion, confusion resolved and reply, while all students have values for session confusion and read confusion. Table 6 reports the descriptive statistics for all the variables entered into the survival regression models.

	Algebra Model 1		Algebra Model 2		Microeconomics Model 1		Microeconomics Model 2	
Variables	HR	S.E	HR	S.E	HR	S.E	HR	S.E
TotalPost	0.60***	0.02	0.47***	0.02	0.71***	0.03	0.60***	0.03
Starter	1.29***	0.02	1.38***	0.05	1.19***	0.02	1.45***	0.11
ExprConfusion			1.17***	0.03			1.18***	0.05
UserExpoConfusion			1.06	0.04			1.05	0.04
OthersExpoConfusion			1.85***	0.06			1.51***	0.06

Table 7: Results of Confusion on Students' Survival (*: $p < 0.05$, **: $p < 0.01$, *** $p < 0.001$). N = 1,523 for Algebra Course, and N = 1,055 for Microeconomics Course.

6.2. Forum Confusion and Survival in the Course Forums

This section reports results from survival analyses testing the influence of forum confusion and its interaction with other variable on the length of time students participate in the forums. Data come from all students who contributed at least one post to the forums in either course. The unit for time is a week, with the timestamp of students' first posts being the starting point their active participation in the course discussion forums, and the date of their last post as the end of their participation. Data are considered right censored if their last post occurred during the last week of a course.

Table 7 shows survival results for both the Algebra and Microeconomics courses. Since the results are consistent across the two courses, we discuss only the results for the Algebra course. Model 1 in Table 7 reports the effects of the control variables of TotalPost and Starter (i.e., number of threads started by the student per week) on student dropout in the Algebra course. is The hazard ratio (HZ) for TotalPost of 0.60, indicating that students who contribute one standard deviation more posts than average are 40% more likely to survive compared to students with an average number of posts $(100 - (100\% \times .60)\%)$. HZ 1.29 for Starter indicates that students who have started one standard deviation more thread starters than average are $29\%((100\% \times 1.29) - 100\%)$ more likely to dropout. Model 2 shows that students who expressed confusion in their posts or were exposed to the confusion expressed by others dropped out more quickly, with exposure to others confusion having a stronger influence (HZ=1.85) than their own expressions of confusion (HZ=1.17). The reason might be be that when students express their own confusion they are not only indicating that they are having some difficulty in the course, but also their desire to learn and overcome an impasse. Exposure to others' confusion might drive students away by communicating that some course materials are low quality. For example, a student in the Algebra course is confused about a quiz and the following discussion shows a response to it by complaining about the same problem:

> "A: *I dont understand how to write square root for my answer. i saw it but it's been rejected.*
>
> B: *I don't understand how to write it either. -16-7sqrt(14). Is this correct? Because it's not working.*
>
> D: *I am still not getting it to go. Through i know that I'm correct*".

In the Microeconomics course, complaints about poor course materials were common, for instance,

> "A: *Hello! Just trying to watch the lecture now but it always stops in the middle for some reason. Does anyone have the same problem as mine?*
>
> B:*my biggest issue is also that the video freezes when I am trying to replay a section to take notes*
>
> C: *I think the quality of videos is very bad. This may lead to lost the interest in the course*".

The research literature we reviewed earlier shows that although confusion might harm learning, confusions that are resolved could enhance the learning experience and even encourage deeper engagement (VanLehn et al., 2003). To test whether confusions that were responded to and resolved increased students' commitment to the course, we added to the survival model interactions of between Users' Expressed Confusion both with the number of replies students' received and an indicator that their problems were resolved. ExprConfusion $\times$ Resolved measures the interaction effect of expressed confusion and confusion resolution; ExprConfusion $\times$ Repl measures the interaction effect of expressed confusion and receiving replies.

	Model 1		Model 2		Model 3	
Variables	HZ	S.E	HZ	S.E	HZ	S.E
TotalPost	0.58^{***}	0.02	0.59^{***}	0.02	0.60^{***}	0.02
Starter	1.11^{***}	0.02	1.22^{***}	0.03	1.30^{***}	0.04
ExprConfusion	1.64^{***}	0.04	1.69^{***}	0.05	1.59^{***}	0.04
Resolved			0.83^{***}	0.03		
ExprConfusion $\times$Resolved			0.87^{*}	0.04		
Reply					0.79^{***}	0.03
ExprConfusion $\times$Reply					0.91^{***}	0.02

Table 8: Results of Survival Analysis for Interaction Effects on Algebra Course (*: $p < 0.05$, **: p <0.01, *** $p < 0.001$) N = 1,523 for Algebra Course, and N = 1,055 for Microeconomics Course.

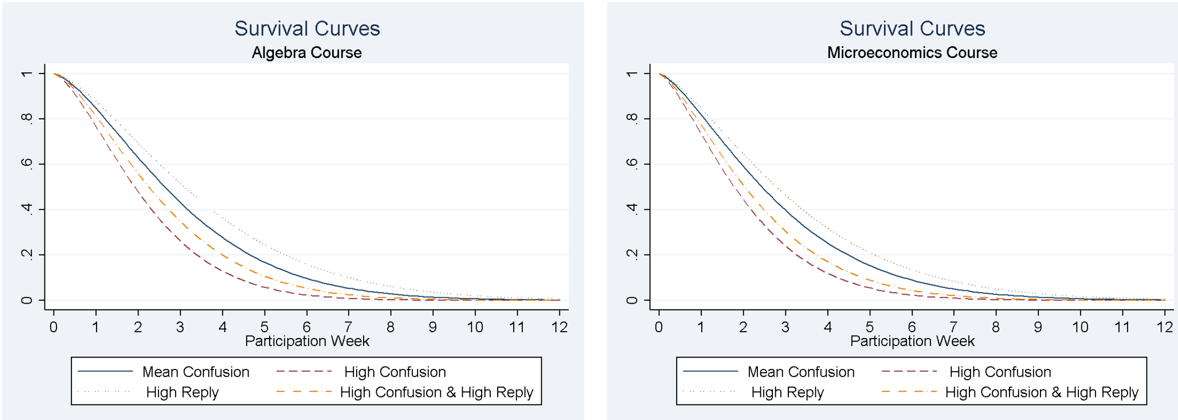
	Model 1		Model 2		Model 3	
Variables	HZ	S.E	HZ	S.E	HZ	S.E
TotalPost	0.66^{***}	0.03	0.67^{***}	0.03	0.68^{***}	0.03
Starter	1.18^{**}	0.04	1.21^{***}	0.04		
ExprConfusion	1.54^{***}	0.04	1.54^{***}	0.04	1.54^{***}	0.04
Resolved			0.87^{**}	0.04		
ExprConfusion$\times$Resolved			0.93^{**}	0.03		
Reply					0.83^{***}	0.04
ExprConfusion$\times$Reply					0.93^{**}	0.02

Table 9: Results of Survival Analysis for Interaction Effects on Microeconomics Course (*: $p < 0.05$, **: p <0.01, *** $p < 0.001$) N = 1,523 for Algebra Course, and N = 1,055 for Microeconomics Course.

The results are presented in Table 8 and Table 9. Taking the Algebra course as an example, the interaction analysis shows that expressing confusion predicted increased risk of dropping out substantially less when students' confusion not resolved than when it was not. The 22% increase in risk of dropping out if their confusion was resolved [6] is less than a third of the 69% increase in risk of dropping out if their confusion was not resolved not resolved. Model 3 in Table 8 shows similar results when comparing the effects of expressed confusion when students' confusion received a reply or did not. Results for the Microeconomics course in Table 9 are similar. Figure 9 and Figure 8 illustrate these results graphically. Taken together, these results suggest that being confused in courses has a negative influence on one's commitment, providing replies to such confusion and resolving the underlying issues at least partly mitigates that negative effect. Especially in a MOOC context, where instructors provide very limited guidance (Only 13% threads have instructor intervention in the Algebra course and 18% in Microeconomics), how to encourage students to offer help to others and even help unresolved threads get answered is a direction for potential positive impact.

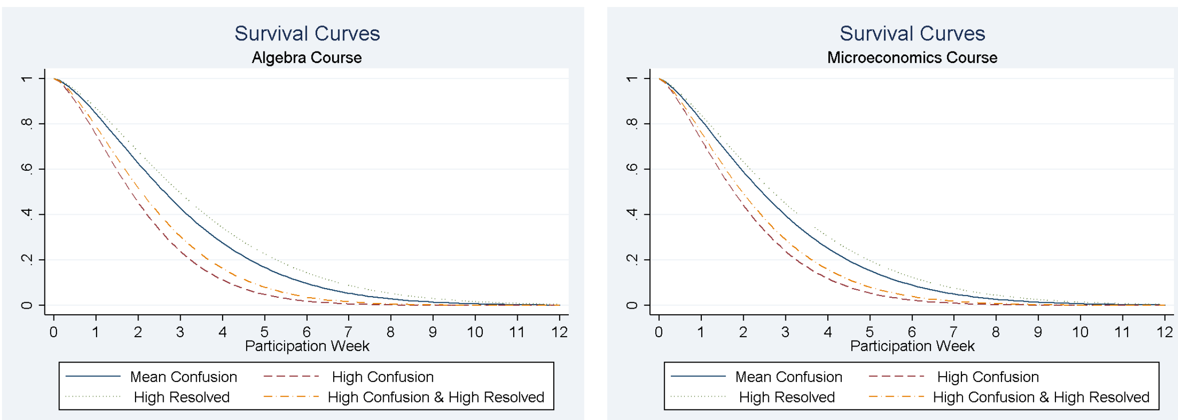
So far, we have have not differentiated types of confusion and only presented some cor-

[6] $1.22 = exp(log(1.69) + log(0.83) + log(0.87))$. For interpretation of interactions in nonlinear models: `http://www.stata-journal.com/sjpdf.html?articlenum=st0194`



(a) Algebra Course (b) Microeconomics Course

Figure 8: Survival Curves for Students Exposed to Different Levels of Confusion Being Replied



(a) Algebra Course (b) Microeconomics Course

Figure 9: Survival Curves for Students Exposed to Different Levels of Confusion Being Resolved

relational confusion analysis between confusion and student dropout. To take the next step and automatically recommend help, learning partners, and even experts in that specific field to confused students (Yang et al., 2014), we not only need to know who is confused, but also what students are confused about. For this purpose, first we designed a simple and effective method for identifying what students are talking about by utilizing their recent click behavior. That is, the content of current post is determined by the type of behavior indicated by the previous click before making that post. The intuition is that a student is more likely to express confusion about recently browsed course material or a recent course experience. The targeted types of contents in the posts and clicks can be divided into three categories: (1) *Course*: content about course tools, course settings, course resources, etc; (2) *Lecture*: content about lectures/videos or discussion on problems/content in lectures; (3) *Quiz*: content about quiz questions, such as quiz answers, submission, and deadlines. Post content that does not belong to any of the above types is categorized into *Other*. To validate the effectiveness of recent click type in determining post content, we randomly sampled 300 posts from each course, asked experts to judge what

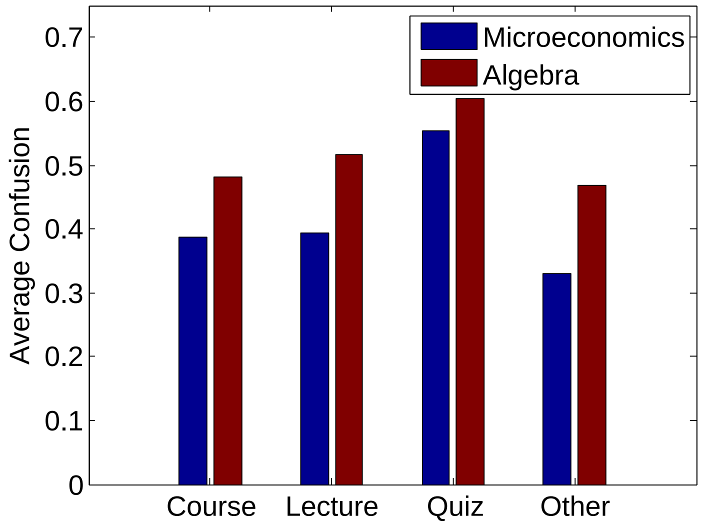


Figure 10: Average Confusion Comparison on Different Content Types.

students are talking about, and checked the correlation between human labeling and our method. Results demonstrated that the automated method that refers to recent click behavior can correctly identify 60% and 64% cases of what students are talking about on Algebra and Microeconomics courses respectively. Phi-Coefficient between human labeling and automatically identification is 0.405 with Kappa 0.300 on Algebra; on Microeconomics course, Phi-Coefficient is 0.632, with Kappa 0.308. We conclude that utilizing the previous click behavior before a post was made captures what students are talking about well enough to offer some visibility into the differential effects of confusion related to different course aspect, so we proceed with the analysis.

Once the context of each post has been identified, we then measure the confusion towards Course, Quiz, Lecture and Other by computing the average confusion scores in a user's Course, Quiz, Lecture, Other posts; denoted as *Course Confusion*, *Quiz Confusion*, *Lecture Confusion*, *Other Confusion* separately. We show the average confusion related to the four categories in Figure 10. We conducted the survival analysis again to observe how confusion within different course contexts influences dropout, as results shown in Table 10. It can be concluded that even though confusion distributions are similar across the two courses (quiz is associated with highest confusion), confusion towards different content types do not have similar influence on dropout. Instead, quiz confusion has the biggest impact on dropout in the Algebra course while confusion towards the course more generally leads to the highest dropout in the Microeconomics course. One explanation is that this pattern is a side effect of the fact that mathematical courses emphasize problem solving whereas less technical courses require much more discussion and opinion exchange. Thus we conclude the influence of different types of confusion on dropout is determined by the specifics of course settings, and we thereby cannot make strong claims generally about specific types of confusion being generally more or less damaging than others across courses.

6.3. Session Confusion and Survival in the Course

The previous section showed that confusion revealed in posting behavior predicted students' subsequent withdrawal from participation in the course forums. However, because only a small

	Algebra		Microeco	
Variables	HZ	S.E	HZ	S.E
TotalPost	0.51^{***}	0.02	0.63^{***}	0.03
Starter	1.06^{*}	0.02	1.08^{**}	0.03
Course Confusion	1.39^{***}	0.03	1.40^{***}	0.04
Quiz Confusion	1.51^{***}	0.03	1.00	0.03
Lecture Confusion	1.15^{***}	0.02	1.26^{***}	0.03
Other Confusion	1.01	0.03	1.03	0.03

Table 10: Confusion towards Different Course Aspects on Survival. N = 1,523 for Algebra Course, and N = 1,055 for Microeconomics Course.

	Algebra		Microeconomics	
Variables	HR	Std.Err	HR	Std.Err
Cohort	0.716^{***}	0.009	0.695^{***}	0.011
Posted	0.325^{***}	0.006	0.339^{***}	0.008
Viewed	0.499^{***}	0.007	0.582^{***}	0.009
Session Confusion	1.093^{***}	0.006	1.277^{***}	0.008
Read Confusion	1.292^{***}	0.007	1.117^{***}	0.006

Table 11: Results of Survival Analysis on All Course Participants(*: $p < 0.05$, **: $p < 0.01$, *** $p < 0.001$) N = 35,855 for Algebra Course, and N = 25,871 for Microeconomics Course.

fraction of students participated in the course forums, we tested the generalizability of the relationship between confusion and survival by examining whether session confusion, revealed in other click stream behavior and available for all students in the courses, also predict students' withdrawal from the course. In this section, we conduct a finer grained analysis on how different types of session confusion affect students' commitment to courses.

As in the previous section, we conducted a parametric regression survival analysis, assuming a Weibull distribution of survival times. This analysis included all the participants in the two courses. Again, survival time is measured in terms of student-weeks. However, unlike the previous analyses, we define the starting point as the timestamp that a student first performed any action in a course, and the end of participation as the last action a student made, with students still active the week of the course consided censored.

Results of the survival analyses are in Table 11. We investigated the influences of session confusion and read confusion, controlling for some important factors such as Cohort, Posting and Viewing. In the Microeconomics course, Cohort has a Hazard Ratio (HR) of 0.695, indicating that students who are one standard deviation higher are 30.5% more likely to survive the course. Students who contributed to one standard deviation more posts than average are 66.1% more likely to survive compared to students who have lower post counts. HR of 0.582 of Viewed indicates the survival rates are 41.8% much higher for those who have viewed one standard deviation more threads than average. This is consistent in the Algebra course.

Session confusion has a HR of 1.277 indicating that students who had one standard deviation higher confusion expressed implicitly in their clickstream data were 27.7% more likely to drop out. The HR of 1.117 for Read Confusion indicates that students who read one standard deviation more confusion than average were 11.7% more likely to drop out. Read confusion is an indirect indication of an individuals confusion via reading others' confusion, which reflects social influence.

7. Discussion and Conclusion

Increasing student engagement and facilitating learning is of great importance to future MOOC deployment. In this work, we investigated the effect of student confusion on survival in MOOC courses. We first measured the confusion demonstrated in students' discussion forum posts and click sessions. Next, we built a series of classifiers to automatically identify both explicit confusion indicated via students' language choices and their implicit and indicated by other click behaviors. We examined the effect of different kinds of confusion on students' continued participation in the MOOC courses. Our results demonstrate that: (1) the more students express their confusion and are exposed to confusion in the MOOC forums, the less likely students are to remain active in the learning community; (2) responding to students' confusing or resolving their problems reduces the impact of confusion on their dropout; (3) the extent to which different types of confusion affect dropout is determined by specific courses; in our data, quiz confusion leads to the highest dropout in a technical course while confusion towards the course in general contributes more to dropout in a less technical course; (4) the more confusion students demonstrate implicitly through their click behavior, or encounter by reading others' confusion, the more likely they are to drop out.

More specifically, compared to traditional classroom learning and many intelligent tutoring systems, students in MOOC courses are less likely to receive immediate feedback when confused. MOOCs have a large number of students in each course and do not often enable face-to-face interaction with instructors or other well-performing students. Such interaction could play an important role in maintaining student engagement because of its directive, facilitative and motivational functions (Shute, 2008). However, once students get confused, they must choose to either remain confused or attempt to continue course learning, or make their confusion explicit and wait for resolution. Remaining in a confused state without support from instructors and others might easily transition into dropout, while expressing confusion in the forum may not receive a timely response, which might eventually cause disengagement. The effect of being confused and then being given support (e.g. resolving their problems or providing students interaction) has a demonstrated effect to partly mitigate the effect of confusion on dropout. Therefore, in the context of MOOCs, encouraging students to help one another in resolving confusion is beneficial for students, instructors and creators of MOOC courses.

7.1. Limitations and Future Work

This study is subject to the limitation that all measures of student expression, including expressed confusion, exposed confusion, session confusion, read confusion, as well as interaction effect of confusion with reply or resolution, are judged by Turkers or computed indirectly via Turkers' ratings who are not MOOC course participants and possibly have no experience in MOOC course forums. There might also be some cascading errors associated with session confusion classifier, since session confusion labels are acquired via estimated forum confusion scores. A natural follow-up is to conduct surveys to collect self-reported confusion when students feel confused in a message level or session level to see whether such data aligns with our findings.

Secondly, our analysis of confusion gives limited support for understanding *what* students are confused about. It might be more helpful if MOOC instructors could not only grasp which students are confused, but also what they are confused about in a finer granularity such as confusion about detailed course concepts. After figuring out the cause of their confusion, we might have a clearer categorization of confusion, such as shallow confusion (e.g. difficulties in submitting homework) or deeper confusion (e.g. incomplete understanding about some theories). Knowing what students are confused about, paves the way for automatically recommending help, solutions, learning partners, and even experts in a relevant field to confused students. This could fully take advantage of our finding that the interaction of being confused and being provided help improves student learning. In future studies, we plan to further explore this area of inquiry. Moreover, we mainly focused on investigating the influences of confusion on students' commitment to courses, and did not explore the relations between confusion and student learning.

Lastly, we used data only from two courses to validate the performance of our confusion detection model and hypothesis about confusion's impact on student survival. More courses with rich information especially information about why a student chooses to attend a MOOC and which group of course participants target at finishing the course learning instead of for fun or meeting new people would bolster the soundness and generalizability of this work. Future work could also extend our studies by applying and evaluating how well these confusion classifiers trained over the whole population could transfer to different populations in terms of age, gender, education or countries. In addition, the confusion detection model is corpus-based, while in real world MOOC scenarios, real time models such as models trained on previous offering of similar courses could be developed to more comprehensively monitor students' confusion across time.

7.2. Implications

Our findings provide guidance for future MOOC deployment. First of all, this study sheds light on how to identify student confusion during the learning process. Tracking and monitoring student confusion helps instructors give appropriate feedback to students. Our machine learning classifier could give an accurate estimation of students' confusion expressed during their learning processes. We validate its performance in two different courses, one economics course and one mathematical course, where students are have different stylistic expressions of confusion. This might also give rise to a real time model to monitor students' confusion across time.

Our LIWC features, question features, and click pattern features can be directly applied to confusion detection in other situations since they are not tailored to any specific domain. Besides, our analysis targets at a lagrer population, regardless of whether users have explicitly made a post or not, compared to existing work, which gives a relatively thorough understanding of confusion that MOOC learners might encounter. In addition, we can recommend videos, slides or online resources to confused learners via a matching between their post questions and the content of these resources.

Thirdly, our results show that getting students' confusion resolved helps reduce their dropout. This provides practical guidance for improving student engagement, especially tailored to the distant nature and the sheer size of MOOCs. More interventions that incorporate resolving problems and providing instructor support could be developed and deployed. For the MOOC practitioner, rather than prioritizing avoiding student confusion, providing immediately support for confused students to create a more interactive environment should be prioritized. Resolving confusion promptly is beneficial for students' MOOC participation experience. To summarize, our findings provide insights into the impact of confusion and its interaction with getting support and practical implications for instructors and creators of MOOC courses.

This work highlights the need to design MOOC environments that are conducive to exchange of help, which includes taking care to encourage and not discourage help seeking, and to facilitate matching help seekers with help providers (Howley, 2015).

REFERENCES

AGRAWAL, A., VENKATRAMAN, J., LEONARD, S., AND PAEPCKE, A. 2015. Youedu: Addressing confusion in mooc discussion forums by recommending instructional video clips. In *National Science Foundation*. Stanford InfoLab.

ALARIO-HOYOS, C., PÉREZ-SANAGUSTÍN, M., DELGADO-KLOOS, C., MUÑOZ-ORGANERO, M., RODRÍGUEZ-DE-LAS HERAS, A., ET AL. 2013. Analysing the impact of built-in and external social tools in a mooc on educational technologies. In *Scaling up learning for sustained impact*. Springer, 5–18.

ANDERSON, A., HUTTENLOCHER, D., KLEINBERG, J., AND LESKOVEC, J. 2014. Engaging with massive online courses. In *Proceedings of the 23rd International Conference on World Wide Web*. WWW '14. ACM, New York, NY, USA, 687–698.

ARROYO, I., COOPER, D. G., BURLESON, W., WOOLF, B. P., MULDNER, K., AND CHRISTOPHERSON, R. 2009. Emotion sensors go to school. In *AIED*. Vol. 200. 17–24.

BAKER, R. S., GOWDA, S. M., WIXON, M., KALKA, J., WAGNER, A. Z., SALVI, A., ALEVEN, V., KUSBIT, G. W., OCUMPAUGH, J., AND ROSSI, L. 2012. Towards sensor-free affect detection in cognitive tutor algebra. *International Educational Data Mining Society*.

BOSCH, N., CHEN, Y., AND DMELLO, S. 2014. Its written on your face: detecting affective states from facial expressions while learning computer programming. In *Intelligent Tutoring Systems*. Springer, 39–44.

BOYER, S. AND VEERAMACHANENI, K. 2015. Transfer learning for predictive models in massive open online courses. In *International Conference on Artificial Intelligence in Education*. Springer, 54–63.

BUHRMESTER, M., KWANG, T., AND GOSLING, S. D. 2011. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science 6,* 1, 3–5.

CALVO, R. A. AND D'MELLO, S. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *Affective Computing, IEEE Transactions on 1,* 1, 18–37.

CAPRARA, G. V., FIDA, R., VECCHIONE, M., DEL BOVE, G., VECCHIO, G. M., BARBARANELLI, C., AND BANDURA, A. 2008. Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology 100,* 3, 525.

CHATURVEDI, S., GOLDWASSER, D., AND DAUMÉ III, H. 2014. Predicting instructor's intervention in mooc forums. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Baltimore, Maryland, 1501–1511.

CONATI, C. AND MACLAREN, H. 2009. Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction 19,* 3, 267–303.

CONOLE, G. 2013. Moocs as disruptive technologies: strategies for enhancing the learner experience and quality of moocs. *Revista de Educación a Distancia 39*, 1–17.

D'MELLO, S. AND GRAESSER, A. 2007. Mind and body: Dialogue and posture for affect detection in learning environments. *Frontiers in Artificial Intelligence and Applications 158*, 161.

D'MELLO, S., OLNEY, A., AND PERSON, N. 2010. Mining collaborative patterns in tutorial dialogues. *JEDM-Journal of Educational Data Mining 2,* 1, 2–37.

DMELLO, S. AND GRAESSER, A. 2012. Dynamics of affective states during complex learning. *Learning and Instruction 22,* 2, 145–157.

DMELLO, S., LEHMAN, B., PEKRUN, R., AND GRAESSER, A. 2014. Confusion can be beneficial for learning. *Learning and Instruction 29*, 153–170.

DMELLO, S. K., CRAIG, S. D., WITHERSPOON, A., MCDANIEL, B., AND GRAESSER, A. 2008. Automatic detection of learners affect from conversational cues. *User modeling and user-adapted interaction 18,* 1-2, 45–80.

EZEN-CAN, A. AND BOYER, K. E. 2015. Understanding student language: An unsupervised dialogue act classification approach. *JEDM-Journal of Educational Data Mining 7,* 1, 51–78.

GUPTA, N. K. AND ROSÉ, C. P. 2010. Understanding instructional support needs of emerging internet users for web-based information seeking. *JEDM-Journal of Educational Data Mining 2,* 1, 38–82.

HAN, J., CHENG, H., XIN, D., AND YAN, X. 2007. Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery 15,* 1, 55–86.

HOWLEY, I. 2015. Leveraging educational technology to overcome social obstacles to help seeking. Ph.D. thesis, Carnegie Mellon University.

HOWLEY, I., TOMAR, G., YANG, D., FERSCHKE, O., AND ROSÉ, C. P. 2015. Alleviating the negative effect of up and downvoting on help seeking in mooc discussion forums. In *AIED*.

HUSSAIN, M. S., MONKARESI, H., AND CALVO, R. A. 2012. Categorical vs. dimensional representations in multimodal affect detection during learning. In *Proceedings of the 11th International Conference on Intelligent Tutoring Systems*. ITS'12. Springer-Verlag, Berlin, Heidelberg, 78–83.

JORDAN, K. 2014. Initial trends in enrolment and completion of massive open online courses. *The International Review Of Research In Open And Distributed Learning 15,* 1.

KIM, J., GUO, P. J., SEATON, D. T., MITROS, P., GAJOS, K. Z., AND MILLER, R. C. 2014. Understanding in-video dropouts and interaction peaks inonline lecture videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*. L@S '14. ACM, New York, NY, USA, 31–40.

KIZILCEC, R. F. AND SCHNEIDER, E. 2015. Motivation as a lens to understand online learners: Toward data-driven design with the olei scale. *ACM Trans. Comput.-Hum. Interact. 22,* 2 (Mar.), 6:1–6:24.

KOEDINGER, K. R., KIM, J., JIA, J. Z., MCLAUGHLIN, E. A., AND BIER, N. L. 2015. Learning is not a spectator sport: Doing is better than watching for learning from a mooc. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*. L@S '15. ACM, New York, NY, USA, 111–120.

KOLLER, D., NG, A., DO, C., AND CHEN, Z. 2013. Retention and intention in massive open online courses: In depth. *Educause Review 48,* 3, 62–63.

LARSON, R. W. AND RICHARDS, M. H. 1991. Boredom in the middle school years: Blaming schools versus blaming students. *American Journal of Education*, 418–443.

LEE, D. M. C., RODRIGO, M. M. T., D BAKER, R. S., SUGAY, J. O., AND CORONEL, A. 2011. Exploring the relationship between novice programmer confusion and achievement. In *Affective Computing and Intelligent Interaction*. Springer, 175–184.

LEHMAN, B., DMELLO, S., AND GRAESSER, A. 2012. Interventions to regulate confusion during learning. In *Intelligent Tutoring Systems*. Springer, 576–578.

LEHMAN, B. AND GRAESSER, A. 2015. To resolve or not to resolve? that is the big question about confusion. In *Artificial Intelligence in Education*. Springer, 216–225.

LEHMAN, B., MATTHEWS, M., DMELLO, S., AND PERSON, N. 2008. What are you feeling? investigating student affective states during expert human tutoring sessions. In *Intelligent Tutoring Systems*. Springer, 50–59.

LIU, C., RANI, P., AND SARKAR, N. 2005. An empirical study of machine learning techniques for affect recognition in human-robot interaction. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*. IEEE, 2662–2667.

NICOLAOU, M. A., GUNES, H., AND PANTIC, M. 2011. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *Affective Computing, IEEE Transactions on 2,* 2, 92–105.

OCUMPAUGH, J., BAKER, R., GOWDA, S., HEFFERNAN, N., AND HEFFERNAN, C. 2014. Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology 45,* 3, 487–501.

PARDOS, Z. A., BAKER, R. S., SAN PEDRO, M. O., GOWDA, S. M., AND GOWDA, S. M. 2013. Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. ACM, 117–124.

PEKRUN, R., GOETZ, T., TITZ, W., AND PERRY, R. P. 2002. Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist 37,* 2, 91–105.

PENNEBAKER, J. W., FRANCIS, M. E., AND BOOTH, R. J. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates 71*, 2001.

RAMESH, A., GOLDWASSER, D., HUANG, B., DAUMÉ III, H., AND GETOOR, L. 2013. Modeling learner engagement in moocs using probabilistic soft logic. In *NIPS Workshop on Data Driven Education*.

ROSÉ, C. P., CARLSON, R., YANG, D., WEN, M., RESNICK, L., GOLDMAN, P., AND SHERER, J. 2014. Social factors that contribute to attrition in moocs. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 197–198.

SAARELA, M. AND KÄRKKÄINEN, T. 2015. Analysing student performance using sparse data of core bachelor courses. *JEDM-Journal of Educational Data Mining 7,* 1, 3–32.

SHUTE, V. J. 2008. Focus on formative feedback. *Review of educational research 78,* 1, 153–189.

SINHA, T., JERMANN, P., LI, N., AND DILLENBOURG, P. 2014. Your click decides your fate: Leveraging clickstream patterns in MOOC videos to infer students' information processing and attrition behavior. *CoRR abs/1407.7131*.

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 *EMNLP*. Association for Computational Linguistics, 254–263.

STATA, C. 2011. Stata statistical software release. In *7.0: Programming. Stata Corporation*. 418–443.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., and Baggett, W. B. 2003. Why do only some events cause learning during human tutoring? *Cognition and Instruction 21,* 3, 209–249.

Wang, Y.-C., Kraut, R., and Levine, J. M. 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*. CSCW '12. ACM, New York, NY, USA, 833–842.

Wen, M., Yang, D., and Rosé, C. P. 2014a. Linguistic reflections of student engagement in massive open online courses. In *Proceedings of the International Conference on Weblogs and Social Media*.

Wen, M., Yang, D., and Rosé, C. P. 2014b. Sentiment analysis in mooc discussion forums: What does it tell us? In *Proceedings of Educational Data Mining*.

Werner, L., McDowell, C., and Denner, J. 2013. A first step in learning analytics: pre-processing low-level alice logging data of middle school students. *JEDM-Journal of Educational Data Mining 5,* 2, 11–37.

Wilson, N. 1989. Learning from confusion: Questions and change in reading logs. *English Journal*, 62–69.

Wixon, M., Arroyo, I., Muldner, K., Burleson, W., Rai, D., and Woolf, B. 2014. The opportunities and limitations of scaling up sensor-free affect detection. In *Educational Data Mining 2014*.

Yang, D., Adamson, D., and Rosé, C. P. 2014. Question recommendation with constraints for massive open online courses. In *Proceedings of the 8th ACM Conference on Recommender Systems*. RecSys '14. ACM, New York, NY, USA, 49–56.

Yang, D., Sinha, T., Adamson, D., and Rose, C. P. 2013. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-Driven Education Workshop*. Vol. 10. 13.

Yang, D., Wen, M., Howley, I., Kraut, R., and Rose, C. 2015. Exploring the effect of confusion in discussion forums of massive open online courses. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*. L@S '15. ACM, New York, NY, USA, 121–130.

Yang, D., Wen, M., and Rose, C. 2014. Peer influence on attrition in massive open online courses. *Proceedings of Educational Data Mining*.