Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains

Xu Wang, Diyi Yang, Miaomiao Wen, Kenneth Koedinger, Carolyn P. Rosé
School of Computer Science, Carnegie Mellon University
5000,Forbes, Pittsburgh, PA, 15213
{xuwang, diyiy, mwen}@cs.cmu.edu, koedinger@cmu.edu, cprose@cs.cmu.edu

ABSTRACT
While MOOCs undoubtedly provide valuable learning resources for students, little research in the MOOC context has sought to evaluate students' learning gains in the environment. It has been long acknowledged that conversation is a significant way for students to construct knowledge and learn. However, rather than studying learning in MOOC discussion forums, the thrust of current research in that context has been to identify factors that predict dropout. Thus, cognitively relevant student behavior in the forums has not been evaluated for its impact on cognitive processes and learning. In this paper, we adopt a content analysis approach to analyze students' cognitively relevant behaviors in a MOOC discussion forum and further explore the relationship between the quantity and quality of that participation with their learning gains. As an integral part of our approach, we built a computational model to automate the analysis so that it is possible to extend the content analysis to all communication that occurred in the MOOC. We identified significant associations between discourse behavior and learning. Theoretical and practical implications are discussed.

Keywords
Massive Open Online Courses (MOOC); Cognitive behavior; Content analysis; Discussion forum; Learning gains;

1. INTRODUCTION
Despite concerns over their effectiveness, MOOCs (Massive Open Online Courses) have attracted increasing attention both in the popular press and academia, raising questions about ptheir potential to deliver educational resources at an unprecedented scale to new populations of learners. With learning through social processes featuring among the potential impacts of MOOC platforms [5], and discussion forums currently the primary means for supporting social learning in typical MOOC platforms, recent research has begun to focus on interventions that might enrich students' interaction in this context [e.g., 30], with the purpose of providing a more engaging and effective learning experience. Previous studies on learning and tutoring systems have provided evidence that students' participation in discussion [e.g., 2, 9, 12] is correlated with their learning gains in other instructional contexts.

However, whether discussion will also contribute substantially to learning in a MOOC context, and what aspects of discussion will ultimately matter most to learning in this new context remain important open questions. Considering the significant connection that has been discovered between discussion behaviors in MOOC forums and student commitment, its potential for enabling students to form supportive relationships with other students, and the potential to enhance social learning through interaction, in depth empirical research is needed to uncover the relationship between student discourse patterns and learning gains in MOOCs.

One challenge to assessing learning in MOOCs, even in cases where formal assessments are integrated with the courses, is that students come into a MOOC with a wide variety of backgrounds [15,20], and it is typically unnatural to make a pretest a natural part of the learning process, especially when activities in the MOOC are all voluntary. However, while inconvenient, it is not impossible. The study reported in this paper took place in an unusual MOOC where a pretest was provided and students were aware that the MOOC data would be used for research purposes. This dataset, from a course entitled "Introduction to Psychology as a Science", thus provides a unique opportunity to begin to address the research questions introduced above.

Many student behaviors have been observed in discussion forums, e.g., question answering, self-introduction, complaining about difficulties and corresponding exchange of social support. A very coarse grained distinction in posts could be on vs. off topic. However, the important distinctions do not stop there and may be substantially more nuanced than that. Other than literal topic features, students' cognitively relevant behaviors, which are associated with important cognitive processes that precede learning may also be found in discussion forums. What those behaviors are in this context, and how frequently they occur are two questions we address.

Specifically, we ask the following research questions in this work:

1. Is a higher quantity of participation in MOOC discussion forums associated with higher learning gains?
2. Is on-task discourse associated with more learning gains than off-task discourse?
3. If certain properties of discussion are associated with enhanced learning, why it is so? What are the higher-order thinking behaviors demonstrated in student discourse and their connection with learning?

We consider that answering these questions has important implications for designing discussion interventions in MOOCs.

Some previous studies on MOOC discussion forums analyzed at a macro-level the quantity of participation [e.g., 1], whereas other work [23] pointed out that quantitative indices of participation does not directly imply the quality of conversation and interaction. Others conducted content analysis of thread topics [17] or used rule-based algorithms to extract linguistic markers [28]. However, students' higher-order thinking behaviors are not well represented or thoroughly and systematically explored in these previous investigations. In this work, we aim to adopt a content analysis approach to hand-code data based on a well-established learning activity classification framework from earlier cognitive science research [8] in an attempt to capture students' discussion behaviors and their underlying cognitive strategies in a MOOC discussion forum. This is the first work we know of that has brought this lens to explore students' discussion behaviors and their association with learning gains in MOOCs.

In particular, we contribute to the existing literature by 1) developing a coding scheme based on Chi's ICAP (Interactive-Constructive-Active-Passive) framework [8] in categorizing students' discussion behaviors in a MOOC context; 2) providing empirical support for the importance of discussion in enhancing learning in a MOOC context. We also contribute to the literature on computer-supported collaborative learning by exploring the relationship between discourse and learning in a multi-user distributed asynchronous discussion environment.

In the remainder of the paper, we first discuss related work and existing theoretical foundations that we leverage in our analysis. Next we introduce our dataset. We then describe our methods, including specifics about the coding scheme, and computational model in the Methods section. We present an extensive correlational analysis and then discuss our interpretation along with caveats and directions for continued work.

2. RELATED WORK

2.1 Research on MOOC discussion forums

Studies in the field of learning science and computer supported collaborative learning have provided evidence that learners' contribution to discourse is an important predictor of their knowledge construction [2, 12]. In offline environments, studies have suggested, for example, that the number of words per utterance [26] and proportion of words produced [14] are correlated with learning gains. Transitioning from traditional classroom to online learning, computer-mediated conferencing has proved to be a gold mine of information concerning students' psycho-social dynamics and their knowledge acquisition [19]. Investigating the usage of discussion forums in MOOCs has been one major theme for research. To give a few examples, at a participation level, Anderson and colleagues [1] found that students who participated in other platform activities (videos, quizzes, etc.) participated more in the forum as well. They also explored patterns of thread initiators and contributors in terms of specific discussion behaviors in the discussion forum. At a content level, Brinton [5] categorized discussion threads into "small-talk", "course logistics", and "course specific" categories. Gillani [17] adopted a content analysis approach combined with machine learning models to discover sub-communities in a MOOC based on user profiles. Anderson [1] used a lexical analysis to see which words predict the number of assignments a student finally turns in.

These studies have set up a good foundation for analyses in MOOC discussion forums. However, to confirm a relationship between discussion and learning, we need to look closer into what aspects of discussion actually contribute to learning from a cognitive perspective.

2.2 Content analysis

We base our work on previous approaches to analyze content of student dialogues in tutoring and computer-supported collaborative learning environments. Chi [6] pointed out the importance of verbal analysis, which is a way to indirectly view student cognitive activity. De Wever [16] further demonstrated that content analysis has the potential to reveal deep insights about psychological processes that are not situated at the surface of internalized collaboration scripts.

Chi's ICAP framework [8] has been considered to be the strongest evidence for the value of a dialogic approach to learning [25], which has been widely adapted and applied to identify learning activities and explain study results [e.g., 24, 27]. The framework has been utilized to explain classical educational experiments [10] and serve as a theoretical foundation for studies on tutoring and computer-supported collaborative learning, for example in a discourse analysis of different kinds of scaffolds [24].

The framework was created through a meta-analysis of 18 studies in which learning activities were classified into 3 categories, namely, interactive activities that involve discussing and co-constructing with a peer or the learning environment, constructive activities that produce a representation of information that goes beyond the presented information, and finally, active activities that show how students are actively engaged in the learning process. The taxonomy suggests the hypothesis that what are referred to in it as interactive activities should generate more learning outcomes than constructive activities, which in turn should generate more learning outcomes than active activities. [8]

MOOCs provide an emerging environment where computer-supported collaborative learning activities might be provided, and where social presence might reflect cognitive presence [27]. Thus, in this context we aim to apply the ICAP framework to explore the relationship between discussion and learning by coding observed student behaviors in the discussion forum.

3. DATASET

In this work, we conducted a secondary analysis of the dataset of the course "Introduction to Psychology as a Science" offered through Coursera collaboratively by Georgia Institute of Technology and Carnegie Mellon University. The course incorporated elements of the OLI (Open Learning Initiative) "Introduction to Psychology" learning environment. One special characteristic of the course was that it administered a pre/post test with the intention to support research.

"Introduction to Psychology as a Science" was designed as a 12-week introductory course. For each week of class, the course targeted a major topic (e.g. Memory, Brain Structures, Nervous System); Course materials include video lectures, assigned MOOC activities, learning activities in the OLI environment, and what are referred to as weekly high-stakes quizzes.

In the first analysis of the dataset [21], researchers found that students who registered for the OLI activities learned more than students who used only the typical MOOC affordances, and further demonstrated that students who did more learning-by-doing activities learn more than students who watch more videos or read more texts. In other words, doing an activity has a much greater effect (6x) on predicted learning outcomes than watching a video or reading a web page. However, students' participation in the discussion forum hasn't been explored yet in that work.

In our preliminary exploration into the dataset, we found that when controlling for students' registration for OLI activities (which serves as a control variable associated with effort and commitment to the course), their quantity of participation in discussion forums significantly predicts learning gains as well. Based on this, we wanted to further explore how students' specific cognitively relevant behaviors in the forums correlate with their learning gains. We observed specific related discourse behaviors in the forum, and present several examples here.

Active behavior: *"According to the OLI textbook, creative intelligence is 'the ability to adapt to new situations and create new ideas or practicality'."*

This is an example of the student actively repeating what's being said in the course materials.

Constructive behavior: *"When I tell my son to wash the dishes, it's much more straightforward to explain his refusal or agreement by some behavioral (e.g. Reward or punishment) or cognitive mechanisms than by an innate instinct to wash or not to wash the dishes."*

This is an example of constructive behavior, when the learner produces output, which could be examples, explanations, etc., that go beyond course materials.

Interactive behavior: *"I agree that language can be an extra difficulty, but it is not a variable with which is counted. Also, depression, work stress…could form extra difficulties for the student in particular."*

This interactive behavior example shows that students not only engage in self-construction, but build their ideas upon their partners' contributions.

Altogether, there are 27,750 registered users in the dataset, and 7,990 posts and comments in the dataset. For the learners who have both pretest and posttest on record, which is our population of interest, there are 3,864 posts in total and 491 users. In addition to forum records, student clicks with course materials are also recorded in the clickstream data. The course has 1,487,665 student clicks. The clickstream logfile provides us with the opportunity to observe each students' interaction with course materials.

4. METHOD

4.1 Unit of analysis

In this paper, our unit of analysis is the message. As proposed in [16], in their review of 15 instruments in doing content analysis of the transcripts of online asynchronous discussion groups, 7 recommended using the message as the unit of analysis.

We first looked at students' quantity of participation, and distinguished on-task discourses from off-task. We then applied a coding scheme on on-task discourse to capture the cognitive behaviors in the discussion forum. We hand-coded half of the dataset, and trained a machine learning model to replicate that annotation approach in the rest of the dataset.

In a MOOC context, the data we usually deal with is student log data [4, 5, 13], which illustrates their participation process. However, students' cognitive behaviors are better represented in their discourse displayed in the discussion forum. In this work, we hand-coded a large sample of the dataset, which may reduce noise in this kind of analysis. Thus the result may be more reliable in demonstrating the relationship between students' cognitive behaviors in the discussion forum and their learning gains.

4.2 Quantity of participation

H1: In response to our first research question, we hypothesized that students who participated more in the discussion forum have higher learning gains.

We quantified students' participation in the discussion forum by the variable PostCountByUser.

PostCountByUser: It is measured by the number of posts a user posted in the discussion forum.

We did not distinguish between posts and comments in this analysis. So the word posts when mentioned in the rest of the paper refers both to posts and comments.

4.3 On-task vs. Off-task discourse

H2: in response to our second research question, we distinguished on-task and off-task discourse in the dataset. And we hypothesized that students' total number of on-task discourse contributions has a positive association with their learning gains.

We distinguished on-task discourse from off-task discourse in the dataset, based on the following definitions. On-task discourse includes posts that talk about course content, the content of quizzes and assignments, comments on course materials, and interaction between students on course content-related issues. Off-task discourse includes posts that talk about administrative issues in the course, e.g., asking for extensions on assignments; technical issues regarding course materials, e.g., asking where to download videos, off-topic self-introductions and social networking.

This feature in the dataset is acquired through hand-coding.

4.4 Cognitively Relevant Discussion behavior

H3: In response to our third research question, we want to investigate what discussion behaviors are demonstrated in the discussion forum, their frequencies and their association with learning. In order to capture these discussion behaviors, we developed a coding scheme based on Chi's ICAP framework [8].

We further hypothesized that students who demonstrated more higher-order thinking behaviors in each of the categories, active discourse, constructive discourse, and interactive discourse have higher learning gains. And according to Chi's work represented in 18 empirical studies, we hypothesized that the effect follows the pattern interactive>constructive>active.

4.4.1 Coding scheme

Students' cognitive behaviors are reflected in the MOOC discussion forums, which is not easily mined through rule-based algorithms due to its scale and informal style. This may pose challenges for computational modeling. In this work, we adopt a hand-coding method to capture higher-order thinking behaviors and follow the hand coding with computational modeling.

Within the category of on-task discourse we divide all posts into 3x3 categories as listed in Table 1 according to Chi's Active-Constructive-Interactive framework [8]. We further offer operational definitions for each category, and provide examples from our dataset. Due to space limitations, we provide abbreviated definitions rather than the full ones provided to the human coders. When defining each category of cognitive behavior, we evaluated how this might contribute to learning. Through empirical observation, we found this coding scheme to be exhaustive of all conditions. The 9 categories are not mutually exclusive. Thus, a post may belong to more than one of these fine-grained categories.

4.4.2 Inter-rater reliability

Two experts separately coded 100 posts randomly selected from the dataset, and applied on- vs. off-task annotation plus the 9 fine-grained categories of discussion behaviors to the sample. The two experts at first reached an agreement statistic of 0.52 (Cohen's Kappa), which is a moderate level of agreement. The two experts then resolved their disagreements through consensus coding by discussing and clarifying some borderline cases. After higher consensus was achieved, one of the experts coded 2000 posts randomly sampled from the whole dataset (3864 posts).

Table 1. Coding Examples

Active Discourse- (1) Repeat	
Operational Definition: The learner explicitly repeats information that's already covered in the material, which could be indicated by quotes.	*E.g. 1: Week 2, I quote from the picture: "The portion of the sensory and motor cortex devoted … as does the entire trunk of the body."*
Active Discourse- (2) Paraphrase	
Operational Definition: The learner paraphrases what's covered in course materials, it could be indicated by words like "it's mentioned in the textbook…", "it's said in the video…"	*E.g. 2: On the chapter about Health Psychology there is a board depicting various factors about Happiness, such as the Inequality of Happiness and then the Inequality Adjusted Happiness.*
Active Discourse- (3) Notes-taking	
Operational Definition: The learner mentions about note-taking and information seeking.	*E.g. 3: I use the text files as a basis for my lecture notes.*
Constructive Discourse- (1) Ask novel questions	
Operational Definition: The learner proposes a novel question or problem based on his/her own understanding.	*E.g. 4: Violence is throughout our history and have shaped societies, is it really as simple as an observed response? or a throwback of survival instinct?*
Constructive Discourse- (2) Justify or provide reasons	
Operational Definition: The learner uses examples and evidence to support a claim he/she has made. Reasoning is explicitly demonstrated in the discourse.	*E.g. 5: It depends on the visual field. Signals from the right visual field come to the left hemisphere, while signals from the left visual field come to the right hemisphere.*
Constructive Discourse- (3) Compare or connect	
Operational Definition: The learner compares cases, connects or shares links to external resources.	*E.g. 6: Here's a link to an article about a lady who stopped dreaming after suffering a stroke: [link]*
Interactive discourse- (1) Acknowledgement of partners' contribution	
Operational Definition: The learner explicitly acknowledges their partners' contribution, which could be indicated by "thanks for pointing that out", "I agree with you there…".	*E.g. 7: That's an interesting point, and it has made me wonder why this example was chosen.*
Interactive discourse- (2) Build on partners' contribution	
Operational Definition: The learner makes a point that builds on what their partner has said.	*E.g. 8: I do agree with what you said to a large degree. Changing a behavior merely to avoid pain or any other form of punishment is not good… Hence it requires a much deeper introspection and understanding…*
Interactive discourse- (3) Defend and challenge	
Operational Definition: The learner challenges his/her partners' ideas, or defends their own ideas, when there is a disagreement. (Note: The partner here can be either a peer or the learning environment)	*E.g. 9: I think I understand what you mean (I am currently doing the statistics course as well). However, as I can see from what you've described, you still have the hypothesis in your psychological experiment which is not null - your prediction that something WILL happen.*

4.4.3 Computational model and data preparation

In order to better visualize the dataset and potentially apply the model to another context, we trained a computational model based on the coded 2000 posts to predict the cognitively relevant discussion behavior categories and expand the coding to the rest of the dataset.

In our hand-coded dataset, we labeled 9 types of cognitively relevant discussion behaviors, but due to the fact that the occurrences of each single category are relatively sparse, we acquired a low accuracy when using the sample to train a model and apply it to the rest of the dataset. Instead, we aggregated the 9 categories into the three major categories—Active, Constructive, and Interactive. All three are binary variables indicating whether the user has a post under this category. We then built models to predict these labels.

Our classifier is designed to predict whether the cognitive behavior expressed in a post belongs to Active (A), Constructive (C) or Interactive (I) by taking advantage of a bag-of-words representation. However, we have to distinguish between on-task discourse and off-task discourse since learning relevant cognitive behaviors will occur primarily in on-task discussion (Among our coded 2000 posts, 558 are on-task discussions).

For this purpose, we built a two-stage classification model. In the first stage, we designed a logistic regression classifier to predict whether a post is on-task or off-task; in the second stage, we classified the posts that were predicted to be on-task into A, C or I categories. The input for each classifier is a bag-of-words feature representation. In the first step, we used the coded 2000 posts as the training set to train a logistic regression classifier to distinguish on-task discourses and off-task discourse, and in the second step, we used 588 on-task messages as our training set to train three logistic regression classifiers to label on-task discourses in the three categories (A, C, I). On the training set, we adopted a 10-fold cross-validation approach to evaluate the model. The classification results presented in Table 2 are the average accuracy and Kappa for this cross-validation. The results show that both accuracy and kappa are within a reasonable range for our further analysis of the whole dataset.

Table 3 shows some top-ranked features identified by the classifiers that are used to predict the three cognitive behaviors. From this table, we can see that in active discourse, students more often mentioned “lectures” “page” “notes”, which indicates they’re actively engaged with the course materials. In constructive discourse, students more often mentioned words associated with reasoning, such as “more” “but” “hence” “examples”, and in interactive discourses, students mentioned “your” “agree” “disagree” more often, which implies interaction. These features are consistent with our initial definitions of these distinct categories of discussion behavior and assumptions about their underlying cognitive processes and strategies.

Table 2. Evaluation metrics of the computational model.

Evaluation Metrics		Accuracy	Kappa
1st Stage **On-task Prediction**	On-task	82.1%	0.527
2nd Stage	Active	74.3%	0.361
Cognitive Behavior	Constructive	75.4%	0.318
	Interactive	75.6%	0.236

Table 3. Performance of Discussion Behaviors Regressors and Top Ranked Features

Categories	**Active**	**Constructive**	**Interactive**
Most Important Word Features (Regression Weight)	***lecture (1.68) page (1.24) what (.84) text (.83) incorrect (.79) answer (.72) says (.72) notes (.68)***	***course (.87) more (.79) give (.75) trying (.68) but (.64) hence (.64) looking (.61) topics (.58) example (.56) because (.56)***	***your (1.56) agree (1.11) our (.99) again (.86) thanks (.76) disagree (.6) response (.6)***

4.4.4 Clickstream Data

In order to explore the relationship between cognitively relevant discussion behaviors and learning, we also need to control for students’ involvement in other activities in the MOOC environment other than the discussion forum. This enables us to isolate, to some extent, the effect of pure effort and engagement in the course from the effects specifically related to discussion behavior. We further generated the following control variables through mining clickstream data of the course.

Video: The variable was computed first by summing the number of unique videos the student started to watch (Based on clicks on unique video urls), and then standardizing the sums.

Quiz: The variable was computed first by summing the number of unique quizzes the student attempted (Based on clicks on unique quiz urls), and then standardizing the sums.

OLI_textbook: The variable indicates reading the OLI textbook, and it’s calculated by summing the number of clicks the student made in the OLI environment and then standardizing the sums.

5. RESULTS

5.1 Participation quantity in the discussion forum

In response to the first research question, we fitted linear regression models to explore the relationship between students’ quantity of participation and their learning gains.

In the dataset, there are 1,079 students out of 27,750 students (i.e., students who registered for the course) who have both pre- and post-test scores on record. And among these students, there are 491 students who have posted in the discussion forum, with a total of 3,864 posts. We now introduce the variables we used in these models.

Dependent variable:

Post-test: The dependent variable in all the following models is students’ posttest score, which is standardized. Post-test score is students’ final exam score composed of 35 multiple-choice questions.

Control variable:

Pre-test: This is a test students took before the course started, which contains 20 multiple-choice questions. We also standardized the pretest score.

OLI_Registration: This is a binary variable capturing whether the student has registered for OLI or not. 1 means the student registered for OLI. As demonstrated in [21], students who registered for OLI learnt more than students who didn’t.

We also controlled for students’ involvement in other activities, including Video, Quiz and OLI_textbook.

Independent variable:

Participation: This is a binary variable indicating whether the student has ever posted in the discussion forum during the course.

PostCountByUser: This is the total number of posts a student contributed in the discussion forum during the course.

From Model 1, we see that whether the student has participated in the discussion forum is a significant predictor of the student’s learning gains. The result from Model 2 shows that for those who have participated in the discussion forum, the more they posted, the higher the learning gains they achieved.

Table 4. Regression results of learning gains on the quantity of participation and on-task discourse

Control/Indep. Variable	Model 1 (N=1079)	Model 2 (N=491)	Model 3 (N=491)
Participation	0.089**		
PostCountByUser		0.005*	0.006*
OnTaskPercent.			0.123**
Pretest	0.196***	0.254***	0.243***
OLI_registration	0.119**	0.107	0.120
Video	0.056*	0.0001	-0.011
Quiz	-0.008	-0.035	-0.037
OLI_textbook	0.050**	0.048	0.044

($p<0.001$***, $p<0.01$**, $p<0.05$*)

5.2 On-task versus off-tasks discourse

In response to the second research question, we looked at whether students' on-task discourse contributes to their learning gains. In this model, we examine the main effect of on-task discourse, which is represented by the variable OnTaskPercent.

Independent variable:

OnTaskPercent. : This is measured by the number of a student's posts that are categorized as on-task divided by the total number of posts the student has made, and the value is standardized.

In this regression model, we also controlled for a student's number of posts, whether they registered for OLI, pretest score, and their involvement in other activities. The regression result is displayed in Table 4-Model 3. The result shows that the quantity of students' on-task discourse in the discussion forum is a significant predictor of their learning gains.

5.3 Cognitively relevant discussion behavior analysis

5.3.1 Active, Constructive and Interactive behaviors

In this section we examine the relationship between students' discussion behavior and their learning gains and attempt to explain why certain behaviors lead to learning. We built linear regression models to explore the relationship between students' active, constructive and interactive discussion behaviors and their learning gains.

In the whole dataset, the number of instances (N=3864) of active, constructive and interactive activities is respectively 269, 744 and 203. And the number of students (N=491) who have demonstrated active, constructive and interactive activities is respectively 114, 230 and 84.

Our independent variables include:

Active, Constructive, Interactive: All three are binary variables indicating whether the student has a post that is categorized in that category.

We also controlled for variables including pretest, the number of posts, whether registered for OLI, students' involvement in other activities, as defined above. The regression result is shown in Table 5.

In Model 4 and Model 5, we found that students who have demonstrated active and constructive behaviors in the discussion forum had significantly more learning gains than students who didn't. From Model 6, we can see that the effect of Interactive discussion behavior is not significant in predicting learning gains. And we then introduced another variable to define whether a user is an active poster by counting the total number of their posts.

Poster profile: This nominal variable categorizing users into active poster and inactive poster. If a user has more than 3 posts (including 3), he/she is categorized as an active poster, otherwise categorized as an inactive poster. 3 is the median of the number of posts.

When nesting interactive behaviors with a poster profile, we found that interactive discussion is a significant predictor of learning gains for students who posted less. We think this might be because the number of posts is a basic measure of a student's social engagement in the discussion forum, which overlaps with some behaviors under the Interactive category. We further fitted a regression model to check the correlation between a student's total number of posts and the number of posts that are categorized as Interactive. The result shows that Interactive posts account for 66% of the variance in the total number of posts. We consider this high correlation could lead to the result described above. The results here show that both active and constructive discussion behaviors significantly contribute to students' learning gains, with active behaviors having higher predictive power. For users who posted less in the discussion forum, interactive behaviors strongly predict their learning gains (coefficient=0.515), however, the effect of interactive behavior disappears on the overall user population.

In addition to the occurrence of different discussion behaviors, we also used the frequency of each behavior as independent variables and did a second round of regression, from which we acquired similar results.

Table 5. Regression results of learning gains on discussion behaviors (part 1, N=491)

Control/Indep. Variable	Model 4	Model 5	Model 6	Model 7
Active	0.125*			
Constructive		0.112*		
Interactive			0.106	
Interactive [inact._poster]				0.496*
Interactive [act._poster]				0.043
Pretest	0.252***	0.246***	0.254***	0.254***
PostCntByUser	0.004	0.004	0.004	0.004
OLI_registr.	0.125	0.109	0.104	0.115
Video	-0.004	0.015	0.003	0.007
Quiz	-0.039	0.036	-0.038	-0.036
OLI_textbook	0.034	0.044	0.040	0.036

(p<0.001***, p<0.01**, p<0.05*)

5.3.2 Specific discussion behaviors

From the hand-coded dataset (N=2000), we summarized the occurrences of the 9 sub-categories of behaviors in Table 6. It shows that the most frequent behavior in the discussion forum is proposing an idea or asking novel questions. And the least frequent behaviors include building on a partner's contribution as well as defending and arguing, which is consistent with our expectation that higher-order thinking behaviors and highly interactive behaviors are relatively rare in MOOC discussion forums, and that the conversations going on in MOOCs are not satisfactorily rich and interactive.

Table 6. Distribution of 9 categories of discussion behaviors

Behavior Type	Freq.	Behavior Type	Freq.
Repeat	53	Notes-taking	28
Paraphrase/ask shallow questions	103	Justify or provide reasons	118
Propose an idea/ask novel questions	315	Compare, connect/ Reflect	59
Acknowledge partners' contribution	101	Build on partners' contribution	23
Defend and argue	14		

We also fitted regression models on this more nuanced coded dataset, but due to the fact that the occurrences of each category is relatively sparse, there was not sufficient statistical power to detect a significant effect of every category on learning gains. We display just the 2 significant predictors (out of 9) in Table 7.

Independent variables:

Constructive-(1): This is a binary variable indicating whether the student has a post that is categorized as "propose an idea/ ask novel questions".

Constructive-(2): This is a binary variable indicating whether the student has a post that is categorized as "Justify or provide reasons".

We controlled for pretest, number of posts, and whether the student registered for OLI in the regression models. We also controlled for students' involvement in other activities, the effects of which aren't significant in the regression models, so we don't report them here in Table 7.

Table 7. Regression results of learning gains on discussion behaviors (part 2, N=399)

	Model 8	Model 9
Constructive-(1)	0.136*	
Constructive-(2)		0.211**
Pretest	0.205***	0.198***
OLI_registration	0.225	0.214
Number of posts	0.007	0.005

($p<0.001$***, $p<0.01$**, $p<0.05$*)

After fitting regression models of learning outcome and all discussion behaviors as main effects, we found that only two categories are significant in predicting learning gains, as shown in Table 7. We consider higher frequencies of both behaviors might be the reason leading to significant effects on learning. Nevertheless, the processes of proposing an idea or a problem, and providing examples and reasons to justify a claim are considered to be higher-order thinking behaviors that have been proved to be instrumental to learning in several studies [e.g., 7, 18, 22], which could also lead to the significant effects.

6. CONCLUSION AND DISCUSSION

In this paper we adopted a content analysis approach and developed a coding scheme to analyze students' discussion behaviors, which are hypothesized as relating to their underlying cognitive processes in the discussion forum of a MOOC. The learning gains measures available for students in this MOOC enabled us to explore the relationship between students' discussion behaviors and their learning, and discuss what aspects of discussion appear to contribute to learning.

We observed that students' active and constructive discussion behaviors are significant in predicting students' learning gains, with active discussion behaviors possessing better predictive power, which is inconsistent with our hypotheses. Interactive discussion behaviors are significant in predicting learning gains only for students who are less active in the forums. This work also provides insight into how students are discussing in the discussion forum now, what behaviors they demonstrate and what the underlying cognitive processes are.

6.1 Active-Constructive-Interactive framework

Based on Chi's framework [8], we hypothesized that students' interactive discussion behaviors will produce more learning gains than constructive behaviors, and constructive behaviors will produce more learning gains than active behaviors. However, in this analysis we found that students' active discussion behaviors are most effective in predicting students' learning gains (coefficient=0.125). In our categorization of active behavior, students are talking about what is already covered in the materials, repeating statements that had appeared in the textbook or video lectures, and asking clarifying questions about definitions, implicitly expressing confusion about course materials, etc. According to Chi's framework [8], constructive activities should provide better learning outcomes than active activities. An example of this is when students need to explain in a constructive condition. However, we consider one reason we may not have seen this pattern in our dataset is that the post-test may not have targeted the skills and concepts students learned from these constructive activities. Assessments of a different nature, for example incorporating more demanding open ended response items, may have been more sensitive to these gains. For example, when the learning task is about design of psychology experiments, an assessment of requiring the students to design an actual experiment might be more telling than multiple-choice questions in measuring students higher-order thinking skills.

6.2 Invisible learning practices

In this paper, we looked at students' overt discussion activities in the forum, however students may be engaged in these higher order thinking activities without articulating their reasoning in a visible discourse. As indicated by [3], reading but not necessarily posting can be a productive practice for some learners. Our estimates of the amount of videos, quizzes and OLI textbook pages attempted could also be improved, for example, using the time spent on each activity, and further details about the attempt of OLI activities could be incorporated, as defined and estimated in [21].

6.3 Design implications

As MOOCs evolve, our focus as a community will transition from a primary concern about retaining users to actively improving the pedagogical effectiveness of this learning environment. Thus we need an empirical foundation to base designs for discussion affordances in MOOCs that might facilitate constructive and interactive conversations. Also, we need to come up with better assessment methods to assess and acknowledge students' higher-order thinking behaviors and skills they acquired through reading others' ideas, explaining and arguing in a discussion forum.

The paper proposes a manual way to hand-code students discussion behaviors, and offers a machine learning model to predict the corresponding behaviors in all communications of the dataset. We haven't had the opportunity to test the model in other courses, as few courses have pre- and post-test measures. If the computational model can be applied, we may provide feedback on students' advanced discussion behaviors in the forum, in terms of their cognitive processes and strategies.

7. ACKNOWLEDGEMENTS

This research was funded in part from NSF DATANET grant 1443068 and a grant from Google.

8. REFERENCES

[1] Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. *In Proceedings of the 23rd international conference on World wide web* (pp. 687-698). International World Wide Web Conferences Steering Committee.

[2] Barab, S., & Duffy, T. (2000). From practice fields to communities of practice. In D. H. Jonassen & S.M. Land (Eds.), *Theoretical Foundations of Learning Environemnts*.

[3] Beaudoin, M. F. (2002). Learning or lurking?: Tracking the "invisible" online student. *The internet and higher education*, 5(2), 147-155.

[4] Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, 8(1), 13-25.

[5] Brinton, C., Chiang, M., Jain, S., Lam, H., Liu, Z., & Wong, F. (2013). Learning about social learning in MOOCs: From statistical analysis to generative model, *7*(4), 346–359.

[6] Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The journal of the learning sciences*, 6(3), 271-315.

[7] Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Erlbaum.

[8] Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73-105.

[9] Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T.,&Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25, 471–533.

[10] Chi, M. T. H., & VanLehn, K. a. (2012). Seeing Deep Structure From the Interactions of Surface Features. *Educational Psychologist*, *47*(3), 177–188.

[11] Clow, D. (2013). MOOCs and the funnel of participation. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp185-189). ACM.

[12] Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, 64, 1–35.

[13] Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14* (pp. 83–92). New York, New York, USA: ACM Press.

[14] Core, M. G., Moore, J. D., & Zinn, C. (2003). The role of initiative in tutorial dialogue. *In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)* (pp. 67–74). Morristown, NJ: Association of Computation Linguistics.

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

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

[17] Gillani, N., Eynon, R., Osborne, M., Hjorth, I., & Roberts, S. (2014). Communication communities in MOOCs. *arXiv preprint arXiv:1403.4640.*

[18] Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American educational research journal*, 31(1), 104-137.

[19] Henri, F. (1992). Computer conferencing and content analysis. *In Collaborative learning through computer conferencing* (pp. 117-136). Springer Berlin Heidelberg.

[20] Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. *In Proceedings of the third international conference on learning analytics and knowledge* (pp. 170-179). ACM.

[21] Koedinger, K.R., Kim, J., Jia Z., McLaughlin E., Bier, N. (2015). Learning is Not a Spectator Sport: Doing is Better than Watching for Learning from a MOOC. *Learning at Scale '15*.

[22] Mestre, J. P. (2001). Implications of research on learning for the education of prospective science and physics teachers †. *Physics Education*, *36*(1), 44–51.

[23] Meyer, K. (2004). Evaluating online discussions: four different frames of analysis. *Journal of Asynchronous Learning Networks*, 8(2), 101–114.

[24] Molenaar, I., Chiu, M. M., Sleegers, P., & van Boxtel, C. (2011). Scaffolding of small groups' metacognitive activities with an avatar. *International Journal of Computer-Supported Collaborative Learning*, *6*(4), 601–624.

[25] Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching*, *50*(3), 315–347.

[26] Rosé, C. P., Bhembe, D., Siler, S., Srivastava, R., & VanLehn, K. (2003). Exploring the effectiveness of knowledge construction dialogues. In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Artificial intelligence in education: Shaping the future of learning through intelligent technologies* (pp. 497–499). Amsterdam, the Netherlands: IOS Press.

[27] Shea, P., & Bidjerano, T. (2012). Learning presence as a moderator in the community of inquiry model. *Computers & Education*, 59(2), 316–326.

[28] Wen, M., Yang, D., & Rose, C. P. (2014, May). Linguistic reflections of student engagement in massive open online courses. *In Proceedings of the International Conference on Weblogs and Social Media*.

[29] Wen, M., Yang, D., & Rosé, C. P. (2014). Sentiment Analysis in MOOC Discussion Forums: What does it tell us. *Proceedings of Educational Data Mining.*

[30] Yang, D., Adamson, D., & Rosé, C. P. (2014). Question recommendation with constraints for massive open online courses. *Proceedings of the 8th ACM Conference on Recommender Systems - RecSys '14*, 49–56.

[31] Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81(3), 329–339.