Current Issues in Emerging eLearning

Volume 2
Issue 1 *MOOC Theoretical Perspectives and Pedagogical Applications*

Article 5

January 2015

Supportive technologies for group discussion in MOOCs

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

Pam Goldman
University of Pittsburgh, peg8@pitt.edu

Jennifer Zoltners Sherer
University of Pittsburgh, jzsherer@pitt.edu

Lauren Resnick
University of Pittsburgh, lauren.resnick@pitt.edu

SUPPORTIVE TECHNOLOGIES FOR GROUP DISCUSSION IN MOOCS

Carolyn P. Rosé[i] Carnegie Mellon University

Pam Goldman[ii] and Jennifer Zoltners Sherer[iii] Institute for Learning

Lauren Resnick[iv] Learning Research and Development Center

VISION

A key hurdle that prevents MOOCs from reaching their transformative potential in terms of making valuable learning experiences available to the masses is that they fail to provide the kind of social environment that is conducive to sustained engagement and learning. This is especially true as students arrive in waves to these fledgling online learning communities. Our research seeks to lay the foundation for meeting this challenge by beginning with a case study and computational modeling of social interaction data in order to yield new knowledge that would inform development of novel, real-time support for building healthy learning communities that foster a high level of engagement and learning.

Interactive, supportive technology exists for effectively supporting small group online collaboration (Kumar & Rosé, 2011; Adamson et al., 2014). Where small group interaction is already used within a MOOC context (such as supported by NovoEd https://novoed.com/), such technology could be imported wholesale. However, in more typical forms of MOOC social interaction, such as in the threaded discussions, the positive effects of these forms of support may not generalize, and therefore the support may need to be substantially adapted. Thus, what we consider in this chapter is how we might build on the prior success with dynamic support for small group collaboration in order to work towards a new generation of more socially supportive MOOCs. The ultimate goal of this new context sensitive support is to yield more resilient massive scale online learning environments. We must consider what such support might look like. To that end, we begin with what exists for supporting small group collaboration (e.g., Kumar & Rosé, 2011; Adamson et al., 2014) and consider what we have learned from a case study in an xMOOC with over 60,000 students enrolled that suggests how such technology might be applied.

The proposal is not to replace humans with technology but to extend the capabilities of human effort through technology in order to use that valuable human contact in a cost effective way. In so doing, we draw from the literature on classroom discussion to motivate design for highly successful facilitative support. The design of such support is consistent with the literature on facilitation of collaborative learning groups (e.g., Hmelo-Silver & Barrows, 2006), and leverages a large body of work that has shown that certain forms of classroom discussion facilitation, termed Accountable Talk®, otherwise known as Academically Productive Talk, are beneficial for learning with understanding (Adey & Shayer, 1993; Bill, Leer, Reams, & Resnick, 1992; Chapin & O'Connor, 2004; Resnick, Asterhan, & Clarke, in press; Topping & Trickey, 2007a, 2007b; Wegerif, Mercer, & Dawes, 1999). This facilitation technique highlights a set of core facilitation moves that teachers employ as tools to encourage group knowledge integration among students in whole class discussions. In classrooms where Accountable Talk is used, students learn to reason together and use each other's thinking as a resource to scaffold their own. Achieving the potential for such an effect in MOOCs would be a substantial improvement over what mainstream MOOCs currently offer.

The unique developmental history of MOOCs creates challenges that cannot be met without insight into the inner-workings of massive scale social interaction. In particular, rather than evolving gradually as better understood forms of online communities, MOOCs have a rather abrupt start. The bulk of the student population that participates in a MOOC signs up for participation substantially before the launch date. When the launch date comes, the community springs up all at once, with potentially tens of thousands of new initiates, and no substantial community core beyond the instructor and TAs. Then they expand in waves as new cohorts of students arrive from week to week to begin the course. As massive communities of strangers that lack shared practices that would enable them to form supportive bonds of interaction, these communities grow in an organic manner. While some students may successfully find like-minded students with whom to bond and find support, when others come they may find an overwhelming amount of communication having already been posted that they feel lost in. Others may find themselves somewhere in between these two extremes. They may begin to form weak bonds with some other students when they join, however, massive attrition may create challenges as members who have begun to form bonds with fellow students soon find their virtual cohort dwindling (Rosé et al., 2014; Yang et al., 2014a). Early attempts to organize the community into smaller study groups may be thwarted by such periodic growth spurts paired with attrition, as groups that initially had sufficient human resources to accomplish their tasks soon fall below that level and then are unable to support the needs of remaining students.

Our research is leading us towards development of new interventions to enhance MOOC platforms, which entails development of new technology (Yang et al., 2014b). An overarching goal is development of a communication medium that can accommodate emerging community/sub-community structure -- supporting the formation and sustenance of supportive relationships. Ultimately, what we aim to foster is communities of inquiry (Garrison, 2010) where students build supportive bonds with one another so that they are able to reason together and use each other's thinking as resources as in Accountable Talk classrooms. We seek to accomplish this through provision of synchronous discussion opportunities and support for more efficient help seeking.

THEORETICAL AND TECHNICAL FOUNDATION

One of the great challenges of making MOOCs more effective is dealing with the massive scale while considering limitations on the instructors and teaching staff for offering support. A key aspect of our vision is to take advantage of the tremendous support resources that the students themselves bring to the table that they could offer each other. Work towards leveraging support from students has already shown promise in the field of Computer Supported Collaborative Learning. Though it is frequently noted that unsupported collaboration may suffer from many interpersonal difficulties, we have observed in a variety of collaborative tasks how students benefit in their pre to post-test gains by almost a full letter grade just from having a fellow student to work with (Gweon et al., 2006; Kumar et al., 2007). Above and beyond that, support for effective collaboration improves learning in collaborative contexts (Gweon et al., 2006).

Until recently, the state-of-the-art in computer supported collaborative learning consisted of static forms of support, such as structured interfaces, prompts, and assignment of students to scripted roles (see Fischer et al., 2013 for a review). Most recently, a decade of research in this vein, including our own (Gweon et al., 2006; Kumar et al., 2007; Kumar & Rosé, 2011; Dyke et al., 2013; Adamson et al., 2014), shows that students can benefit from their interactions in learning groups when automated support is provided, especially dynamic, interactive and context sensitive support. Dynamic forms of collaboration support "listen in" on student conversations in search of important events that present opportunities for discouraging dysfunctional behavior or encouraging positive behavior using automated analysis of collaborative learning processes (Kumar et al., 2007; Adamson et al., 2014).

The theoretical foundation for the work begins by considering what makes discussion for learning effective. The answer to the question will vary depending upon whether one comes from a cognitive (Webb, 2013), developmental (Golbeck & El-Moslimany, 2013), or sociocultural (Hakkarainen et al., 2013) perspective. In our work, we adopt a sociocognitive perspective (Howley et al., 2013), emphasizing the value in making reasoning public and encouraging students to listen carefully to and build on one another's reasoning (de Lisi & Golbeck, 1999). As students engage in this type of exchange, they have the opportunity to observe discrepancies between their own mental model and those of other students. This exchange thus provides opportunities to experience cognitive conflict and learning. They also have the opportunity to take ownership over knowledge and position themselves as valuable sources of knowledge within the interaction (Howley et al., 2011).

In the classroom discourse community, work on teacher facilitation techniques suggests methods for encouraging these types of interactions between students. One such framework, termed Accountable Talk, or Academically Productive Talk, has demonstrated striking results in terms of precipitating steep increases in achievement in the domain of instruction, with transfer to other domains that persists for years (Adey & Shayer, 1993; Bill, Leer, Reams, & Resnick, 1992; Chapin & O'Connor, 2004; Resnick, Asterhan, & Clarke, in press; Topping & Trickey, 2007a, 2007b; Wegerif, Mercer, & Dawes, 1999). In our recent work, we have adapted some of these teacher facilitation practices in automated support for collaborative learning that uses intelligent software agents to provide the facilitation for student groups (Adamson et al., 2014). When the automated facilitation provided support that is well adapted to the needs of the student groups, we find pockets of intensive interaction among students following facilitation moves (Dyke et al., 2013; Adamson et al., 2014). In addition to the local effect on small groups of collaborating students, we see that students who have experienced a supported small group collaboration experience bring the positive effect back to whole class discussion (Clark et al., 2013). This result connecting small group and large group effects suggests that technology for supporting small group online collaboration may have a direct application in MOOCs where students are assigned to work in small groups for some assignments. However, our larger goal is to apply the general principle more directly to all forms of social interaction in MOOCs. For example, if we provide intensive synchronous collaboration experiences in the midst of the MOOC learning experience, we hope the positive effects will be felt also within the asynchronous threaded discussion forums.

An essential enabling technology in our work on dynamic support for collaborative learning has been development of technology for automated analysis of discussion for learning (Donmez et al., 2005; Rosé et al., 2008; Mu et al., 2012; Gweon et al., 2013). In this work we focus on automatic application of multi-dimensional frameworks for characterizing collaborative learning processes (Howley, Mayfield, & Rosé, 2013; Howley, Mayfield, Rosé, & Strijbos, 2013). More recently we have developed techniques for automated analysis of threaded discussions in MOOCs, including analysis of emergent discussion groups (Yang et al., under review; Kumar et al., under review), indicators of motivation and cognitive engagement (Wen et al., 2014a), and expressed attitude towards the course and course tools (Wen et al., 2014b).

In our work using automated analyses to dynamically trigger agent based facilitation to support collaboration, as well as to support our sensemaking in interpretation of data for research purposes, we have made use of supervised, semi-supervised, and unsupervised machine learning methods. In our experience, supervised models, which are models trained with data hand annotated with theory motivated labels has stronger predictive validity (Rosé et al., 2014b; Wen et al., 2014a; Wen et al., 2014b), but it is also helpful from a research perspective to employ exploratory models that can offer a bird's eye view of the data that do not require pre-determined hand labels. To that end, some of our recent work has focused on the development of probabilistic graphical modeling techniques for identifying emergent community structure in MOOC discussions (Yang et al., 2014a; Yang et al., under review) and other types of threaded discussions (Kumar et al., under review). The goal here is to use cues from who is interacting with whom (using social network structure that can be constructed from reply links within threads) integrated with models of which themes are discussed (using topic modeling tools) in order to identify emerging groups with topical foci that reflect the values they share when they are interacting.

From a technical perspective, our modeling approach integrates two types of probabilistic graphical models. First, in order to obtain a soft partitioning of the social network of the discussion forums, we used a Mixed Membership Stochastic Blockmodel (MMSB) (Airoldi et al., 2008). The advantage of MMSB over other graph partitioning methods is that it does not force assignment of students solely to one sub-community. The model can track the way students move among sub-communities during their participation. Some of our earlier work made use only of this portion of the model (Rosé et al., 2014a). Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is a probabilistic topic model that we used to estimate for each person a distribution of identified communicative themes that mirrors her pattern of interconnectivity in the community. In our recent work, the fuller integrated exploratory technique was useful for identifying emerging groups within discussion threads with significantly higher or lower attrition than average (Yang et al., under review). The text component of the model provides some measure of anchoring and allows us to identify the topics of discussion shared by students in groups that exhibit higher or lower attrition than average. If we can identify these vulnerable sub-communities as they emerge, we can channel more support and hopefully mitigate some of the dropout risk.

In addition to basic research in machine learning applied to problems in conversation analysis, an important contribution of this work we offer to the MOOC research community is development of two publically available tool kits that are in wide use, namely TagHelper tools (Rosé et al., 2008) and LightSIDE

(Mayfield & Rosé, 2013), each of which have been downloaded over 5,000 times from over 70 countries. A recent survey sent out to the 200 most recent downloaders of LightSIDE indicates that 39% of respondents continue to use LightSIDE regularly after downloading. Both tool kits provide a convenient graphical user interface environment for novice users of text classification technology that easily runs text extraction and classification experiments with a few clicks. On top of that, LightSIDE serves as a vehicle for dissemination of new techniques for effective application of machine learning to text mining, including novel feature extraction techniques (Gianfortoni et al., 2011). The newest version (LightSIDE 2.0) includes a model specification panel that enables easy use of multi-level modeling techniques from applied statistics as domain adaptation and multi-domain learning approaches. One of the most unique capabilities of LightSIDE is its sophisticated support for error analysis.

Another important aspect of our work has been architectures for managing real time support for online discussion. A major aspect of this research began with the development of the Basilica architecture, which facilitates rapid development of multi-party collaboration environments. A recently published journal article (Kumar & Rosé, 2011) describes a series of collaborative environments developed through this architecture using reusable components. An improved version of the architecture, referred to as Bazaar, is now freely available online and includes instructional materials that enable others to learn to use it quickly (Adamson & Rosé, 2012). It has been used over the past couple of years in tutorials and short training courses with undergraduates and graduate students. Experience in these venues demonstrates that even undergraduate students are able to learn to use Bazaar and build prototype dynamic support for collaborative learning within one or two weeks. A key contribution of this research is development of a method for coordinating real time automated behavior monitoring with decision making about providing support and advice for shaping behavior. It's effective coordination algorithms simultaneously consider the history of participation, the current state of the student's engagement, and planned future events related to instruction and support.

The most productive realization of this technology to MOOCs may not be the most straightforward mapping of its usage in small group chat to this environment, namely using intelligent software agents to submit automated messages to threaded MOOC discussions. These contributions might be regarded as spam. Beyond that, it would not be clear which students would even see the messages or when. Thus, a more effective realization of such technology might be to use automated analyses of student behavior in MOOCs to trigger personalized recommender mentor agents that would provide opportunity suggestions to students as a side bar. Alternatively, automation could involve trigger events that are delivered to facilitators or instructors about discussions that need attention. Solutions to problems related to balancing multiple concerns in triggering supportive interventions that have been an important part of the development of the Bazaar architecture could still be leveraged nevertheless.

CASE STUDY

A course developed by the University of Pittsburgh's Institute for Learning, entitled 'Accountable Talk®: Conversation that Works' provides data for our case study. It was developed as a seven-week course in the Coursera platform and launched in October of 2013. This course was conceived as a teacher professional development course primarily for K-12 instructors. An earlier version of the course had been offered a number of times in Moodle to groups of 20 or fewer K-12 teachers or administrators. It was always a highly interactive course in which threaded discussions were a central learning activity and attrition was all but non-existent. The course instructors anticipated that offering the course through Coursera would reduce the potential for intensive teacher-student engagement. In order to partly compensate for this, additional instructional resources were offered to students to supplement the materials that were part of the original course.

The structure of the course followed a traditional xMOOC pattern. Students took an optional demographic pre-course survey at the beginning of their participation consisting mainly of standard Coursera pre-course survey questions. Each week, they did readings, watched videos, took quizzes, and participated in two types of threaded discussion. One was meant to focus on thoughtful discussion about the week's theme; the other was meant to be related to personal experience putting the principles taught by the course into practice. There were two substantial peer graded assignments that students completed. At the end of the course, students took an optional post-course survey.

More than 60,000 students signed up for the course, approximately 51,000 of which were still enrolled in the course at its completion, although fewer than 3% of students completed all the assignments. Of the 60,000 who signed up, only about 25,000 students accessed the course materials at least once. Of those students, only about 5% of the students ever posted to the discussion forums. At the completion of the course, 4,709 posts had been contributed, and the average number of posts contributed by forum participants was four. Contrary to expectation, of the students who self-declared a profession, only 18% of them were K-12 instructors. Instead, the very diverse student population (both in terms of national origin and profession) included not only the full spectrum of instructors (including college professors), it also included doctors, lawyers, other professionals, retired people, and a large contingent of parents of teen-agers. Participants who were not teachers frequently declared that their hope in participating was that they would gain better communication skills that would be valuable in their work and families.

What was unusual about the administration of the course was the level of dedication of the two course instructors in supporting the threaded discussions. They took shifts watching the discussion around the clock, and frequently stepped in to offer feedback and guide the discussions. Based on their observations, they created new videos to offer just in time instruction and advice, and to address issues that came up along the way. For example, in response to a trend of student remarks about difficulties keeping up with the large amount of material offered in each week, the instructors offered suggestions for selective participation. When the instructors noticed interesting discussions occurring in the discussion threads, they posted podcasts where they highlighted these discussions and offered their own commentary. Furthermore, the instructors very quickly noticed the difference between the expected student profile and the diverse population of students who indeed participated in the course. They were quick to respond by setting up special threads in each sub-forum dedicated to specific student populations such as professionals, parents, K-12 instructors, post-secondary instructors, and non-traditional students. The goal was to aid students in finding like others with whom to interact.

The instructors noticed that the depth of discussion in the threads each week was far less deep and substantial than what they had encountered in earlier Moodle deployments. In particular, interspersed with serious, thoughtful contributions were purely social contributions. Sometimes discussions about teaching practices degenerated into complaint sessions about school administration. These contributions did not go unappreciated, however. Some of these complaint posts were among the most highly up-voted posts. Nevertheless, there were also comments in the post-course survey that indicated that some students were dismayed that other students were not taking the threaded discussions as seriously as they were.

Though the discussions were peppered with off topic discussion about life issues, the tone remained largely jovial. On the other hand, some antisocial behavior was also noted. One participant in particular started a number of belligerent threads, expressing opinions in a disrespectful way. The instructors took note of this student and eventually attempted to offer feedback to curb his behavior after other students started to file complaints. Unfortunately, this intervention was not successful in shaping his behavior for the better, and ultimately he was removed from the course. The posts he contributed in the meantime were nevertheless left within the discussion forums, and precipitated reactions from other students, some of which were disrespectful as well. We see here the dangers of antisocial seeds being sewn. The lingering effect eventually dissipated; however, it is not yet clear what the ripple effect of this antisocial behavior was on other students who interacted directly with the posts or indirectly as they read. Antisocial behavior degrades the quality of the environment for others. However, if we put a positive spin on it, we can take the existence of such behavior as a research challenge and seek to turn dysfunction into learning opportunities (or at least contain it so that it is not harmful to others).

A frequent positive comment in the post-course survey was to note and appreciate the dedication of the course instructors in terms of the intensity of their involvement in the threaded discussions. Nevertheless, the course instructors were overwhelmed by the sheer volume of contribution despite the fact that such a small percentage of enrolled students ever participated in the discussion forums. Despite their best efforts, there were as many complaints from students in the post-course survey describing that they felt they were speaking into the air when they posted and that they struggled to find people to have a meaningful discussion with or they felt as overwhelmed and lost in the sea of posted messages as the instructors also felt trying to keep up with it.

The work of the instructors in supporting the discussions was aided by regular technology enhanced reports of behavior trends within their MOOC, which were discussed at weekly or biweekly meetings throughout the duration of the course. Using data mining techniques such as survival analysis and topic analysis applied to data scraped from the discussion forums, the frequent meetings discussed trends in discussion focus, apparent student interests and concerns, and student dropout. Figure 1 displays a rudimentary representation that was commonly used to support these discussions.

In order to create this representation, Latent Dirichlet Allocation (LDA) was applied to the data that accumulated since the course's launch date. Each post was treated as a separate document. LDA was configured to identify 10 themes. Each message was coded as discussing whichever of these 10 themes was rated by LDA as explaining the source of the largest percentage of words in the post. For each day of the course, we computed a distribution of topics based on the percentages of messages assigned to each topic. The line graph in Figure 1 displays the topic trends over time. Though the pattern appears messy, what is clear is that which topics are dominant varies from week to week, although new students were starting the course in each of these weeks, and few students survived until the end, which resulted in week 1 relevant topics remaining prominent throughout. The meaning behind the color coding came out by examining the thread titles for threads that exhibited a high proportion of each associated topic. This association enabled the instructors to make meaning out of the graphs. In addition to lists of prominent thread titles, exemplar messages exhibiting some concrete ideas represented by each topic were also presented to the instructors.

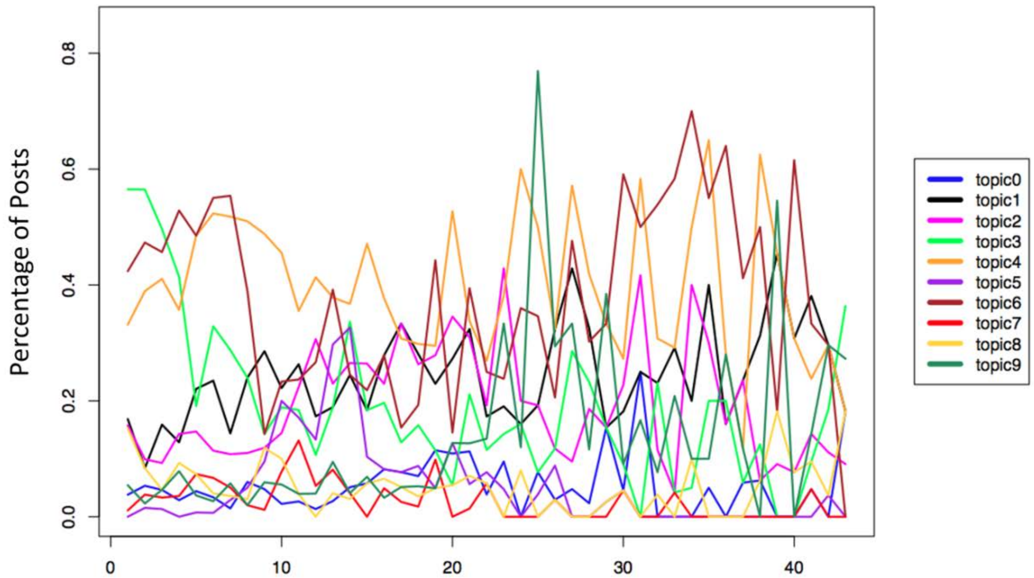

Figure 1: This line graph displays trends in discussion topics over time from the day the course was launched (day 0) until midway through week 7.

One prominent topic early in the course and then again late in the course was Topic 6, which was discussion focused on one particular video that the students found particularly inspiring and thought provoking. Topic 4, which remained consistently prominent throughout the course, was focused on discussion of nontraditional students (i.e., students other than K-12 instructors) problem solving together about how to apply the material of the course to their lives. Prominent threads in this topic included stories from a variety of types of nontraditional students applying the principles, one thread about getting shy people to talk, and another one about managing corporate meetings effectively. Topic 5, where participants talked about their experiences applying the principles, had a spike in the middle of week 2 and then diminished. These threads focused on a variety of types of instructors and their stories. Other threads focused on adopting a growth mindset and emotional intelligence, topics introduced in Week 2 of the course.

We observed the usual pattern of a significant spike in activity during the initial week of the course. However, upon closer inspection, one lesson learned through application of simple data mining techniques as the course was in progress was that the pattern of rate of dropout associated with the week when a student began active participation in the course appears quite different. In particular, students who began their active participation in the first week showed a pattern of lower attrition over time than their counterparts in later cohorts. In order to make this assessment, we used a survival analysis. Survival analyses are known to provide less biased estimates than simpler techniques (e.g., standard least squares linear regression) that do not take into account the potentially truncated nature of time-to-event data. In a survival model, a prediction about the likelihood of a failure occurring (in this case, ceasing to actively participate) is made at each time point based on the presence of some set of predictors. The estimated weights on the predictors are referred to as hazard ratios. The hazard ratio of a predictor indicates how the relative likelihood of the failure (e.g., dropout) occurring increases or decreases with an increase or decrease in the associated predictor.

In order to distinguish the attrition rate from week to week of student groups that started their participation on different weeks, we constructed 7 binary variables we refer to as cohort variables. For each data-point for each student, the cohort variable referring to the week of the student's initial active participation was set to true for each of the student's data points while the other cohort variables were set to false for that student. The inclusion of these variables made it possible within the survival analysis to assess the association between rate of attrition and week of first active participation. Only the variable that indicated that a student began their active participation in the first week of the course made a significant prediction. Its hazard ratio was .65, which means that students who began their active participation in the first week of the course were 35% less likely to drop out on each subsequent time point than the population average. Based on demographic information gleaned from the surveys of those who completed them, this pattern did not appear to be explained by any significant difference in the type of student who joined late versus early. Some comments from the post-course survey indicated however, that students who arrived late struggled to find their place in the large amount of communication that had been contributed before their arrival. This suggests some potential importance to the design and management of the threaded discussion for keeping students engaged. Consistent with this, participation in the discussion forums appeared to be an indication of commitment to the course. Of those students who completed the course, 50% of them participated in the discussion forums at least once, whereas only 1% of non-completers did. The design space for more conducive threaded discussions integrated with the needed support is still wide open. Exploring this space is an important direction for continued research in this area.

FUTURE DIRECTIONS

The specific goal of our research is to develop technology capable of supporting effective participation in conversation to achieve a positive impact on human learning, growth, and well-being. Our conviction is that in order for the technology to achieve maximum impact, it must first be capable of processing, generating, and engaging in conversation. Second, its behavior should be designed with a deep understanding of the mechanics of what makes conversation work in different settings as well as an understanding of what properties of conversation add to or detract from its positive impact on important outcomes of conversation. Finally, its design should be based on knowledge of what external stimuli manipulate these properties of conversation and in what ways.

In our prior work we followed this research program to develop technology for effective support of online small group collaborative learning. A decade of such research has produced both substantial empirical results as well as publicly available tools and technologies to share with other researchers. For MOOC providers that already offer small group synchronous interaction as an affordance for course participation, the work we have already done may already contribute to their foundation. In this chapter we begin to follow the same research program, but target discussion for learning in a more typical xMOOC context where the primary mode of social interaction is asynchronous threaded discussions. A case study from one specific MOOC discussed above offers insights into the limitations of learning through discussion in current xMOOCs, as well as the challenges in instructor ability to overcome those limitations without support.

We identify two main areas for potential impact of new technology. The first need is to engage more of the MOOC participants in discussion. Research from the collaborative learning community (e.g., Hmelo-Silver et al., 2013) as well as the classroom discourse community (Resnick et al., in press) offer convincing evidence for the value of discussion for learning. Correlational results suggest that active participation in MOOC discussion forums is associated with lower attrition. However, the great majority of MOOC participants never post even once to the discussion forum. Research on vicarious learning offers demonstrations that students may learn by watching other students engaging meaningfully in interaction for learning (Chi et al., 2008), and one might conjecture that students who did not actively engage in the discussions nevertheless were able to benefit from them. However, responses to a post-course survey question related to why the student did not participate or did not participate more in the discussion forums suggest that students who did not participate did not find value in the threaded discussions for a variety of reasons that would challenge this conjecture. Frequent student responses identified a lack of time to engage or difficulty finding opportunities that were of interest in the overwhelming numbers of available threads. This suggests that participation in the forums could be increased through a computer-generated personalized recommendation that would draw attention to opportunities for engagement that might be of interest to each individual student.

Another issue uncovered in the case study is that some students who did participate in the forums were dissatisfied with the response they received. On the positive side, in this course where the instructors were unusually dedicated and involved, their interventions to direct students to focus more deeply on the course material and to engage more substantively with one another were met with universal positive response wherever they were mentioned in the post course survey, and led to positive effects locally within the conversations where they were involved. This suggests that some form of facilitation in a MOOC context may have value. However, we see that human effort can go only so far. And if we are successful in eliciting more participation, the problem will be exacerbated. The personalized recommendation paradigm mentioned above might be useful for this problem as well. In particular, in this case a recommendation that draws attention to one type of opportunity may also serve to draw attention away from other experiences. Thus, a form of social recommendation could be used as a major organizing force for interaction, creating spaces where the contribution participants are prepared to make is directed to recipients who will appreciate it.

While recommendation technology has been used widely in online instruction (Manouselis et al., 2013), work on social recommendation has been sparse. In our work we have already begun to lay the foundation for social recommendation in MOOCs to address the issues raised above (Yang et al., 2014b). We invite the larger MOOC research community to join in this endeavor, as we believe more and better participation will improve learning for MOOC students.

REFERENCES

Adamson, D. & Rosé, C. P. (2012). Coordinating Multi-Dimensional Support in Conversational Agents, *ITS 2012 Proceedings of the 11th International conference on Intelligent Tutoring Systems, Lecture Notes in Computer Science Volume 7315*, Springer-Verlag, pp 346-351.

Adamson, D., Dyke, G., Jang, H. J., Rosé, C. P. (2014). Towards an Agile Approach to Adapting Dynamic Collaboration Support to Student Needs, *International Journal of AI in Education* 24(1), pp91-121.

Adey, P., & Shayer, M. (1993). An Exploration of Long-Term Far-Transfer Effects Following an Extended Intervention Program in the High School Science Curriculum. *Cognition and Instruction, 11*(1), 1 - 29.

Airoldi, E., Blei, D., Fienberg, S. & Xing, E. P. (2008). Mixed Membership Stochastic Blockmodel, *Journal of Machine Learning Research, 9(Sep):1981—2014.*

Bill, V., Leer, M., Reams, L., & Resnick, L. (1992). From cupcakes to equations: The structure of discourse in a primary mathematics classroom. *Verbum, 1, 2,* 63-85.

Blei, D., Ng, A. and Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research* (3) 993-1022.

Chapin, S., & O'Connor, C. (2004). Project challenge: Identifying and developing talent in mathematics within low-income urban schools *Boston University School of Education Research Report No* (Vol. 1, pp. 1-6).

Chi, M., Roy, M., & Hausmann, R. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning, *Cognitive Science* 32(2), pp 301-341.

Clarke, S., Adamson, D., Howley, I., Rosé, C. P. (2013). The Impact of CSCL Beyond the Classroom, *Proceedings of Computer Supported Collaborative Learning*

de Lisi, De Lisi, R. & Golbeck, S. L. (1999). Implications of Piagetian Theory for Peer Learning, in O'Donnell & King (Eds.) *Cognitive Perspectives on Peer Learning*, Lawrence Erlbaum Associates: New Jersey

Donmez, P., Rose, C. P., Stegmann, K., Weinberger, A., and Fischer, F. (2005). Supporting CSCL with Automatic Corpus Analysis Technology**,** *CSCL '05 Proceedings of the 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!*, pp 125-134

Dyke, G., Adamson, A., Howley, I., & Rosé, C. P. (2013). Enhancing Scientific Reasoning and Discussion with Conversational Agents, *IEEE Transactions on Learning Technologies* 6(3), special issue on Science Teaching, pp 240-247.

Fischer, F., Kollar, I., Stegmann, K., Wecker, C., Zottmann, J. (2013). Collaboration scripts in computer supported collaborative learning, in Cindy Hmelo-Silver, Angela O’Donnell, Carol Chan, & Clark Chin (Eds.) *International Handbook of Collaborative Learning*, Taylor and Francis, Inc.

Garrison, D.R., Anderson, T., Archer, W. (2010) The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education*, 13(1–2), 5-9

Gianfortoni, P., Adamson, D. & Rosé, C. P. (2011). Modeling Stylistic Variation in Social Media with Stretchy Patterns, in *Proceedings of First Workshop on Algorithms and Resources for Modeling of Dialects and Language Varieties*, Edinburgh, Scottland, UK, pp 49-59.

Golbeck, S. & El-Moslimany, H. (2013). Developmental approaches to collaborative learning, in Cindy Hmelo-Silver, Angela O’Donnell, Carol Chan, & Clark Chin (Eds.) *International Handbook of Collaborative Learning*, Taylor and Francis, Inc.

Gweon, G., Rosé, C. P., Zaiss, Z., & Carey, R. (2006). Providing Support for Adaptive Scripting in an On-Line Collaborative Learning Environment, *Proceedings of CHI 06: ACM conference on human factors in computer systems.* pp 251-260, New York: ACM Press.

Gweon, G., Jain, M., Mc Donough, J., Raj, B., Rosé, C. P. (2013). Measuring Prevalence of Other-Oriented Transactive Contributions Using an Automated

Measure of Speech Style Accommodation, *International Journal of Computer Supported Collaborative Learning* 8(2), pp 245-265.

Hakkarainen, K., Paavola, S., Kangas, K., Setaama- Hakkarainen, P. (2013). Sociocultural perspectives on collaborative learning: Toward collaborative knowledge construction, in Cindy Hmelo-Silver, Angela O'Donnell, Carol Chan, & Clark Chin (Eds.) *International Handbook of Collaborative Learning*, Taylor and Francis, Inc.

Hmelo-Silver, C. E. & Barrows, H. S. (2006). Goals and Strategies of a Problem-based Learning Facilitator. *The Interdisciplinary Journal of Problem Based Learning*, 1(1), pp 21-39.

Hmelo-Silver, C., O'Donnell, A., Chan, C. & Chin, C. (2013) *International Handbook of Collaborative Learning*, Taylor and Francis, Inc.

Howley, I., Mayfield, E., Rosé, C. P. (2011). Missing Something? Authority in Collaborative Learning, in *Proceedings of the 9th International Computer Supported Collaborative Learning Conference, Volume 1: Long Papers*, pp 336-373

Howley, I., Mayfield, E. & Rosé, C. P. (2013). Linguistic Analysis Methods for Studying Small Groups, in Cindy Hmelo-Silver, Angela O'Donnell, Carol Chan, & Clark Chin (Eds.) *International Handbook of Collaborative Learning*, Taylor and Francis, Inc.

Howley, I., Mayfield, E., Rosé, C. P., & Strijbos, J. W. (2013). A Multivocal Process Analysis of Social Positioning in Study Group Interactions, in Suthers, D., Lund, K., Rosé, C. P., Teplovs, C., Law, N. (Eds.). *Productive Multivocality in the Analysis of Group Interactions*, edited volume, Springer.

Kumar, A., Palakodety, S., Wang, C., Wen, M., Rosé, C., Xing, E. (under review). Large Scale Structure Aware Community Discovery in Online Forums, submitted to *Knowledge Discovery and Data Mining.*

Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007). Tutorial Dialogue as Adaptive Collaborative Learning Support, *Proceedings of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, pp 383-390

Kumar, R. & Rosé, C. P. (2011). Architecture for building Conversational Agents that support Collaborative Learning, *IEEE Transactions on Learning* 4(1), pp 21-34

Manouselisr, N., Drachsler, H., Verbert, K., & Duval, E. (2013). Recommender systems for education, Springer

Mayfield, E. & Rosé, C. P. (2013). LightSIDE: Open Source Machine Learning for Text Accessible to Non-Experts, Invited chapter in the *Handbook of Automated Essay Grading*.

Mu, J., Stegmann, K., Mayfield, E., Rosé, C. P., Fischer, F. (2012). The ACODEA Framework: Developing Segmentation and Classification Schemes for Fully Automatic Analysis of Online Discussions. *International Journal of Computer Supported Collaborative Learning7(2)*, pp285-305.

Resnick, L. B., Asterhan, C. A., & Clarke, S. N. (in press). *Socializing Intelligence through Academic Talk and Dialogue*. Washington, DC: American Educational Reserach Association.

Rosé, C. P., Wang, Y.C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F., (2008). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning, *The International Journal of Computer Supported Collaborative Learning* 3(3), pp237-271.

Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P. & Sherer, J. (2014). Social Factors that Contribute to Attrition in MOOCs, *Proceedings of the First ACM Conference on Learning @ Scale (poster)*

Topping, K. J., & Trickey, S. (2007a). Collaborative philosophical enquiry for school children: Cognitive effects at 10-12 years. *British Journal of Educational Psychology, 77*(2), 271-288.

Topping, K. J., & Trickey, S. (2007b). Collaborative philosophical inquiry for schoolchildren: Cognitive gains at 2-year follow-up. *British Journal of Educational Psychology, 77*(4), 787-796.

Webb, N. (2013). Information processing approaches to collaborative learning, in in Cindy Hmelo-Silver, Angela O'Donnell, Carol Chan, & Clark Chin (Eds.) *International Handbook of Collaborative Learning*, Taylor and Francis, Inc.

Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: an empirical investigation of a possible socio-cultural model of cognitive development. *Learning and Instruction, 9*(6), 493-516.

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

Wen, M., Yang, D., Rosé, D. (2014b). Linguistic Reflections of Student Engagement in Massive Open Online Courses, in *Proceedings of the International Conference on Weblogs and Social Media*

Yang, D., Wen, M., Rosé, C. P. (2014a). Peer Influence on Attrition in Massively Open Online Courses, *Proceedings of Educational Data Mining*

Yang, D., Piergallinin, M., Howley, I., & Rosé, C. P. (2014b).Forum Thread Recommendation for Massive Open Online Courses, *Proceedings of Educational Data Mining*

Yang, D., Wen, M., Kumar, A., Xing, E., Rosé, C. P. (under review). Towards an Integration of Text and Graph Clustering Methods as a Lens for Studying Social Interaction in MOOCs, submitted to *The International Review of Research in Open and Distance Learning*.

[i] **Dr. Carolyn P. Rosé** is an Associate Professor of Language Technologies and Human-Computer Interaction in the School of Computer Science at Carnegie Mellon University. She is also the director of DANCE: and open source effort to design and build out affordances for collaborative learning in MOOCs (http://www.cs.cmu.edu/~cprose/DANCE.html). Her research group's highly interdisciplinary work, published in over 170 peer reviewed publications, is represented in the top venues in 5 fields: namely, Language Technologies, Learning Sciences, Cognitive Science, Educational Technology, and Human-Computer Interaction, with awards or award nominations in 3 of these fields. Specifically, her research program is focused on better understanding the social and pragmatic nature of conversation, and using this understanding to build computational systems that can improve the efficacy of conversation between people, and between people and computers. She serves on the executive committee of the Pittsburgh Science of Learning Center and the co-leader of its Social and Communicative Factors of Learning research thrust. She also serves as President Elect of the International Society of the Learning Sciences and Co-chair of its Computer Supported Collaborative Learning Community Committee. She serves as Associate Editor of the International Journal of Computer Supported Collaborative Learning and the IEEE Transactions on Learning Technologies. **(Corresponding Author)** Email: cprose@cs.cmu.edu

[ii] **Pam Goldman** is the Senior Instructional Design Fellow at the Institute for Learning, University of Pittsburgh. She led the development of the Institute's Coursera course. She she designs, authors, co-authors, produces, and edits a series of learning tools for educators based on the Institute's theory of action and practice of district and school reform and oversees the development of the IFL's suite of online workshops and a series of new online initiatives. Email: peg8@pitt.edu

[iii] **Jennifer Zoltners Sherer, PhD**, is an Instructional Design Fellow for the Institute for Learning at the University of Pittsburgh. In addition to designing tools and learning environments, she teaches Accountable Talk online workshops, writes, and supports the IFL's data and evaluation process. She co-designed and co-taught the IFL's Coursera course: Accountable Talk: Conversation that Works and co-developed the Process Engineering for Educational Results (PEER) project. Her research interests focus on distributed leadership and improving teaching and learning in K-12 systems through tool design and implementation, professional development, reform initiatives, and curriculum. She has published and presented on distributed leadership, organizational routines, and improving scientific understanding through the development of literacy skills. Prior to receiving her Ph.D. in Learning Sciences from Northwestern University, she was a teacher in Oregon. Email: jzsherer@pitt.edu

[iv] **Lauren Resnick** is a Distinguished University Professor of Psychology and Cognitive Science and also of Learning Sciences and Education Policy at the University of Pittsburgh. Dr. Resnick is founder and Co-Director of the Institute for Learning, which bridges the domains of research and practice by conveying to educators the best of current knowledge about learning processes, principles of instruction, and the design of school systems. She is an internationally known scholar in the cognitive science of learning and instruction and was Director of the Learning Research and Development Center at the University of Pittsburgh from 1977 to 2008. She has researched and written widely on the learning and teaching of literacy, mathematics, and science. Her recent work focuses on school reform, assessment, effort-based education, the nature and development of thinking abilities, and the role of talk and discourse in learning. Dr. Resnick is a prolific author, a respected editor, and a frequent consultant, with appointments to many national education boards, commissions, and associations. Recognized both nationally and internationally, Dr. Resnick has received multiple awards for her research. Email: lauren.resnick@pitt.edu