Should Your MOOC Forum Use a Reputation System?

Derrick Coetzee dcoetzee@eecs.berkeley.edu **Armando Fox** fox@cs.berkeley.edu **Marti A. Hearst** hearst@ischool.berkeley.edu **Björn Hartmann** bjoern@eecs.berkeley.edu
Computer Science Division and School of Information
University of California, Berkeley, CA 94720, USA

ABSTRACT

Massive open online courses (MOOCs) rely primarily on discussion forums for interaction among students. We investigate how forum design affects student activity and learning outcomes through a field experiment with 1101 participants on the edX platform. We introduce a reputation system, which gives students points for making useful posts. We show that, as in other settings, use of forums in MOOCs is correlated with better grades and higher retention. Reputation systems additionally produce faster response times and larger numbers of responses per post, as well as differences in how students ask questions. However, reputation systems have no significant impact on grades, retention, or the students' subjective sense of community. This suggests that forums are essential for MOOCs, and reputation systems can improve the forum experience, but other techniques are needed to improve student outcomes and community formation. We also contribute a set of guidelines for running field experiments on MOOCs.

Author Keywords

Massive open online course; MOOC; forum; reputation system.

ACM Classification Keywords

H.5.3. Information Interfaces and Presentation (e.g. HCI): Group and Organization Interfaces; K.3.1. Computers and Education: Computer Uses in Education

INTRODUCTION

Massive open online courses (MOOCs), such as those currently operated by Coursera, edX, and Udacity, typically include hundreds to thousands of students in each course. Traditional social structures of on-campus courses, such as direct interaction of teachers with each student, become infeasibly expensive at this scale. As Chamberlin and Parish lamented, "The number of people in a course can be overwhelming! How does someone connect with 2000-plus people?" [4].

CSCW'14, Feb 15-19 2014, Baltimore, MD, USA
ACM 978-1-4503-2540-0/14/02.
http://dx.doi.org/10.1145/2531602.2531657

A scalable alternative is interaction among students in an online discussion forum, in which students can assist one another. This model has been pervasive and successful in learning communities on the open web [36]. However, some forums are effective in creating learning communities while others are not, and a poor sense of community has been linked to high student attrition [39], which is observed in practice in MOOCs [8].

Large, popular question answering forums on the web, such as Yahoo! Answers and Stack Overflow, have been successful in generating communities with high participation, quick responses, and quality content [1, 24]. These forums generally rely on *reputation systems* [29], which provide extrinsic rewards such as "reputation points" for timely, high-quality contributions. We investigate whether forums in general — and reputation systems in particular — can help to promote learning and a successful student community in the MOOC context. We target five main research questions:

1. *Is forum participation correlated with student learning outcomes*, such as student retention and final grades? Online forums benefit students in traditional courses [5] — do they have similar effects in a MOOC?
2. *Do reputation systems improve forum participation metrics* like number of unanswered questions and response time? Prior work showed that Stack Overflow's reputation system helps to keep response times close to the minimum possible [24]. Can a reputation system achieve similar benefits in a MOOC forum?
3. *Do reputation systems affect the type of content posted on class forums?* Different incentive systems change the kinds of questions that are asked [19]. Will the reputation system alter how students ask questions?
4. *Do reputation systems affect student learning outcomes?* If reputation systems increase forum participation, students may have access to alternative explanations of concepts and spend more time explaining material to others (as in reciprocal teaching [28]), which could lead to improved learning as measured by course assessments.
5. *Do reputation systems affect the student's subjective sense of community*, their sense that they belong to a community of learners who they can rely upon? The incentives provided by reputation systems may have a positive effect if they lead to increased engagement, or a negative effect, if they interfere with students' intrinsic motivations and promote competition over cooperation [19, 12].

We investigate these questions in a between-subjects field experiment on a single course on the edX platform. $N = 1101$ participants were assigned to forums with and without reputation features. In line with previous work on traditional courses [5], we find that forum participation is correlated with higher grades and higher retention. Presence of forum reputation features significantly affects forum participation, increasing the number of responses and decreasing the time to first response. Reputation features also curtail certain language usage patterns like appeals for help. However, reputation features did not have a significant impact on students' assessment outcomes or their sense of community. Overall, these results suggest that reputation systems can improve the forum experience, but do not contribute substantially to assessment outcomes or community formation in the current MOOC context.

RELATED WORK

Relevant prior work includes research on forums in traditional classes; answering sites for informal learning communities; language use in online forums; and the current generations of MOOCs. We discuss each area in turn.

Students and Forums

Online discussion forums have been used to support on-campus courses since the early 1990s, when they were based on mailing lists and Usenet [7]. Effective forums contribute to learning through community building and collaborative dialogues. The cognitive apprenticeship model in learning theory, wherein more experienced peers and experts provide the minimal assistance needed to complete tasks until autonomy is achieved, helps to explain the benefits of forums for problem solving [20].

Students who voluntarily participate in forums have been found to achieve better exam grades and perform better overall, even when only viewing content and with minimal instructor interaction [5]. We expect that forum use will also improve student outcomes in MOOCs. Intent to participate in forums has been linked to two motivational factors: the perceived importance of learning, and peer pressure, the sense that peers are putting effort into the forum [41]. The importance of peer pressure suggests that even interventions that generate small changes to participation can have a larger indirect effect.

Q&A Sites and Reputation Systems

Question and answer (Q&A) sites provide responses to specific questions and often discourage discussion not related to a question. Nevertheless, they are used both for informational purposes (as intended) as well as for conversational purposes [18]. Popular examples include Yahoo! Answers, for general topics, and Stack Overflow, for software development. Investigations of existing sites typically focus on characterizing usage patterns through large scale data analysis, without comparing design alternatives, as researchers are unable to modify the system. For example, Adamic et al. characterize activity on Yahoo! Answers by clustering categories of questions based on text and user features [1]; Mamykina et al. measure response times on Stack Overflow and categorize users based on quantitative behavior metrics [24].

Q&A sites rely on reputation systems, in which participants receive extrinsic rewards such as "reputation points" when their contributions are rated as useful by others. Reputation systems are thought to encourage participation, increase contribution quality, and identify reliable information. They originate on electronic marketplaces where they help identify trustworthy participants [29], and have been used to support both student discussion forums [6] and social collaborative learning systems [32]. In Tausczik's investigation of Math Overflow [34], a Q&A site for mathematics research, "[u]ser behavior suggests that building reputation is an important incentive, even though users do not report this in the survey."

Reputation systems as extrinsic motivators may also have negative consequences. Dearman found that Yahoo! Answers users employed reputation as a cue to withhold answers – if the questioner was known for assigning low scores to increase their own points, then others would be reluctant to respond [11]. Tangible extrinsic motivators such as money or candy can also diminish feelings of community and bonding [12]. On the other hand, in a comparison of a for-fee market-based answer forum to a system with a top-10 leader board [19], the reputation-based system had more socially-conducive community building questions, suggesting that the discouraging effects of reputation points are not as severe as those of tangible extrinsic motivators.

Despite their ubiquity, there has been little systematic investigation into whether and how reputation systems benefit sites which use them. Researchers have however systematically studied the impact of other design features on online communities. Harper [16, 17] and Ludford [21] report on online field experiments motivated by social design to increase participation at the MovieLens film recommender site. Our work similarly is a field experiment designed to study the impact of specific design features (reputation systems) on course participation and assessment outcomes.

Communication in Forums

Our work investigates language differences between the two experimental forums. In particular, in the absence of the extrinsic incentives of a reputation system, students may have to take extra steps to encourage responses. Studies from social psychology have shown that demonstrations of gratitude can increase efforts to assist someone even when those efforts are costly, and that gratitude can increase assistance provided to strangers [2, 14]. Other language usage such as ending an information need with a question mark, explicitly scoping the audience, and being succinct, can also increase the likelihood of quickly receiving many high-quality answers on social networks [35]. As described in the results, we focus on appeals for help such as "thanks for any help," which can be interpreted as pre-emptive expressions of gratitude.

MOOCs

Massive open online courses (MOOCs) are online courses with open enrollment that typically have thousands of active students. They originated with the University of Man-

itoba's Connectivism and Connective Knowledge MOOC in 2008 [22], and have precursors in instructional video sites like Khan Academy [37]. Forums are pervasive in MOOCs and have been characterized as "an essential ingredient of an effective online course" [23], but early investigations of MOOC forums show struggles to retain forum users over time. In one example, half of forum users ceased participation due to factors such as lack of forum facilitation, an overwhelming number of messages, and rude behavior of other students [23]. Published research on MOOCs has thus far focused on reflections on the MOOC experience and the great number and diversity of students who participate [25, 9]. In contrast, we provide a systematic, comparative study of design alternatives and their impact on MOOC students.

Recently, researchers have begun to develop and evaluate new technology to facilitate learning in MOOCs. ARTful [33] uses EEG technology to track student attention and later uses this information to suggest material for review, while Pex4Fun [38] uses automated test generation techniques to assist in automated grading in programming courses. Although these techniques are valuable, they are presently limited by special hardware requirements and a restricted topic domain, respectively. We aim instead to evaluate enhancements to existing forum systems that are easy to deploy in current MOOC settings.

METHOD

We conducted a between-subjects field experiment on a seven-week, open-enrollment software engineering course offered on the edX platform ("CS169.1x: Software as a Service" from University of California, Berkeley).

Course Testbed

The course features 11 video lectures recorded during a corresponding on-campus course, 4 homeworks and 4 quizzes. Lectures incorporate short assessments that test the material just presented. The course has been offered three times on the Coursera platform and three times on the edX platform, each time with similar content.

All edX courses include a simple forum designed for answering specific factual questions with limited thread depth (at most three). Posts can be upvoted or reported for misuse, but not downvoted. Posts are listed in the order they were posted. Each post shows the author's name, when it was posted, and the number of upvotes. Users can be manually marked as moderators. Each user has a user page listing all their contributions; there are no reputation scores or badges. Threads are searched using full-text search, rather than tags, and can be ordered either by date or number of votes. Other MOOC providers such as Coursera have similar forum designs.

In the particular course under study, the forum is self-moderated by students and a set of community teaching assistants (community TAs), volunteers who took the previous offerings of the course. Instructors generate content but do not participate in the forum.

We modified the course software to replace the standard edX forum with two variants of an experimental forum. When students and community TAs first visit the forum, they are randomly assigned to one of the two experimental forum websites based on a hash of their username, and are unable to access the other. The assignment of participants to forums remains stable throughout the course, and the two forums share no content. Although using separate content databases leads to less available content for each forum, and divides the efforts of the community TAs, it avoids complex interactions between students using different features, and allows differences in passive participation to be examined.

Both sites were based on the open-source OSQA software [27], which is similar to Stack Overflow (see Figure 1). Like the edX forum, OSQA is designed for answering specific factual questions with thread depth at most three, rather than for extended complex discussions. This model encourages asking concrete questions on individual assignments over general class-related discussions. On one of the experimental forums, the *basic forum*, students can post questions, answers, and comments, search content using tags and full-text search, and follow or send private messages to other users. The *full-featured forum* additionally provides a reputation system with the following features:

- Users can upvote and downvote questions, answers, and comments to indicate the most useful ones. Answers with more upvotes are listed first.
- Users receive a reputation score based primarily on how many upvotes their posts receive. Reputation points reward users and highlight which users are reliable to others. As users gain reputation points, they gradually acquire access to moderation features on the forum.
- Users who ask questions can select a best answer among the answers.

An *anchor* is a piece of course material that naturally stimulates discussion, such as a homework or lecture. [15] We follow the best practice of making relevant discussion easily accessible from anchors by inserting forum links into anchor pages which show posts with related tags.

After the course completed, we sent all participating students an optional survey gathering qualitative information about the usefulness of the forum, usefulness of the reputation system features, their motivation for contributing to the forum, and their sense of community.

Participants

MOOCs are characterized by varying participation over the duration of the course, with considerable attrition between milestones [3]. The March 2013 course we examined exhibited this pattern, starting with 5985 enrolled students and ending with 532 passing students; students who use the forum depart somewhat more slowly, with 1572 students starting the course and 442 passing (see Table 1). First-Class-to-Course-Completion-Rate (FCCR), a metric devised by Russell measuring the ratio of the number of students who received a passing grade to the number who viewed the first video lecture, was 11%, lower than the 19% and 36% reported by Google for their Power Searching course, [13] suggesting higher attrition.

Figure 1. The full-featured forum has several voting and reputation features; the basic forum does not.

Milestone	Total students	Forum users	Non-forum users
Enrolled	5985	1572	4413
Viewed any lecture/HW/quiz	4963	1305	3658
Viewed first video lecture	3266	791	2475
Completed first quiz	2594	1127	1467
Started first homework	1874	985	889
Watched final lecture	1078	638	440
Started final homework	996	625	371
Completed final quiz	925	594	331
Received passing grade	532	442	90

Table 1. Number of students reaching various milestones during the course. Attrition occurs rapidly at the beginning and continues over time. Forum users (using either the experimental or edX forums at any point in the course) abandon the course more slowly, starting at 25% of students and ending at 64%. Most passing students (83%) used the forum.

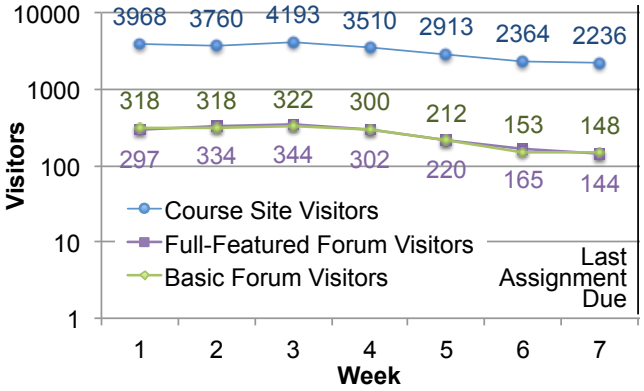

Figure 2. Forum and class visitors by week.

1101 students opted in to participate in the experiment, with 548 in the full-featured forum and 553 in the basic forum; 471 students who opted out instead used the standard edX forum. 3391 students (68%) viewed at least one lecture or assessment, but never visited any forum. Students opt in when they first visit the forum, and all subjects retained access to the standard edX forum to enable them to opt out of the experiment at any time. All students, regardless of whether they participated in the forum experiment, agreed to the edX terms of service which allow researchers to study their edX data. In any given week, 13%–17% of students who accessed the course website also accessed the experimental forum website, as shown in Figure 2.

Different information was available for different course participants. Aggregate grade and participation information was available for all students. For the 1101 students who consented to the experiment, detailed information about their forum usage is available. Additionally, 216 students opted in to a more permissive consent form that allowed us to correlate their detailed forum activity with their edX grades.

Hypotheses

Prior work indicates that students who use forums achieve better grades. Our first goal is to establish whether in the MOOC setting there is a connection between mere use of the forum and course outcomes. Because forum use was self-selected by students and not manipulated in our experiment, we can only show correlation, but such a correlation would support the potential usefulness of performing forum interventions and provide context for our other results. This gives two hypotheses:

H1. Forum participants will have higher course grades than non-participants.
H2. Forum participants will have higher retention than non-participants (they drop out later).

Next, we exploit our controlled forum intervention to seek strong connections between forum reputation features and important student outcomes: course grades and course retention. In addition, we also study factors directly linked to forum participation such as time to first response and number of responses. Such factors can help explain changes in outcomes (or the lack thereof). We also investigate the students' subjective *sense of community*. Reputation features could promote sense of community by encouraging more participation and by providing more information about which users are experienced in the context of a forum discussion.

H3. Reputation features will improve forum participation metrics, such as time to first response.
H4. Reputation features will improve grades.
H5. Reputation features will improve student retention.
H6. Reputation features will affect the student's subjective *sense of community*.

Although useful, simple quantitative forum metrics are unable to distinguish forum threads that have similar structures. Subjective metrics like sense of community are subject to bias as they are based on self-reported and incomplete data. For this reason we also sought to investigate differences in the *content* of the forums using text analysis techniques. Although many questions can be investigated with text analysis, for the purposes of this work we focus on one example that is straightforward to analyze, namely, whether students in the basic forum use more linguistic cues such as "can anyone help me" to persuade others to assist them. As noted in related work, we expect this because reputation systems offer extrinsic rewards to users posting answers, reducing the need for social rewards.

H7. Users of the basic forum will use more appeals for help in their questions.

Dependent Measures

Course Outcomes: Final grades and retention

We operationalize *retention* as days from the beginning of the course until the last student activity, defined as any access to any webpage associated with the course. We extract student activity from clickstream data provided by edX and forum activity, recorded by our own forum server.

Similarly to on-campus courses, a student's *final grade* is a summative assessment of their work throughout the course on a 0-100 scale, and students with a score of 50 pass. However, due to low completion rates, final grades in a MOOC have a different interpretation: they are a combined measure of the portion of the course that the student completed and the student's performance in that portion. For example, a student who scored 100% on the first assignment and then dropped out, and a student who scored 25% on all four assignments would receive the same final grade.

Although final grades remain useful as a combined metric, we also wish to effectively separate performance on completed work from attrition. One straightforward and flexible way to do this is to compare results on individual assessments, including only students who completed the assessment.

Forum Participation

We operationalize forum participation by following the model of prior analyses, e.g. [24]. In particular, we investigate *number of responses to questions*. A high number of questions without any responses indicates an inactive forum and has been linked to a low sense of community [10], while questions that receive many answers are indicative of multiple solutions or discussions. We also investigate *time to first response* – the interval between the time a new question is posted, and the time the first answer to that question is posted.

Sense of Community

We use Rovai's Classroom Community Scale [30] to measure subjective sense of community by users of each forum. In this self-report survey instrument, participants answer 20 questions on a five-point Likert scale about their perceptions of community, e.g. *"I feel that students in this course care about each other."* The results are aggregated into a numerical rating from 0 to 80. Rovai's Scale has been used, for example, to connect certain quantitative forum metrics to sense of community [10]. We assessed sense of community as part of the end-of-course survey.

	Full-featured forum	**Basic forum**
Total posts	819	587
Questions (threads)	115	144
Answers	358	286
Comments	346	157
Total users	546	555
Posters or upvoters	152	—
Post authors	119	152
Question authors	70	88
Answer authors	81	108
Comment authors	59	58

Table 2. Basic descriptive statistics for each forum. Deleted content is included.

RESULTS

The two experimental forums had roughly equal number of students (562 vs. 578), with more posts on the full-featured forum (819 vs. 587). Additional overall statistics are shown in Table 2. Each thread has one question (original post) and zero or more answers (responses to questions giving solutions) and comments (other posts).

Fisher's exact test shows that users on the full-featured forum are less likely to post ($p \approx 0.036$)—it has more posts overall but less users posting questions and answers. The basic forum also has more questions, and a post on the basic forum is more likely to be a question than a post on the full-featured forum ($p \approx 10^{-6}$). Although fewer users answered questions on the full-featured forum, it also has less users posting overall—as a result we found no evidence that posters on either forum are more likely to answer questions (Fisher's exact test, $p > 0.5$).

The top 10 posters in each forum contribute a substantial portion of overall posts: 54% and 32% in the full-featured and basic forums, respectively. The behavior of these authors can have substantial influence on the forum as a whole, particularly in the full-featured forum where the top two posters are responsible for 23% and 9% of all posts. Posts not by the top 10 posters (46% and 68% respectively) are contributed by a long tail of authors; see Figure 3. Because of the holistic nature of the forum, where all posts are available to all users and affect behavior of others, there is no simple way to account for or exclude top posters during analysis. For example, examining the average number of answers per question after removing top posters would produce a misleading underestimate, because the software is designed to discourage answering questions that already have a good answer.

Users applied an average of two tags to each question (at least one is required), with 43% of tag uses related to homeworks or quizzes, 31% related to the course topic (software engineering terms), and the remainder largely related to course administrative issues.

In analyzing our results, since the distributions being compared usually cannot be shown to be normal, we rely on the

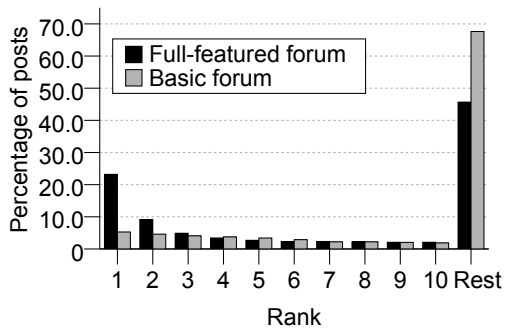

Figure 3. Number of posts contributed by top posters. The top two posters in the full-featured forum stand out with 23% and 9% of all posts. Both distributions are long-tailed.

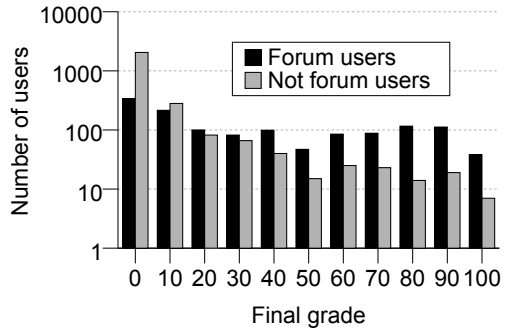

Figure 4. Comparison between final grades of users who visited forum at least once and users who did not. Note that vertical axis is logarithmic. Grades for students who visited the forum are significantly higher.

nonparametric Mann-Whitney U rank-sum test to determine whether one tends to have larger values than the other.

Forum Use Correlated With Higher Grades and Retention

Cheng et al. [5] have shown that students who voluntarily participate in forums perform better and receive higher exam grades. In our setting, we define a *forum user* as a user who visited the experimental forum or the edX forum (1572 users), and a *non-forum user* as a user who is not a forum user but viewed at least one lecture, homework, or quiz (3221 users); this excludes the 1192 users who enrolled in the course but never participated. With these definitions, students who participate in the forum, even by just visiting it once, have considerably higher grades than those who don't: 78% of those who don't visit the forum receive zeros as final grades, as compared to only 26% of those who visit the forum, and the median score is 22% as opposed to 0%, as shown in Figure 4. This difference is significant (Mann-Whitney U = 701305.5, n_1 = 1321, n_2 = 2622, $p < 0.001$), and hypothesis H1 is supported. The difference remains significant even if the non-forum users are restricted to the 1321 highest grades in the set (Mann-Whitney U = 480786.0, $p < 0.001$).

Moreover, students who visit the forum participate in the course longer before dropping out, lasting for a median of 39 out of 42 days as opposed to 19 of 42 days for students who don't visit the forum, as shown in Figure 5. This difference

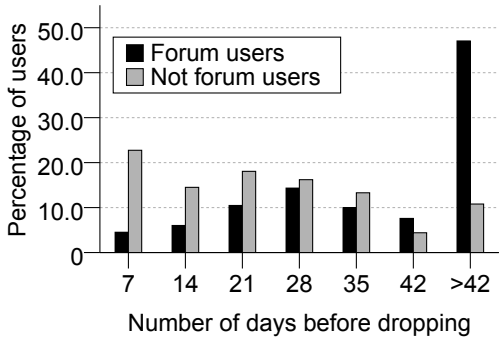

Figure 5. Percentage of forum users and non-forum users, respectively, who drop each week of the course, with the last bar showing those who are still participating at the end of the course after six weeks. Forum users are defined as users who visit the forum at least once. Forum users participate significantly longer. In particular, over 20% of non-forum users quit the first week, and nearly 50% of forum users participate until the end.

Assessment	# forum users, median		# non-forum users, median		U	Significance
HW1	737	700	216	610	60099	$p < 0.001$
HW2	471	300	85	300	17308	$p < 0.02$
HW3	350	500	56	500	9110	$p > 0.1$
HW4	259	500	47	500	5658	$p > 0.2$
Quiz 1	703	8	290	7	84436	$p < 0.001$
Quiz 2	476	11	164	10	31025	$p < 0.001$
Quiz 3	401	8	290	8	19839	$p < 0.008$

Table 4. Comparison of forum users and non-forum users on particular assessments, examining only students who completed those assessments. Mann-Whitney U tests show that forum users receive better assessment scores on most assessments, but are inconclusive on the last two homeworks. More forum users than non-forum users attempted every assessment.

is significant (Mann-Whitney U = 5609, n_1 = 115, n_2 = 144, $p < 0.001$), supporting H2. Note that because forum users were self-selected, both these results are solely correlational.

Forum Use Correlated With Higher Assessment Scores

As noted above, retention is poor for non-forum users, and this severely skews their final grade distribution, with 78% receiving zeroes. To separate the effects of retention from student understanding of the material, we also investigate whether non-forum users also received lower scores on assessments that they completed.

By examining particular assessments and the students who completed them, we found that forum participation was associated with significantly better scores on all assessments except for the last two homeworks, where no difference could be shown. For example, the median score of both forum and non-forum users on Homework 3 was 500, the maximum, and no significant difference was found (Mann-Whitney U=9109.5, n_1 = 350, n_2 = 56, $p > 0.1$). Table 4 summarizes results for all assessments. Again, because forum users were self-selected, this result is correlational.

Reputation Features Improve Forum Response Metrics

Passive participation in the forum is similar in the two groups: most students view only a few discussion threads during the

	Hypothesis and Result	Data	Test	Significance
H1	**Forum Use → Higher Grades.** Supported.	88% of students who don't visit forum receive zeros as final grades, compared with 26% of those who visit at least once.	Mann-Whitney U = 701305.5, n_1 = 1321, n_2 = 2622	$p < 0.001$
H2	**Forum Use → Higher Retention.** Supported.	Students who visit the forum participate in the course for a median of 39 out of 42 days, versus 19 out of 42 for those who don't visit.	Mann-Whitney U = 1722209 n_1 = 1528, n_2 = 4634	$p < 0.001$
H3	**Reputation Features → Number of Responses.** Supported.	Questions in the full-featured forum received a mean of 3.5 responses, versus 2.3 in the basic forums.	Mann-Whitney U = 5609 n_1 = 115, n_2 = 144	$p < 0.001$
	Reputation Features → Time to First Response. Supported.	Median time to first response for an answered question in the full-featured forum was 59 min, versus 2 hr 21 min for the basic forum.	Mann-Whitney U = 4626 n_1 = 110, n_2 = 120	$p < 0.001$
H4	**Reputation Features → Higher Grades.** Not Supported.	Median final grades in the full-featured forum and basic forum respectively were 73 and 71.	Mann-Whitney U = 4235.5 n_1 = 89, n_2 = 97	$p > 0.4$
H5	**Reputation Features → Higher Retention.** Not Supported.	The median forum retention in the full-featured and basic forums respectively was 24.5 and 26.1 days.	Mann-Whitney U = 147315 n_1 = 546, n_2 = 555	$p > 0.2$
H6	**Reputation Features → Sense of Community.** Not Supported.	The full-featured and basic forums had mean community scores of 48.5 and 49.8 out of 80, respectively.	$t(81) = -0.35$	$p > 0.7$
H7	**Reputation Features → Less Appeals for Help.** Supported.	The full-featured forum had 24 of 819 posts with appeals for help; the basic forum contained 43 of 587 such posts.	Fisher's exact test	$p \approx 0.0002$

Table 3. Summary of the quantitative support for each hypothesis.

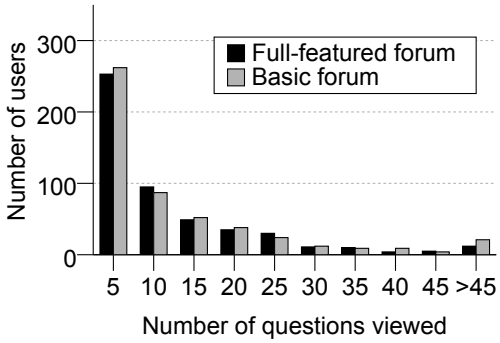

Figure 6. Histogram of number of thread views per user in each forum, showing similar distributions for both forums. Neither forum's users were found to view significantly more threads.

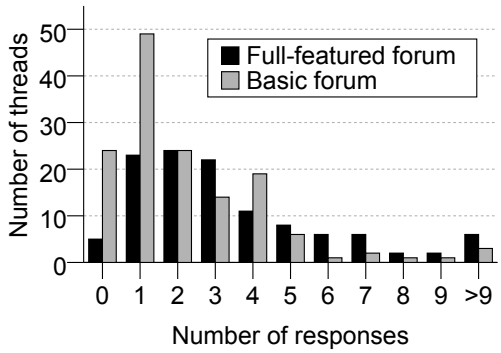

Figure 7. Number of responses questions received in each forum. Questions in the full-featured forum received significantly more responses. In particular, unanswered questions were much more common in the basic forum (21 or 15% as opposed to just 1 in the full-featured forum).

course (over 70% viewed 10 or fewer threads), and we could show no significant difference in the number of threads users of each forum viewed (Mann-Whitney U = 150996, n_1 = 546, n_2 = 555, $p > 0.4$). The histogram in Figure 6 shows strikingly similar distributions.

Active participation shows clear differences across forums. Questions in the full-featured forum received more responses and received their first response more quickly. Questions received a mean of 3.5 responses as opposed to 2.4 in the basic forum, and this difference was significant. To measure the time to first response we exclude questions that received no response, which is conservative since these were rare in the full-featured forum. Median time to first response was 59 minutes as opposed to 2 hours and 20 minutes in the basic forum, and this was also significant (Mann-Whitney U = 4626, n_1 = 110, n_2 = 120, $p < 0.001$), supporting H3. See Figures 7 and 8.

No Significant Effect of Reputation System on Student Outcomes

The median final grade in the full-featured forum and basic forum respectively were 73 and 71 (see Figure 9). No significant difference was found (Mann-Whitney U = 4235.5, n_1 = 89, n_2 = 97, $p > 0.4$), and H4 was not supported. Since the grade of a student incorporates both their progress in the course and assessment outcomes, it falls into a bimodal distribution where both forums include students who quit the course with zero grades and students who completed the course with more typical grades.

The comparison above used only the subset of users who opted in to have their forum activity correlated with their course outcomes. This subset of users had very high retention (more than 85% in both forums participated until the end of the course), providing insufficient data to confirm or reject H5. A more meaningful comparison can be made without

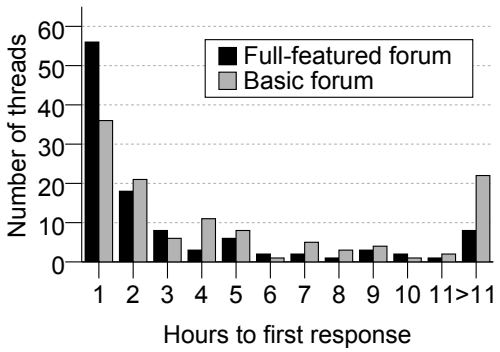

Figure 8. Number of hours for a question to receive its first response in each forum, among questions receiving at least one response. The full-featured forum has a significantly faster time to first response, as evidenced by the much larger number and proportion of questions answered in under one hour (49 or 50% of all answered questions as opposed to 35 or 30%), as well as having less than half as many posts requiring 12 or more hours.

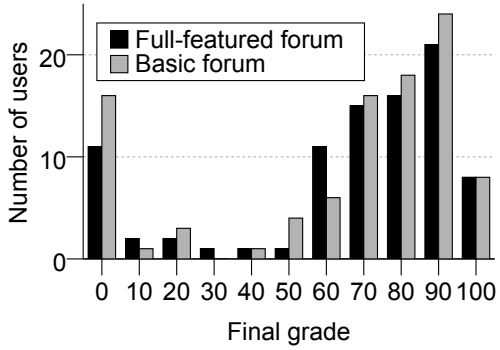

Figure 9. Histogram of final course grades for users of each forum. Neither group was found to have higher scores.

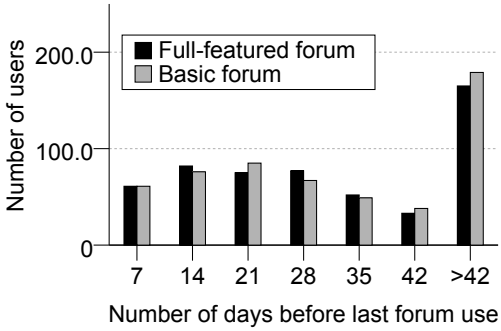

Figure 10. Number of forum users who cease to visit the forum each week, with the last bar showing those who are still participating at the end of the course. No significant difference was shown.

correlating forum activity and course outcomes if we instead compare only *forum retention* of all users on the two forums; that is, the number of days from the beginning of the course to their last forum access, shown in Figure 10. The median forum retention was 24.5 and 26.1 days in the full-featured and basic forums respectively, and no significant difference was shown (Mann-Whitney U = 147315, n_1 = 546, n_2 = 555, $p > 0.2$), providing no evidence for H5.

No Significant Effect of Reputation System on Sense of Community

Participants' self-report score to Rovai's Classroom Community Scale [30] follow a normal distribution (SciPy omnibus normaltest, [31] k_2 = 0.25, 2.8, $p > 0.2$), so a parametric test can be used. The Rovai Scale questions were included with the qualitative survey, and the full-featured and basic forum groups had 43 and 40 responses respectively. Mean scores were similar: the full-featured and basic forums had mean scores of 48.5 and 49.8 out of 80, respectively, suggesting a moderate sense of community. See Figure 11. No significant difference between forums was found ($t(81) = -0.35$, $p > 0.7$), and H6 is not supported.

Reputation Features Decrease Appeals for Help

We used the exploratory text analysis tool WordSeer [26] to explore a number of plausible hypotheses regarding differences in language between the forums. Among the examined phrases were expressions of community and discussions of reputation features, but these had insufficient data to analyze.

We isolated one feature suitable for analysis: appeals for help. Our hypothesis is that, in the absence of a reputation system, users will rely on linguistic cues such as "any help is appreciated" to persuade others to answer them. The following words, with stemming, were used to search for sentences in the questions containing appeals for help: "appreciate," "help," "anyone," "anybody," "somebody," "someone," "thank." Posts containing these words were identified and then the posts were rated blindly (without knowledge of the forum it came from) by three independent raters as either representing or not representing an appeal for help, and the final classification was determined by majority. Fleiss' kappa gives a strong inter-rater reliability of $\kappa = 0.80$. An example sentence that all raters agreed was an appeal for help: "Any and all help is greatly appreciated." An example that all raters agreed was not: "It helps me in the homework 1."

In the full-featured forum, 24 of 819 posts (2.9%) by 19 of 546 unique users (3.5%) contained language appealing for help. The basic forum contained 43 of 587 such posts (7.3%) by 31 of 555 unique users (5.6%). This is a significant difference (Fisher's exact test, $p \approx 0.0002$), and H7 is supported. Note however that there is some subjectivity in the rating process, and we cannot necessarily infer that users intended to persuade others to help them; for example, these comments might have been expressions of frustration.

Qualitative Results

Students using the full-featured and basic forums provided 43 and 45 responses, respectively, to the optional end-of-course survey, a response rate of 8%. All questions were optional. 83% of students gave their pseudonym; this lack of anonymity may introduce bias.

No Significant Difference in Usefulness

On a scale of 0 to 4, the median rating of usefulness of both forums was 2.0, and we found no significant difference (Mann-Whitney U = 629.0, n_1 = 37, n_2 = 39, $p > 0.1$). Similarly, when asked how often they find the answers they were

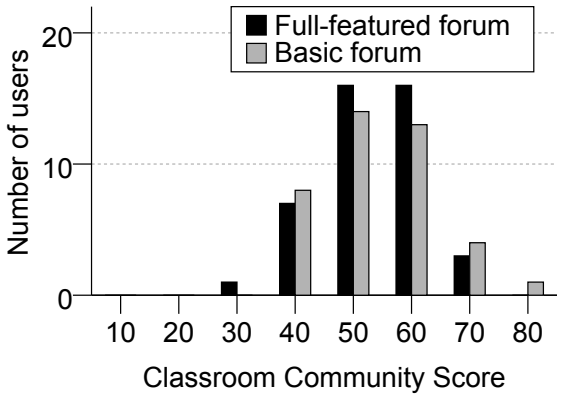

Figure 11. Histogram of Rovai's Classroom Community Scale scores for users of each forum. Neither group was found to have higher scores.

looking for on a 0-4 scale, median ratings were 2.5 and 3.0, and again there was no significant difference (Mann-Whitney U = 665.5, $n_1 = 36$, $n_2 = 39$, $p > 0.3$).

Reputation Features Moderately Useful

Full-featured forum users were asked some questions about their experience with the reputation system. When asked how often features were useful, both upvotes and the "select best answer" feature were given a median rating of 2.0 on a 0 to 3 scale, with 60% of students rating them "always" or "usually" useful. 62% of students said "Yes" or "Usually" when asked whether the reputation system was "fair."

Motivating Students to Care About Reputation Score

One limitation of our study is that students have limited incentive to increase their reputation score when it has no impact beyond the forum of a single short course. 40% of students said they would be more motivated to increase their reputation score if it carried over into other edX courses, and 23% said they would be more motivated if it carried over into future iterations of the same course. This suggests that the use of a sitewide reputation system that preserves scores over time could be more effective. 23% of students also said they would be more motivated to increase their score if it affected their final grade in the course.

Motivated By Altruism, Not Self-Interest

As shown in Figure 12, students self-reported that their main reason for responding to forum posts was in order to be helpful to others, rather than to increase their reputation score or benefit their own learning. Similarly, 89% of students reported it was "unimportant" or only "a little important" to increase their reputation score. On the other hand, [34] found that users of the Math Overflow forum tended to understate their desire to increase their reputation score on surveys, when compared with their behavior. Particularly since surveys were not fully anonymous, students may have avoided answers that make them appear selfish. Some motivators may be underrepresented due to being excluded from the survey, for example having the answer with the most upvotes or getting an answer selected as the best answer.

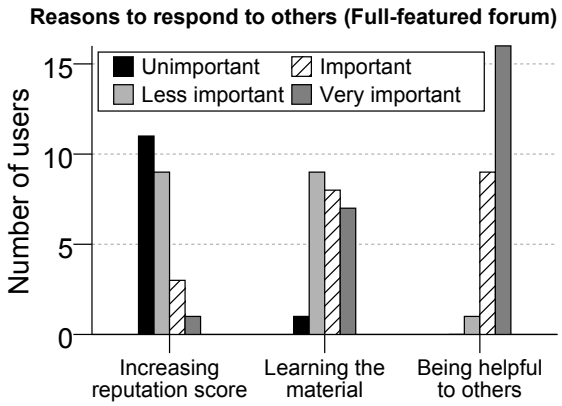

Figure 12. Though presence of reputation features did impact behavior, students self-reported that increasing reputation was not a major goal.

DISCUSSION

We review our results in context, list important limitations, and give advice for running field experiments on MOOCs.

Interpretation of Results

Reputation systems provide concrete, yet limited benefits for MOOC forums. They can result in faster response times and larger numbers of responses per post, improving the students' objective forum experience; and surveys suggest that students found the reputation features useful, improving their subjective forum experience. But we could not show a significant difference in the students' final grades or their subjective sense of community, suggesting that reputation systems do not contribute substantially to assessment outcomes or student community formation.

Users are more likely to post on the basic forum, but posts are more likely to be questions, and a larger proportion of questions are not answered, so these posts might not be productive. The number of users who posted or upvoted on the full-featured forum is also the same as the number who posted on the basic forum (see Table 2), suggesting that upvotes may be displacing some types of posts on the full-featured forum.

The strong correlation between forum usage and student outcomes like retention and grades is difficult to interpret due to a range of possible confounding factors, which we explore below, but to the extent that a causal link exists, it may be attributed to the use of the forum to get "unstuck" when a student is unable to make progress in the course on their own. In learning theory terms, the student is operating in their zone of proximal development [40], which is considered important for advancing individual learning. In the absence of a forum, students frustrated by an obstacle are, intuitively, more likely to cease participation. It is therefore unsurprising that forum software is widely deployed in MOOCs today.

The finding that forum users receive better scores on most individual assessments but not the last two homeworks can be explained in several ways. Steadily increasing p-values

from one homework to the next suggest that as students depart, sample size shrinks and statistical tests can no longer detect differences, but they still exist. Additionally, assessments that are too easy can fail to distinguish two groups even if their competency is very different; perfect median scores like we observed suggest low difficulty. The overall disadvantage of non-forum users on individual assessments suggests that poor retention is not the sole cause of their lower grades.

Limitations and Biases

Our study has a number of important issues that limit validity and generalizability.

Single Course Offering

We studied a single offering of a single course. It is not clear to what extent our results would generalize to other settings; e.g., student demographics may be different in other course subjects. Because software engineers are tech savvy, they may be more familiar with discussion forums than other types of students (18% of participants use them "frequently" and 21% "on a daily basis"). In addition, class size may lead to different forum behavior. The most popular MOOCs have an order of magnitude more students than the class we studied.

Small Number of Very Active Users

As in many other systems, participation resembles a power law distribution: the top contributors author a substantial proportion of posts. In the full-featured forum, 43% of posts were by the top 5 users, and in the basic forum 21% of posts were by the top 5 users. The behavior of these particular users may have had a strong effect on metrics like average number of responses, and it is difficult to determine to what extent their behavior is based on the presence or absence of the reputation system. There is no simple way to analyze the forum without the effects of top contributors because the forum is a holistic system in which every post affects all users.

Self-Selection Bias

The study required subjects to consent to participation, and only 67% of forum visitors did so. Students who opted out were directed to the traditional edX forum. This self-selection may have resulted in a biased sample, for example because students who opted out preferred the traditional edX forum experience, or because they had experience with reputation systems and disliked them. Of these participants, only 20% also consented to have their grades correlated with forum activity, and only 8% responded to the survey at the end of the course. Because these interventions were conducted late in the course, they are more likely to have excluded students who had already abandoned it. In particular, this suggests that sense of community scores may be biased upwards.

No Claims of Causation

Although participation in forums was correlated with increased retention and higher grades, there is no evidence of a causal relationship that forum participation leads to improved outcomes. Several alternative explanations are possible. First, students who are more seriously committed to completing the course and the assignments may be more likely to encounter difficulties which require forum support to resolve. Second, students who drop out very early in the course, and so exhibit very low retention and grades, may not ever have encountered a situation in which the forum would have been useful. Finally, students who are unable to use the forum due to limited English-language proficiency or technical literacy may also struggle with the rest of the course.

External Validity and Contamination

Although our experimental forum was used in a real class setting, it does not provide the fully-integrated forum experience that students would normally receive. When they first join the forum, users must visit another website at another domain, log in again using their edX credentials, and complete a consent form before accessing it. These extra steps may deter some users. Additionally, students may have interactions outside of the experimental forum, such as in the edX forum or in private study groups on Facebook or Skype. This could result in the spread of information between users of the two experimental forums, affecting both course outcomes and subsequent forum activity.

Lessons for MOOC studies

MOOCs are attractive for controlled experiments because of their large number of participants and the ability to deploy designs rapidly. However, such experiments face unique challenges that are not encountered in traditional classroom studies. In this section we outline some advice for MOOC researchers and for operators of MOOC platforms.

Advice for MOOC researchers

1. **Study retention and performance separately**: In a MOOC, high attrition is the norm. Metrics like grades, although still valuable, conflate student progress in the course and understanding of material. To separate these factors, it is useful to examine retention directly, to compare performance on individual assignments, and to break subjects into cohorts for analysis based on retention.
2. **Use clickstreams to infer retention**: By logging every page load, clickstreams simplify computing student retention, since a large variety of interactions could potentially indicate continued interest in a course.
3. **Don't rely on late surveys**: Because of attrition, late surveys tend to capture a small, biased set of "survivors" who participated until the end. Early surveys can capture information about other students and their intent for taking the course, and also function as a pretest, enabling measurement of individual improvement over the course.
4. **Use established platforms, but work around limitations**: Reaching large subject pools means leveraging existing platforms with large market share and limited extensibility. In our study, we inserted external links that directed students to another site hosted by the experimenters, sacrificing an integrated experience for more control. We authenticated with edX's OpenID interface and reused edX usernames, simplifying signup and data integration.

Advice for MOOC platforms

To enable continual innovation, MOOC platforms need to empower individual instructors and researchers to deploy new designs and technologies in classes with as few barriers as possible. Our field study ran up against several limitations that a more research-friendly platform could have avoided:

1. **Enable extensibility**: Instructors and researchers should have the opportunity to build and deploy sophisticated extensions and client-side scripts on the MOOC platform, without resorting to the use of an external site. These should be able to transparently replace or extend existing functionality in courses, and be easy to share with others.
2. **Enable interoperability**: When new technology can't be integrated directly with the platform, it still has to interact with it. Support federated identity, and make it easy for external sites to request information about a user or take action on their behalf (with their permission). All functionality available on the web should also be exposed through a service API.
3. **Make data easy to access**: To the extent it is possible, MOOC platforms should secure explicit and broad prior permission from all students to disseminate their data for research purposes, and should make less sensitive data freely available to the general public. When researchers need to go beyond this blanket permission, the platform should make it easy to direct students to a consent form before they can participate in a course.

CONCLUSION AND FUTURE WORK

We presented results from a field experiment on the impact of forums and forum reputation systems on MOOC students. Our work focused on a single offering of a single course, and an important question is the extent to which our results generalize to other courses. Replicating the results would provide evidence that they are not due to the influence of individual prolific users or the topic being studied. Another open question is whether reputation systems would be useful in student forums that center around long, complex discussions of course concepts, rather than the simple question-answer format explored here. Noticeable differences between our forum and success stories like Stack Overflow include the community size and duration of forum use, suggesting that a forum used with a larger course or set of courses over a longer period of time may be more successful.

Our analysis focused primarily on significant relationships between factors; an interesting question is whether metrics of participation such as number of posts and number of threads viewed can be related by regression to course outcomes. Another important question is how to develop analyses of holistic systems like forums that are robust against undue influence from the most prolific users.

The forum used in this work represents a single forum design, and the possibility remains that other forum designs, including novel designs, may produce different results. In particular, due to the short timespan of this experiment, we used OSQA's "bootstrap mode" which allows users to earn privileges with less points. Both bootstrap mode and standard mode in OSQA have different point rewards from other sites like Stack Overflow, and another choice of point rewards may be more effective in encouraging participation. A more ambitious direction that may better improve course outcomes like grades is to create a new forum design based on sound pedagogical principles such as conceptual change, metacognition, and project-based learning.

Beyond forums, there is a wide array of MOOC interventions that could be studied using a similar setup to ours, ranging from simple web design changes, to changes in course materials and assessments, to novel interaction technologies. The strong correlation between forum usage and student outcomes suggests that making an overall improvement in student outcomes requires new interventions that will effectively target the 66% of students who never set foot in the forum. We hope that our work can provide a blueprint for future field experiments in live MOOCs.

ACKNOWLEDGMENTS

We thank Sam Joseph and the World TAs of CS169.1x for coordinating support of students on the forums. This work was funded by the National Science Foundation under award IIS-1149799 and partially funded by the National Endowment for the Humanities under grant HK-50011.

REFERENCES

1. Adamic, L. A., Zhang, J., Bakshy, E., and Ackerman, M. S. Knowledge sharing and Yahoo Answers: everyone knows something. In *Proceedings of WWW 2008*, ACM (New York, NY, USA, 2008), 665–674.
2. Bartlett, M. Y., and DeSteno, D. Gratitude and prosocial behavior helping when it costs you. *Psychological science 17*, 4 (2006), 319–325.
3. Belanger, Y., and Thornton, J. Bioelectricity: A Quantitative Approach — Duke Universitys First MOOC. *Duke University Report* (2013).
4. Chamberlin, L., and Parish, T. MOOCs: Massive open online courses or massive and often obtuse courses? *eLearn 2011*, 8 (Aug. 2011).
5. Cheng, C. K., Par, D. E., Collimore, L.-M., and Joordens, S. Assessing the effectiveness of a voluntary online discussion forum on improving students course performance. *Computers & Education 56*, 1 (2011), 253 – 261.
6. Cheng, R., and Vassileva, J. Adaptive reward mechanism for sustainable online learning community. *Artificial Intelligence in Education* (2005), 152–159.
7. Chizmar, J. F., and Williams, D. B. Altering time and space through network technologies to enhance learning. *CAUSE EFFECT 19* (1996), 14–21.
8. Clow, D. MOOCs and the funnel of participation. In *LAK '13: 3rd International Conference on Learning Analytics & Knowledge* (2013).
9. Cooper, S., and Sahami, M. Reflections on Stanford's MOOCs. *Commun. ACM 56*, 2 (Feb. 2013), 28–30.
10. Dawson, S. P. Online forum discussion interactions as an indicator of student community. *Australasian Journal of Educational Technology 22*, 4 (2006), 495–510.
11. Dearman, D., and Truong, K. N. Why users of Yahoo! Answers do not answer questions. In *Proceedings of CHI 2010*, ACM (2010), 329–332.

12. Deci, E. L., Koestner, R., and Ryan, R. M. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin 125*, 6 (1999), 627.

13. Google. Power searching with Google: second course FCCR. `https://plus.google.com/+ResearchatGoogle/posts/cSTrtKHyQh6`, Oct. 2012. [Online; accessed 28-May-2013].

14. Grant, A. M., Gino, F., et al. A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. *Journal of personality and social psychology 98*, 6 (2010), 946–955.

15. Guzdial, M., and Turns, J. Effective discussion through a computer-mediated anchored forum. *Journal of the Learning Sciences 9*, 4 (2000), 437–469.

16. Harper, F. M. *The impact of social design on user contributions to online communities*. PhD thesis, University of Minnesota, 2009. AAI3358616.

17. Harper, F. M., Frankowski, D., Drenner, S., Ren, Y., Kiesler, S., Terveen, L., Kraut, R., and Riedl, J. Talk amongst yourselves: inviting users to participate in online conversations. In *IUI 2007*, ACM (2007), 62–71.

18. Harper, F. M., Moy, D., and Konstan, J. A. Facts or friends? Distinguishing informational and conversational questions in social Q&A sites. In *CHI 2009*, ACM (2009), 759–768.

19. Hsieh, G., and Counts, S. mimir: A market-based real-time question and answer service. In *Proceedings of CHI 2009*, ACM (2009), 769–778.

20. Luca, J., and Mcloughlin, C. Using online forums to support a community of learning. In *ED-MEDIA 2004*, AACE (2004), 1468–1474.

21. Ludford, P. J., Cosley, D., Frankowski, D., and Terveen, L. Think different: increasing online community participation using uniqueness and group dissimilarity. In *CHI 2004*, ACM (2004), 631–638.

22. Mackness, J., Mak, S., and Williams, R. The ideals and reality of participating in a MOOC. In *International Conference on Networked Learning 2010*. 266–275.

23. Mak, S., Williams, R., and Mackness, J. Blogs and forums as communication and learning tools in a MOOC. In *International Conference on Networked Learning 2010*. 275–285.

24. Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., and Hartmann, B. Design lessons from the fastest Q&A site in the west. In *Proceedings of CHI 2011*, ACM (2011), 2857–2866.

25. Martin, F. G. Will massive open online courses change how we teach? *Commun. ACM 55*, 8 (Aug. 2012), 26–28.

26. Muralidharan, A., and Hearst, M. A. Supporting exploratory text analysis in literature study. *Literary and Linguistic Computing* (2012).

27. OSQA: The open source Q&A system. `http://www.osqa.net/`, 2013. [Online; accessed 28-May-2013].

28. Palinscar, A. S., and Brown, A. L. Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and instruction 1*, 2 (1984), 117–175.

29. Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. Reputation systems. *Commun. ACM 43*, 12 (Dec. 2000), 45–48.

30. Rovai, A. Development of an instrument to measure classroom community. *The Internet and Higher Education 5*, 3 (FebMar 2002), 197–211.

31. SciPy v0.12 Reference Guide: scipy.stats.normaltest. `http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html`, May 2013. [Online; accessed 28-May-2013].

32. Sterbini, A., and Temperini, M. Socialx: Reputation based support to social collaborative learning through exercise sharing and project teamwork. *Int. J. Inf. Syst. Soc. Chang. 2*, 1 (Jan. 2011), 64–79.

33. Szafir, D., and Mutlu, B. ARTFul: adaptive review technology for flipped learning. In *Proceedings of CHI 2013*, ACM (2013), 1001–1010.

34. Tausczik, Y. R., and Pennebaker, J. W. Participation in an online mathematics community: differentiating motivations to add. In *CSCW 2012* (2012), 207–216.

35. Teevan, J., Morris, M. R., and Panovich, K. Factors affecting response quantity, quality, and speed for questions asked via social network status messages. In *ICWSM 2011* (2011).

36. Thomas, D., and Brown, J. *A New Culture of Learning: Cultivating the Imagination for a World of Constant Change*. CreateSpace, 2011.

37. Thompson, C. How Khan Academy is changing the rules of education. *Wired Magazine 126* (2011).

38. Tillmann, N., De Halleux, J., Xie, T., Gulwani, S., and Bishop, J. Teaching and learning programming and software engineering via interactive gaming. In *ICSE 2013* (2013), 1117–1126.

39. Tinto, V. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. University of Chicago Press, 1993.

40. Vygotsky, L. L. S., et al. *Mind in society: The development of higher psychological processes*. Harvard University Press, 1978.

41. Yang, X., Li, Y., Tan, C.-H., and Teo, H.-H. Students' participation intention in an online discussion forum: Why is computer-mediated interaction attractive? *Information and Management 44*, 5 (July 2007), 456–466.