Using anonymity and rounds-based structure for effective online discussions in STEM courses

Mr. Swaroop Joshi, The Ohio State University

Swaroop Joshi is a Senior Lecturer of Computer Science and a Ph.D. candidate in Computer Science and Engineering at Ohio State University. He is interested in a range of topics in Education Technology and Software Engineering, including but not limited to: Computer Supported Collaborative Learning, Game Based Learning, Programming Languages, Compiler Construction and Optimization.

Dr. Neelam Soundarajan, Ohio State University

Dr. Neelam Soundarajan is an Associate Professor in the Computer Science and Engineering Department at Ohio State University. His interests include software engineering as well as innovative approaches to engineering education.

Using Anonymity and Rounds-based Structure for Effective Online Discussions in STEM Courses

1 Introduction and Background

Over the last several years, there has been an increasing emphasis on teamwork and *collaborative learning* activities across all engineering disciplines. These activities are intended to better prepare students in two important ways. First, they help students develop important team skills which will be essential in nearly every large industry project that the students may be involved with in their future careers. Capstone team projects which have become a standard part of (nearly) every engineering and computing program have been especially successful in helping to achieve this goal. The second intended goal of such activities is to help students learn the technical, conceptual material by engaging in suitable activities with their fellow-students rather than just listening passively to lectures. At the same time, many engineering and computing faculty have serious concerns about introducing such activities to any serious extent in their courses; primary among these concerns is the potential negative impact of such activities on topic coverage. Trying to arrange such activities *outside* the regular classroom poses numerous problems, especially given the work and school schedules as well as family responsibilities that many students have to juggle.

A natural answer, especially for millennials who are constantly online, would seem to be to organize such collaborative learning tasks online. And, indeed, there have been numerous attempts at doing just this. But the results have been quite disappointing. Thus Cole's[1] course on *information technology* with 75 students in it was organized so that lectures were in alternate weeks, the other weeks being intended for students to discover new material and post to the class wiki. Fully one quarter of the questions on the final exam were to be from the material that students posted. The expectation was that, given this, students would eagerly post content, edit each others posts, and engage in collaborative learning. Halfway through the course there had been *no* posts to the wiki! Rick and Guzdial[2] report that although they obtained positive results using wikis in architecture and English composition classes, the results in STEM classes were "overwhelmingly disappointing". For example, they report that fully 40% of math and engineering students settled for a zero on an assignment rather than engage in collaborative learning! Rick and Guzdial suggest that one of the reasons why STEM students did not engage in their collaborative learning activity was, they simply weren't sure what they were supposed to do when they were "asked to collaborate". We will consider these and other challenges that seem especially acute in the context of STEM later in the paper.

Returning to the notion of in-class activities, Mazur and others[3,4] have developed a technique known as *peer instruction* (PI) that they have used in college-level STEM courses. In this approach, first, each student answers a conceptual multiple choice question submitting the answer via a *clicker* or other similar device; then the students turn to their neighbors and, in groups of 3 or 4, discuss the question; after a few minutes of discussion, each student again answers the same question. During the discussion time, the instructor may walk around the room but deliberately does *not* participate. Mazur reports that the percentage of students who, following discussion with their peers, change their answer from a wrong choice to the correct one far exceeds the percentage who change from the correct choice to a wrong one. This is not surprising since the power of *socio-cognitive* conflict and its resolution by interaction among small groups of learners *without* involvement of an instructor, which is the idea underlying PI, has been well studied over the years in the context of children's learning. The notion of cognitive conflict was first formulated by Piaget [5], and later expanded upon by several others [6,7].

However, there are a number of limitations with PI and other similar approaches, mostly related to the fact that it is a classroom technique. First, there is no way to ensure that the students in a given group include ones who picked different possible answers because the grouping is based essentially on where students are seated. Second, some students tend to dominate their groups even if they don't necessarily have the right answers; this may be especially problematic for students from underrepresented groups, including women, whose opinions may be ignored in such groups. Third, the amount of time spent in the discussion is, naturally, limited; hence, students who are not quick to speak and take time to formulate precise and deliberate arguments may not contribute effectively to the discussions. Most importantly, there is also the concern, already mentioned, regarding potential negative impact of such activities on topic coverage.

Our work is based on the idea that it would be possible to develop an *online* approach and system that will allow small groups of students in computing and engineering courses to engage in focused discussions on specific conceptual topics with the group including students with different conceptions of the topic to help each student develop deep understanding that will overcome all of these problems and provide a number of additional benefits. We should stress that our interest is not in systems that will allow students to debate questions of broad social interest such as, say, immigration policies with the goal of helping students develop into responsible citizens. Rather, the motivation behind our work is to help students develop deep understanding about specific, important concepts related to computing, engineering, and other STEM subjects. As concept inventories in a number of these subjects have demonstrated, students harbor surprising misconceptions even though they demonstrate an ability to solve standard textbook problems [8,9,10].

While moving the discussion online would seem to be the obvious answer, there are still some important challenges. Consider a scenario where some students, possibly because of their work schedule, may not yet have studied the topic in question sufficiently well, while others have done so. If the discussion were to start at that point, the students who have studied the topic may charge ahead while the others haven't had a chance to even *log into* the system, let alone participate in the discussion in any depth [11]. As a result, the point of the discussion, that all students in the group will develop deep understanding, will be seriously compromised. Moreover, there is also the risk of *free-riding* [12]; in other words, some students may simply wait for others to work on the assigned problem and simply copy the answers arrived at on the basis of the discussion among the students who do participate. The effects of such a scenario are much more acute in the context of STEM courses since the problem that the group is supposed to discuss, as noted by Rick and Guzdial[2], is quite likely to have *one* correct answer or only one best or optimal approach, unlike questions in, say, social sciences, where there may be more than one possible answer, all equally reasonable. Further, there is also the possibility of one or two members of the group dominating the discussion by making long and repeated posts, thereby drowning out the posts of the other, possibly more thoughtful and insightful, students in the group. Hew and Cheung[13] review many of these difficulties.

In this paper, we report on our experience with using CONSIDER, an online system that we have developed to enable such discussions while addressing many of these problems; in particular, the results of two specific aspects of the system, i.e., *anonymity* of the discussions which ensures that students in a group are not aware of the identities of the other students in the group; and the *round-based* structure of the discussions which ensures that each student participates effectively in the discussion.

2 The CONSIDER Approach

We have developed a novel approach to online learning, employing the three key concepts we explored above: inducing cognitive conflict to form groups, anonymity to mitigate any prejudices that may hinder participation, and a rounds-based structure that overcomes several limitations of face-to-face and existing online discussion approaches. The approach, named CONSIDER, short for CONflicting Student Ideas to be Discussed, Evaluated, and Resolved, works as follows: the instructor posts an appropriate question or problem based on some key concept in the course with the students' individual (initial) answers to the question being used to form groups of students with conflicting or differing answers; the students in the group then engage in a structured, rounds-based discussion as explained below with the other students in the group, critically evaluating each others posts and trying to resolve the conflicts; following this discussion phase, each student again answers the question individually, and submits, along with her final answer, a summary of the discussion in her group. A key point is that the *grade* of the student depends *entirely* on this final submission, i.e., the correctness of her final individual answer to the question and the quality of her summary of the discussion in her group. It does *not* depend in any way on the correctness of her initial answer or on whether her final answer is the same as or different from her initial answer. This means that there is every reason for each student in the group to pay careful attention to the contributions of *every* student in the group since any of them, including herself, may be correct or wrong! And, of course, all of this is done outside of the classroom so that there is little or no impact on the classroom time; in fact, once the students have submitted their final individual answers and summaries, the instructor may decide to have a brief in-class discussion to address any interesting misconceptions that may have been common to multiple groups. Thus our online approach enables students to engage in effective collaborative learning, participating in the discussion at any time, from anywhere, and using any computing device (laptop, tablet, smartphone, etc.).

Phase 1 The first phase of the CONSIDER approach focuses on identifying the conflicting concepts students may have and forming groups based on that. To identify the conflicts, the instructor posts a question addressing a core concept on the online platform. This phase typically lasts 24 to 48 hours. Each student is required to answer the question on the web app, individually. The question may be either a short answer question or a multiple-choice question but typically includes both (see examples below); for the multiple-choice component, the student is required to not only pick one of the given options, but also to justify the choice. All the individual answers are recorded and available to the instructor, via the instructor interface of the web app, once the deadline for this phase expires. At this point, the instructor goes through all the answers, and based on the differences in the individual student's answers, creates *heterogeneous* groups of 4–5 students each, i.e., with each group having at least two students whose individual answers conflict with each other. Once the groups are formed, the CONSIDER system assigns aliases S1, S2, etc. to the students in each group. Note that the discussion in each group is entirely independent of the discussions in the other groups; students in a given group will see only the submissions of the students in that group and these submissions will be labeled by the aliases of the corresponding students. If there are no substantial differences in the explanations and if most students pick the right answer and give the right explanations, it may be an indication that the topic is well understood, and the instructor can simply move on to the next topic. On the other hand, if most students get the answers or the explanations wrong, it may be an indication that additional lectures or other resources might be required so that students understand the topic better. It is also possible that the question itself was not framed properly, and the instructor should look into all these possibilities if there is not enough conflict in the answers.

Phase 2 This is the discussion phase and is the piece that enables collaborative learning in the CONSIDER activity. This phase is organized into a series of *rounds*. For convenience, we will think of Phase 1, the initial submission phase, as Round_0, the actual discussion rounds being labeled, Round_1, Round_2, ... The number of rounds in the discussion as well as the duration of each round is determined by the instructor and entered into the system via the instructor interface when the assignment is originally posted; this information will be available to students when they first see the question. In our experience, a reasonable duration for each round is 24 hours; the number of rounds probably depends on the particular question but two to three rounds seems appropriate for most problems.

Once the instructor has formed the heterogeneous groups, when students next log into the web app, they will be in Round_1 of the discussion. They will see the submissions from Round_0, i.e., the *initial* answers submitted by each student in the group, each submission is labeled with the identifier (S1, S2, S3, S4) of the student whose submission it is. The student is now required to click, for each answer in his group, one of the three buttons: `Agree with all key points`, `Disagree with at least one key point` or `Did not understand some key points`. Then she is also required to explain, in a *textbox*, why she agrees or disagrees with the particular submission; if appropriate, the student may click "`Did not understand some key points` and ask for additional clarification. Explicating their (dis)agreement in this manner forces the students to read through each of the other student's posts and helps them identify disagreements they may have, which they can try to address in their explanation. An important aspect of the rounds-structure is that, posts made in the current round are *not* available to other students in the group *until* the deadline for the current round expires. In fact, a student may log in again before that deadline expires and make any changes to what he submitted previously, possibly because the student may have realized, after further thought, that what she said previously contained a mistake; it allows students to edit out any snarky comment they may have made in the heat of the moment. Thus, in each round, for each student, the system will retain only the last submission that the student makes before the deadline for the round expires; any earlier submissions the student may have made will be overwritten by later submissions as long as they are made before the deadline expires. This allows all students, whether they are quick to react or take their time to formulate their responses carefully, to participate equally. And the anonymity ensures that external factors such as gender or ethnicity etc. have little or no impact; nor can students simply adopt whatever answer is given by the "smartest" student in the group since, anonymity means that, for all a given student knows, he or she might be the one who has the right answer, all the others being wrong!

The 24 hour window serves another, practical purpose. Many college students, as noted earlier, have work schedules and family obligations they have to juggle along with their school work. Thus different students may be on very different schedules with some being able to work on class tasks early in the morning, others in the middle of the day and yet others in the dead of night. The CONSIDER discussion structure enables *all* these students to participate effectively. The discussion moves forward at a pace that works with *all* of their schedules.

When each new round starts, students see the *previous* round's posts from their group members listed at the top of the page and they have to indicate whether they agree/disagree with the latest, possibly modified, position of each of their group members. Again, a student must do this for *her own* previous round post as well. This is important because, it may so happen that she finds another student's last round post convincing enough to make her realize her mistake and change her position. That, improvement in an individual student's understanding during the process of resolving a cognitive conflict through peer discussions, is the heart of our approach.

Phase 3 Once the final discussion round ends, each student is required to submit his/her final answer to the question. This may be completely different from, or a refined version of, his/her initial answer. The complete log of the discussion in the group is accessible to the student and can be consulted when they craft their final answer. This provides another opportunity to the student to (re-)consider his/her answer and refine it in appropriate ways. The student is also required to submit a summary of their group's discussion. As noted earlier, the student's grade for the activity depends only on his/her final answer and the summary. The point is to encourage students to engage in a discussion where they are presented with conflicting ideas and to digest the various view points to finally come up with a carefully considered solution. The duration of this phase is also determined by the instructor and, in our experience, should be 24 to 48 hours; again, student may edit their submissions until the deadline expires.

We have developed a platform independent, scalable, responsive web app to implement this approach, using Google App Engine, Python and HTML. We used it in an undergraduate course on Principles of Programming Languages and compared the effectiveness of the discussion using our approach, with another discussion that was conducted on a popular tool Piazza, which uses forum-like discussion format. The experiment and results are described in the next section.

One important concern regarding the use of CONSIDER in STEM courses, as an anonymous reviewer noted, is that "in non-STEM the questions are usually asking for opinion, whereas, in STEM-related majors, the answers might be very clear. A simple search using Google can provide a simple answer for any STEM-related questions …". There are, in fact, two issues here. First, many questions that instructors in STEM courses might consider using have specific right/wrong answers; Rick and Guzdial[2] also express a similar concern when discussing why online collaborative learning doesn't seem to work well in STEM courses. Second, answers to many of these questions are easily available online. Based on our experience, both concerns can be largely mitigated by carefully chosen questions. Clearly, questions taken directly from standard textbooks or minor variations thereof are likely to be susceptible to both concerns and hence unsuitable for use with CONSIDER. However, if the question is appropriately tailored to be grounded in the context of, say, a project that are often key components of more advanced engineering and computing courses, then the question can become not only suitable for a CONSIDER discussion but also helps to effectively tie together the conceptual aspects of the course with the practical considerations that tend to dominate the project component. Thus, for example, in our junior/senior-level principles of programming languages course, the main project consists of having the students implement an *interpreter* for a simple programming language. A main goal of the project is to illustrate many of the concepts discussed in the course but most students tend to focus, in their actual projects, on getting their implementation to work correctly for various possible inputs rather on the conceptual questions. Once the students have completed (or nearly completed) the project, a CONSIDER discussion that explores the application of a new concept introduced in the course such as *type inferencing* to the interpreter project can be extremely effective in helping students see the practical implications of such advanced concepts. There is no risk that simple answers can be found online since the answers depend on the details of the particular project; in fact, for such discussions, there may well be more than one legitimate approach and the resulting discussion can contribute substantially to the students' understanding.

3 Method

3.1 Subjects

A total of 37 junior/senior level Computer Science and Engineering undergraduate students, recruited from a Principles of Programming Languages course in a large public university in the Midwestern United States, participated in this study. 24 of them answered a post-activity questionnaire which reflected, among other things, the demographic information. The respondents consisted of 83% CS majors and 17% non-majors. Three-fourths of the respondents were males. About 46% of them identified as Caucasians and an equal number were Asians, while 4% of the respondents were African-Americans and 8% Hispanics.

3.2 Procedures

The students of the course were given two assignments in the form of online-discussions on the two tools: (1) Piazza (`http://piazza.com`), a popular online-discussion forum used in thousands of courses across the world, including CSE courses at this university, and (2) CONSIDER, the web app we developed to implement our approach (described in Sec. 2). The two assignments were on the topics discussed in the class in the preceding lectures and were of comparable difficulty. The students were given a *practice session* so they could get used to the CONSIDER interface. They were already familiar with the Piazza interface, so it was not necessary to give a practice session for that.

For readers who are not Software Engineering experts, some brief comments on the background for the two activities might be useful. Scala is a relatively new programming language that is based on Java but incorporates a number of new ideas from other cutting-edge languages. All students in the course have a fairly strong background in Java (since it is the language used in several of the earlier courses) but (almost) none of them has much knowledge or experience of Scala.

Here's a 2-part problem:

1. Explain how exactly the order in which a class C inherits from various traits T1, T2, ..., influences the behavior of C, i.e., how this order influences what a given method applied to an instance of C will actually do when executed.
2. Consider the following statement: "A `super` call that appears in the body of a method defined in a Scala class is treated differently by the Scala compiler from the way a `super` call that appears in the body of a method defined in a trait is treated by the Scala compiler." Is the statement true or false? If true, explain why this difference; if false, briefly describe how the Scala compiler treats such `super` calls.

Figure 1: Traits Question for Piazza activity

The Piazza activity (Figure 1) concerns what are called *traits* in Scala which adds to the power of *inheritance* in the language. That is, traits enable the Scala programmer to use inheritance in ways that are much more effective than it can be used in a Java program. However, this comes with added complexity. The first, easier, part of the question relates to the added complexity and is based almost directly on the class discussion; so most students are expected to (and do) answer it correctly. The second part is more challenging and requires the student to think through the details of how inheritance works in the simpler Java setting and try to work out the complexities introduced by Scala's traits. The correct answer to the question in this part is that the given statement is indeed true but justifying it takes careful thinking.

The CONSIDER activity (Figure 2) concerns what is called *type inferencing* in which the compiler for the language automatically deduces some or all of the information about the *type* of a given variable in the program. By contrast, in Java, this information has to be explicitly specified by the programmer. Type inferencing is expected to make programming faster – once the programmer is comfortable with how it works. This activity tries to get students to think about what exactly happens during type inferencing in the Scala system. For comparison, the activity also asks about Ruby and Python which are two languages that are quite different from Java/Scala; Ruby and Python are somewhat similar to each other and most (or all) students are reasonably familiar with at least one of them. The goal of including this part of the activity is to get students to go beyond what was explicitly presented in the class discussion. The correct answers to the two questions in this activity are that, indeed, depending on the situation, a Scala programmer may omit some, all, or none of the type information. And, for the second part, that Ruby and Python don't use type

Type inferencing is a major part of Scala. This is a 2-part question.

1. Consider the following: "In some situations, the Scala programmer can omit *all* type information; in others, he/she has to provide *some* but not all; in yet others, complete type information (as you might in an equivalent Java program) has to be provided." True or false?
   - If true, provide examples of each of the three cases.
   - If false, explain what part of the statement is false and why.
2. Do languages such as Ruby (or Python) use type inferencing?
   - If yes, explain how they do it along with some simple examples.
   - If not, explain why they do not do so.

Figure 2: Type Inferencing Question for CONSIDER activity

inferencing in the same manner as Scala but there is something vaguely analogous that happens when the Ruby/Python programmer is executed (while type inferencing in Scala happens *before* the program is executed).

Both activities were conducted as graded homeworks as part of the regular course. Only those students' data who signed a consent form was considered for this research. Students' actual grade was computed based only on their final answer, although we (later) analyzed both the initial and final answers for the purpose of studying the effectiveness of our approach. In both cases, students were asked to submit their initial answer within 48 hours of posting the question online. Based on their initial answers, groups of students with conflicting ideas about the solution were formed. Type of discussion was the independent variable, which was manipulated by assigning one activity to one type of discussion and the other activity to the other discussion tool. While in the Piazza activity, they engaged in a *forum-based* discussion and critiqued each others answers on Piazza for the next 48 hours, the CONSIDER discussion phase was organized as two 24-hour rounds, where students engaged in a *rounds-based* discussion and posted their responses *anonymously* as described in Section 2. Figure 3 shows an example discussion in CONSIDER. The student whose alias is S2 *disagrees* with S1's initial post (indicated by the red background for that post) and provides explanation for why she disagrees with S1 in the text box at the bottom of the screenshot. In Phase-3 for both conditions, students were asked to submit their final answers to the same questions they discussed. For both conditions (Piazza and CONSIDER), the initial and final answers were evaluated on a 4-point scale, and the difference in the two scores was measured. After both activities were completed, an optional, anonymous, online questionnaire was given to all the students who had signed the consent form. Data from both the assignments and the questionnaire was de-identified before we began any analysis.

3.3 Instruments

3.3.1 Improvement in Learning

The individual submissions were evaluated on a 4-point scale using the following general rubric:

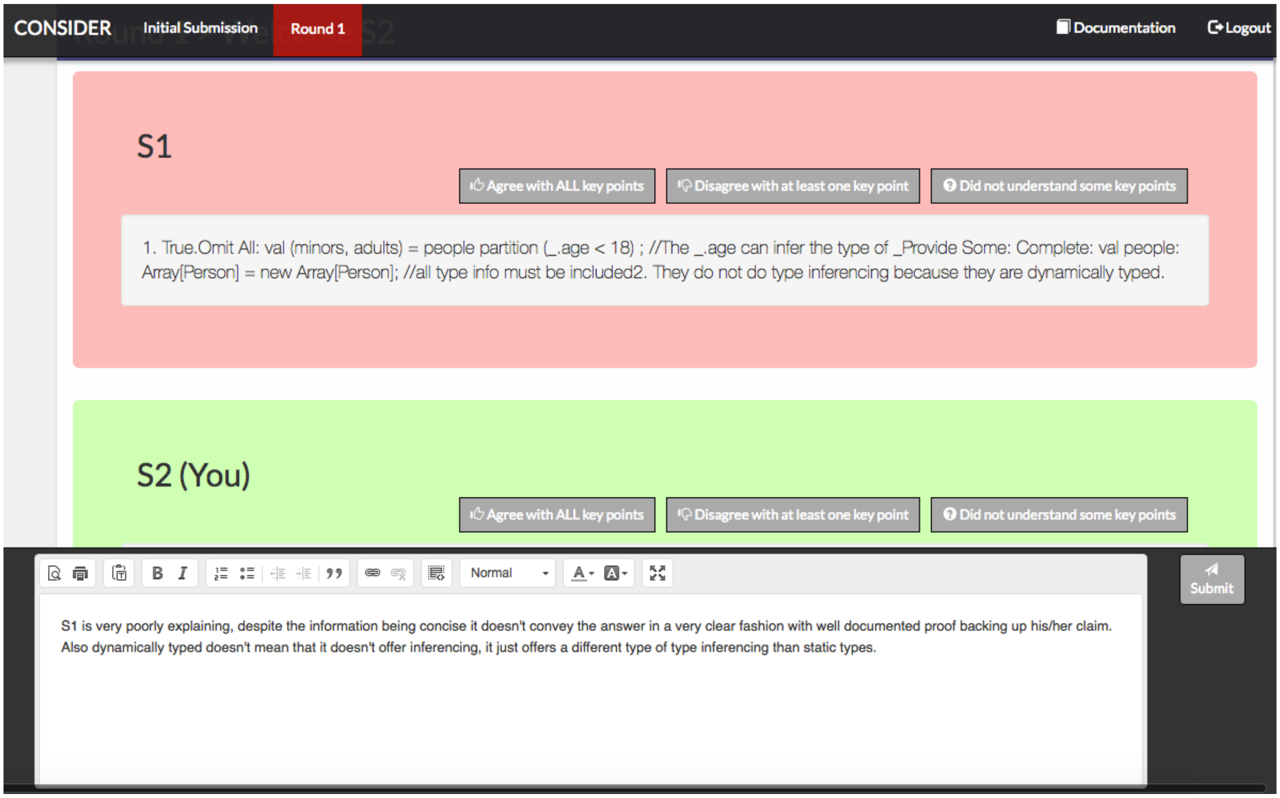

Figure 3: Example of a Discussion Round in CONSIDER

0 Did not attempt
1 Wrong answer, no interesting points
2 Wrong answer but interesting explanation
3 Right answer but without good explanation
4 Right answer with good explanation

We measured the improvement in learning by subtracting each student's score for the initial answer from that for the final answer. Then we tested each improvement distribution for normality using Shapiro-Wilk test ($\alpha = .05$). For both distributions $p < .05$, confirming the distributions are not normal. Hence we employed a non-parametric paired difference test for comparing the vectors using Wilcox signed-rank test.

3.3.2 Survey

An optional, anonymous online survey was administered after both activities were finished to get the participants' opinion on the two approaches. 24 out of the 37 participants completed the survey (65% return rate). In addition to the demographic data (reported above) and feedback on the user interface, it contained questions about students' opinion on how the two features of the CONSIDER approach helped improve their learning. We report their responses to the following Likert items measured on a 5-point Agree-Disagree likert scale, and related open-ended comments. Cronbach's $\alpha > .90$ for these Likert items indicates that the internal reliability of this scale is high.

L01 The CONSIDER activity provided me the opportunity (and the time) to develop a better understanding of the topic than other small group discussion activities (in-class as well as on-line).

L02 Not knowing the identities of the other students in the group had a positive impact on the quality of the discussion.

L03 Organization of the discussion into a series of rounds had a positive impact on the quality of the discussion.

4 Results

4.1 Improvement in Learning

	Piazza	CONSIDER
Mean	0.15	0.65
SD	0.29	0.72
Median	0	1
Min	-0.25	0
Max	0	2
Range	1.25	2

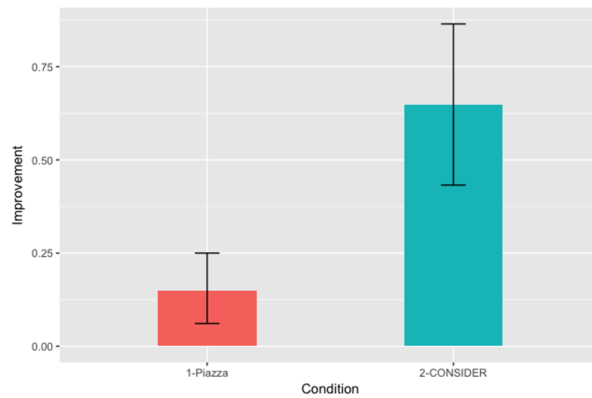

Figure 4: Improvement in Learning, measured under two conditions

The mean improvement, i.e., the difference in the final and the initial answer score on a 4-point scale, is 0.15 for the Piazza activity, and 0.65 for the CONSIDER activity. Since the data does not meet the assumption of normality, a non-parametric statistical comparison was performed using the Wilcox signed-rank test. The analysis shows that the improvement in learning was significantly higher in the CONSIDER activity ($M = .65, Mdn = 1$) compared to the improvement in learning in the Piazza activity ($M = .15, Mdn = 0$), $p = .00025, r = -.60$. The bar graph in Figure 4 shows error bars with 95% confidence interval.

4.2 Survey Questions

75% of the students said that the CONSIDER activity provided them a better opportunity to develop their understanding of the topic compared to other small group discussion activities, like in-class discussions or the Piazza discussion (by responding *Agree* or *Strongly Agree* to Likert item (L01). The bar graphs in Figure 5 show their responses to all three Likert items. 83% respondents said that anonymity had a positive impact on the quality of their discussions (L02) while 75% thought the rounds-based structure also played an important role (L03).

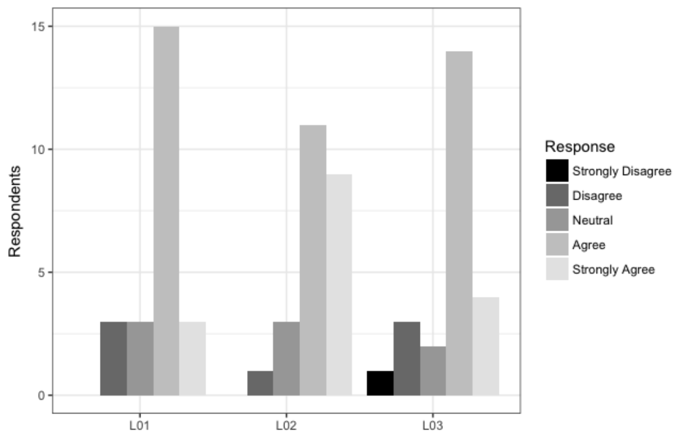

Figure 5: Student responses to the Likert Items

5 Discussion

The results indicate that there is a significant improvement in student's learning using the CONSIDER approach as compared to the discussion-forum approach using Piazza. Looking at the students' open ended comments in the survey responses gives some insights into why CONSIDER approach, with its unique features, is more effective than the regular discussion-forum approach. Their comments are in line with the reasons why we designed those features and included in our approach. One of the key reasons for incorporating a 24 hour rounds-based structure, where posts are made available only at the end of each round, is that it avoids any knee-jerk reactions and frivolous discussions. A participant is *forced* to think through her response (which she can keep editing till the round ends) which results in more meaningful, deeper, and well-reasoned posts. Some students shared this experience in their comments. See, for example: "For me the one post per 24 hour requirement made me think more about each post and force a more thoughtful discussion", and another student's comment: "There is not a need to have an immediate response to others opinions, so it gives more time to research/read up on whatever topic is being discussed to better evaluate correctness". While some students were not pleased with multiple deadlines and felt the rounds structure interfered with the "natural flow of discussions", and even "annoyed" that something was "due every 24 hours for 3-4 days", some of them also expressed that the deadlines "made it far more likely that everyone addressed everything in the discussion . . . ". The use of agree/disagree buttons in CONSIDER served as an ice breaker of sorts, as students could begin by describing why they agreed or disagreed with other students' point of view, whereas in Piazza, some of them found themselves "mostly just waiting for someone else" to point out something in their post, to which they could respond to. Recall we discussed, in Section 1, a similar problem noted by Rick and Guzdial in their research. Like their study, lack of a "starting point" led to lower participation in the Piazza discussion. It must be pointed out, however, that some students clearly preferred the forum-like format of Piazza where the turn around time was much quicker than the 24 hours wait in CONSIDER. We were surprised to find that none of the students mentioned anything about the feature of anonymity in their comments, although more than four-fifths of them clearly thought that not knowing the identities of the other students in the group had a positive impact on the quality of the discussion (Figure 5). We are inclined to believe that they found it trivial to explicitly comment on something that most of them readily agreed with.

We would also like to share some comments that reflect the importance of cognitive conflict, which is at the backbone of our approach. These comments pertain to Piazza and CONSIDER activity, both. Several students expressed that "being exposed to disagreeing opinions was the most useful feature" and it encouraged them "to re-evaluate my answers and think about the problem in more depth than if it had just been presented in a lecture". One of the students said that (s)he enjoyed this activity because "this level of communication is unexpectedly infrequent in college".

Finally, we would like to point out what one of the students expressed about the user interface: he liked the simple, uncluttered, and focused user interface of the CONSIDER app over Piazza, which has, like many other professional softwares, "tons of bells and whistles that don't actually help you learn and just detract from the user experience". While this is perhaps not representative of a large number of students, the point about having too many features that contribute little to the learning is something worth noting for all designers and developers of educational software.

6 Conclusion and Future Work

We have developed a novel approach to collaborative learning, called CONSIDER, that exploits the affordances of online technologies to implement unique features such as structured, rounds based discussions and anonymous posting, which, not only address various challenges other discussion approaches face, but also provide important advantages. Use of cognitive conflict to form the small groups ensures that when the students try to resolve the conflicts, they will develop deep learning of the concepts involved. Anonymity lets students participate more freely, mitigating any personal, ethnic, or academic preconceptions some of them may hold about each others abilities. The round structure gives each student the time and opportunity they need to make meaningful contributions in an equitable way. We have implemented it as a scalable, platform-independent web application using Google App Engine and Python. We studied the effectiveness of our approach in an undergrad programming languages course in Computer Science and Engineering by comparing it with a popular discussion tool, Piazza, which is very commonly used in college courses across the world. Two graded homework assignments were given as part of the experiment, one on Piazza and the other on CONSIDER, to the whole class in two different weeks. 37 students participated in the study. Their answers to the questions posed in the discussions were evaluated on a 4-point scale, and the differences in the post-discussion and pre-discussion scores was compared across two conditions (Piazza v. CONSIDER). The improvement in the scores as a consequence of the CONSIDER discussion was significantly higher than that in the Piazza discussion.

24 students completed the post-activity online survey on how the approach helped their understanding of the concept. Three-fourths of participants responded that CONSIDER provided them a better opportunity to learn than any other in-class or online activity. A very high number of participants said that the two unique features of CONSIDER approach –anonymity and rounds-based structure– helped improve the quality of discussion in their groups (83 and 75% respectively). Their text comments to the reflective questions highlight the importance of the unique features of

CONSIDER.

We plan to further evaluate the efficacy of the features of CONSIDER by designing careful experiments in coming semesters and using the tool in different engineering classrooms. This set of experiments will help us evaluate the effectiveness of these features of CONSIDER. We would also like to perform a detailed analysis on the discussion data from these experiments (including the one reported in this paper) to see how exactly the discussions in the two conditions differ and whether, and if yes, how that affects individual student's learning.

Our tool is available as an open source software (`go.osu.edu/consider`), which other educators can download and configure to use in their courses. It is highly customizable in terms of features such as number of rounds, duration of rounds, group size, etc., to suit their specific needs.

References

[1] M. Cole. Using Wiki technology to support student engagement. *Computers & Education*, 52(1):141–146, January 2009.

[2] J. Rick and M. Guzdial. Situating CoWeb: a scholarship of application. *I. J. of Computer-Supported Collaborative Learning*, 1(1):89–115, March 2006.

[3] C. Crouch and E. Mazur. Peer Instruction: Ten years of experience and results. *American Journal of Physics*, 69(9):970–977, September 2001.

[4] R. J. Dufresne, W. J. Gerace, W. J. Leonard, J. P. Mestre, and L. Wenk. Classtalk: A classroom communication system for active learning. *Journal of Computing in Higher Education*, 7(2): 3–47, March 1996.

[5] J. Piaget. *The early growth of logic in the child*. Routledge, London, 1964.

[6] W. Doise and G. Mugny. Sociocognitive conflict. In *The Social Development of the Intellect*, volume 10, pages 77–101. Pergamon, Amsterdam, 1984.

[7] S. D'Mello, B. Lehman, R. Pekrun, and A. Graesser. Confusion can be beneficial for learning. *Learning and Instruction*, 29:153–170, February 2014.

[8] L. C. Kaczmarczyk, E. R. Petrick, J. P. East, and G. L. Herman. Identifying student misconceptions of programming. page 107. ACM Press, 2010.

[9] K. C. Midkiff, T. A. Litzinger, and D. L. Evans. Development of Engineering Thermodynamics Concept Inventory instruments. pages F2A–F23. IEEE, 2001.

[10] D. Hestenes, M. Wells, and G. Swackhamer. Force concept inventory. *The Physics Teacher*, 30(3):141–158, March 1992.

[11] K. Rollag. Teaching business cases online through discussion boards: Strategies and best practices. *Journal of Management Education*, 34(4):499–526, 2010.

[12] G. Salomon and T. Globerson. When teams do not function the way they ought to. *International Journal of Educational Research*, 13(1):89–99, January 1989.

[13] K. F. Hew and W. S. Cheung. *Student Participation in Online Discussions*. Springer New York, New York, NY, 2012.