Innovative Approach to Online Argumentation and Models for Structuring the Arguments

Neelam Soundarajan
Computer Science & Eng
Ohio State University
Columbus, OH 43210
Email: soundarajan.1@osu.edu

Swaroop Joshi
Computer Science & Eng
Ohio State University
Columbus, OH 43210
Email: joshi.127@osu.edu

***Abstract*—Researchers have stressed the importance of argumentation among small groups of students in STEM courses to help them develop deep understanding. But it is not widely used in college courses due to such challenges as finding time in already packed courses, effective organization of argumentation in large classrooms, etc. This paper presents a novel online approach to enable argumentation to be adopted widely.**

One interesting question we investigated in a junior-level computing course concerned the structure of such arguments. Common experience with *online forums* in courses suggests that a handful of students dominate them while others hardly participate. So we expected that *round-based* discussions where each student in the group made *one* submission in each round, the submission not being available to the others until the start of the next round, would be more effective than *forum-based* discussions where students made as many submissions as they wished and whenever they wished to, and saw each submission as soon as it was made. But to our surprise, the results showed that both were equally effective! We present the details of our approach, the unexpected results from our course, some hypotheses that may explain the results, and future plans to investigate this further.

I. Introduction

Argumentation plays a central role in effecting advances in nearly every STEM discipline. Professionals in such disciplines as Computer Science engage in vigorous arguments about different approaches to address specific technical problems before accepting any particular solution. Online forums such as *stackoverflow* provide excellent venues for such discussions and often host long threads in which professionals in CS and other disciplines argue the merits/demerits of alternative approaches to addressing important questions. Not surprisingly, a number of researchers (e.g., [1], [2], [3], [4], [5], [6]) have stressed the importance of developing the skills of students in STEM discipline to engage in *argumentation*. Much of this work has, however, been at the K-12 level.

Argumentation is even more important for undergraduates in computing and engineering and other STEM fields as they get prepared for their professional careers where they can expect to engage in vigorous arguments not only with the broader professional communities in their respective disciplines, but also with other members of their project teams and other interested constituents to defend their specific choices in design and implementation projects. Further, argumentation will help undergraduate students develop deep understanding of new technical concepts. Indeed, in a real sense, argumentation plays a very similar role in helping students develop understanding of new concepts as it does in professionals' online arguments on alternative solutions to new technical problems. While in the latter case the professional is challenged to critically consider and debate alternative approaches to (possibly) stretching the state of the art to address a new problem, in the former case the student is in the process of stretching his/her conceptual understanding of the field; and, as in the professional context, an effective way to help ensure this is to have the student discuss/debate the topic in question with peers, i.e., other students, especially those who seem to have a different conception of the topic.

Prior research has shown that some key requirements must be met to ensure that argumentation is most productive: The argumentation must be in small groups of 4–5 students each; each group must include students with *different* approaches to the topic; and the instructor should *not* participate in the discussion. The last requirement may seem surprising but it is critical since, otherwise, the students may simply accept what the instructor says without careful analysis and the goals of helping them acquire *deep* understanding as well as preparing them for engaging in effective discussions as professionals will be compromised.

Even if we succeed in meeting these requirements, there are a number of challenging issues that must be addressed if argumentation is to be widely used in computing/engineering courses. First, how would faculty find time in their already packed courses to accomodate small-group argumentation? Second, wouldn't the most vocal students dominate such discussions while others stay in the background? Third, wouldn't stereotypical biases some students may harbor concerning the abilities of others seriously affect the discussions? Etc.

We have developed a highly innovative approach and online system, *CONSIDER*, to address these and other problems. The name was chosen to stress that students in a group are expected to carefully consider the positions of the other students and, possibly, revise or refine their own positions based on those; it is also an acronym, see below, intended to capture other aspects that are central to how the system is designed to function. A CONSIDER discussion starts with the instructor posting, on the system, a suitable problem. Each student then

978-1-5386-1174-6/18/$31.00 ©2018 IEEE

submits his/her individual answer by a specified deadline. Next, the instructor uses the system to form groups (typically consisting of 4 or 5 students each) based on these submissions, with each group including students with *conflicting* ideas about/approaches to the problem; and the discussion begins. The discussion may be customized in various ways. It may be specified to be *anonymous* with students in each group being labeled S1, S2, S3, S4 or they may know each other's identities; the discussion may be organized in a series of *rounds* with each student making one submission in each round and the other students not seeing the submission until the start of the next round or it may be organized in a more *forum-like* manner with each submission becoming available to the group as soon as it is made; etc. In each case, the student should specify whether he/she agrees or disagrees with the positions of each of the students in the group. The name CONSIDER is an acronym for *con*flicting *s*tudent *i*deas *d*iscussed, *e*valuated, and *r*esolved (or *r*efuted!). As the acronym suggests, the central goal is to ensure that students in the group carefully consider *conflicting* conceptions held by the other students in the group and, as appropriate, refine/correct their own conception of the topic in question.

We should note here that while participating in CONSIDER discussions will help students develop strong argumentation skills, that is not the primary goal. After all, when STEM professionals engage in argumentation, the goal is to arrive at the most appropriate scientific or technical answer to the question being discussed, not win the argument as may be the case in, say, a legal setting. Rather, the goal is to help students develop deep conceptual understanding while also sharpening their skills at recognizing and being open to well-justified technical positions that may differ from their own. This is not to say that the ability to win debates is not important, just that CONSIDER-based activities will not contribute to those abilities. We should also note, as pointed out by one of the referees, CONSIDER-based activities will not contribute to developing students' *oral* argumentation and presentation skills. Again, this is not to say that these are not important; just that other parts of the student's curriculum, such as in-class, formal debates in courses focused on professional ethical issues, will hopefully address these important skills.

We have used CONSIDER in some junior-level computing courses. One key question that we were interested in investigating was the most effective way of structuring the argumentation to best help students develop deep understanding. The reason the question is important is that *online discussion forums* used by many faculty to encourage students to engage in discussions about the course topics have been surprisingly ineffective, see, e.g., [7], [8]. Often, a handful of students dominate the forum while others are passive observers or ignore it altogether. Our hypothesis was that the reason for this was the way that the forums were structured so that, when a topic was being discussed, each student made as many or as few *posts* to the forum as he/she chose and whenever the student chose to do so. Further, each post became available to the entire class as soon as it was made. With this structure, a common occurrence is for a handful of, often just two, students to engage in a back-and-forth discussion while the others in the class, even the interested students, soon lose track of the key points being discussed.

We proposed an alternative structure, a *round-based* one, in which each student in the group is required to make *one* post in each *round* (whose duration will likely depend on the topic in question and course logistics). The student's post for a round will *not* become available to other students in the group until after the end of the current round; indeed, the student would be able to, if he/she wished to do so, *edit* the post until the current round ends since no others would have seen the post until that point. Such a round-based structure rather than the forum-based structure of discussions, we hypothesized, would ensure active and engaged participation by all students in the group and result in better learning.

We tested the hypothesis in a junior-level course of principles of programming languages (PL). Using the customization facilities of CONSIDER, we investigated the effectiveness of the two structures in the context of two (fairly typical) topics in the course. To our surprise, the results showed that that both were equally effective. In Section II, we summarize the framework underlying the approach and other related work. In Section III, we describe the CONSIDER system; as we explain, the system was developed over several semesters, following the methodology of *design-based research*. In Section IV, we present our research question, the experimental design that we used in the principles of PL course, and the unexpected results we obtained. In Section V, we discuss the results, present some hypotheses that may explain our results, and future plans to investigate this further; we also summarize our plans for future evolution of the CONSIDER system.

II. Background and Theoretical Framework

Socio-cognitive conflict, a key concept underlying the CONSIDER system, originates in Piaget's classic work [9] on children's learning. The key point of Piaget's theory was that peer interaction is a potent component of a learner's grasp of new concepts. In particular, socio-cognitive conflict, i.e., disagreements with *other learners'* conception of the same problem or topic and interaction with peers to resolve the disagreements is fundamental since it highlights alternatives to the learner's own conception. In resolving the conflict, the learner is forced to consider and evaluate these alternatives on *equal* terms. Note that a *teacher* is *not* involved except, possibly, as an observer. This is critical since, as noted earlier, if a teacher were to participate in the discussion, the learners are likely to simply accept what the teacher says without careful analysis, thereby compromising the depth of learning. As Howe and Tolmie [10] put it, "conceptual growth depends on *equilibriation*, that is the reconciliation of conflicts between prior and newly experienced conceptions".

Although Piaget was concerned mainly with the intellectual growth of children, his ideas are very relevant for adult learners as well, including undergraduate STEM students. Indeed, resolving socio-cognitive conflicts should be more effective for college students than for young children since college students may be more capable than young children of analyzing and evaluating others' ideas that might conflict with their own. And given the serious problem of *misconceptions* harbored by students in many STEM disciplines that researchers have investigated, see, e.g., [11], an approach that may help address such misconceptions is clearly worth pursuing.

A different approach, one that has been commonly used, to trigger cognitive conflict is for the *instructor* to present *anomalous data*, i.e., data that conflicts with the students' prior misconceptions. The expectation is that the conflict between the presented information and the student's prior conception will trigger *disequilibrium* and cause the student to revise his/her conception. But Chinn and Brewer's work [12] showed that this approach failed to trigger conceptual change in a large majority of college students in STEM courses. Given the authority of the teacher, many of the students seem to simply accept whatever the teacher says without much analysis. As a result, deep down, there was no real conceptual change. To put it differently, in cases where a student's understanding conflicts with the explanation provided by the instructor, the student simply accepts the explanation without critical evaluation. By contrast, when the (cognitive) conflict is between a given student's conceptualization of the topic and those of her *peers*, the student is forced to evaluate the alternatives critically and pick one[1] after *careful deliberation* since, as far as the student knows, she, rather than the peer, may be the one whose explanation is correct!

It may be useful to note the important distinction between this approach and Vygotsky's notion of *zone of proximal development* (ZPD) [13]. ZPD stresses the importance of a "*more* competent other" in the interaction; thus, according to Vygotsky, interaction is most fruitful when one of the members of the group is more competent than the others and can help the other members move beyond their current abilities (into their ZPD). Interestingly, while some researchers have confirmed the importance of Vygotsky's "more competent other", the results of other researchers (see, e.g., [14]) suggest that what matters most is the cognitive conflict that a student experiences because of disagreements with other students' conception of the same problem or topic. In any case, the approach in CONSIDER is based on socio-cognitive conflict, not ZPD.

It is also worth noting that in the last two decades or more, there has been considerable focus on *collaborative learning* in STEM education. Thus, e.g., team projects in capstone design courses are often considered an essential part of undergraduate engineering programs. While collaborative learning is indeed important, it is not directly relevant to our work since it does not, for the most part, involve students in a team trying to resolve cognitive conflicts. Indeed, students in such teams often go out of their way to not criticize the ideas offered by other members of the team for fear of offending them.

[1]More commonly, the student will revise her original conception incorporating ideas from other students' conceptions rather than simply abandoning her original conception and picking one of the others.

Driver *et al.* [1] make a strong case for *argumentation* as a central component of STEM education. To quote, "[a]s argument is a central feature of the resolution of scientific controversies, it is somewhat surprising that science teaching has paid so little attention to a practice that lies at the heart of science. It is our contention that this significant omission has led to important shortcomings ... if science education is to help young people engage with the claims produced by science-in-the-making, science education must give access to these forms of argument through promoting appropriate classroom activities and their associated practices." Given that CONSIDER, as we will see, requires students to offer arguments defending their positions, it has the potential to help students develop strong argumentation skills. But we should note that our *primary* goal is to help students to develop deep conceptual understanding; the fact that, in the process, they will develop strong argumentation skills is an added bonus not the main goal. In Nussbaum's terms [15], our interest is in having students "arguing to learn," not "learning to argue".

Socio-cognitive conflict is also the primary driving force behind the in-class *peer instruction* (PI) technique developed by Mazur and others [16], [17]. In PI, 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 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. However, there are a number of limitations with this and similar approaches, mostly related to the fact that it is a classroom technique, the activity being interspersed with regular lectures. First, since the topic in question was just discussed in the lecture, students may not have thought about the underlying ideas carefully. Second, there is no way to ensure that students in a given group include ones who picked different possible answers because the grouping is based essentially on where students are seated. Third, some students, not necessarily the ones with the most developed understanding of the topic, may dominate their groups; and any stereotypical biases that students may harbor, perhaps subconsciously, may also compromise the discussion. Further, the amount of time spent in the discussion is, naturally, limited; hence, students who take time to formulate precise and deliberate arguments may not contribute effectively to the discussions. As we will see in Section III, the CONSIDER approach addresses all these problems effectively.

While CSILE [18], perhaps the earliest online approaches used in science courses was developed before the web, more recent systems that attempt to have students build knowledge in somewhat similar ways use *wikis*, see, e.g., [19]. Unfortunately, many of these efforts have not been effective in improving learning even if wiki-based knowledge-building efforts outside the classroom have been quite successful, the best example being the Wikipedia. Thus Cole's [8] course on *information systems* 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 and discuss the material. Students were told that fully one quarter of the questions on the final exam would be from the material that students posted. The expectation was that, given this, students would eagerly post content, edit each other's posts, and engage in active discussions. Halfway through the course there had been *no* posts to the wiki! Leung and Chu [20] in a course on *knowledge management* and Judd *et al.* [21] in a large course on psychology report equally poor results of the use of a wiki. Rick and Guzdial [7] report that although they obtained positive results using wikis in architecture and english composition classes, the results in STEM classes were "overwhelmingly disappointing". Thus they report that fully 40% of math students settled for a zero on an assignment rather than engage in such discussions!

Over the last few years, a number of systems specifically designed to support online argumentation in courses have been developed, see, [22], [23], [24]. Many of these, though, are part of larger systems intended to help students, for example in high-school chemistry classes, to engage in collaborative knowledge *construction*, following principles of constructionism. As such, they often include elaborate graphical (and, often, video) facilities to enable students to engage in the necessary experimentation, literature search, etc. The entire course is often designed around the system in question. By contrast, CONSIDER is intended for use in standard undergraduate computing and engineering courses to help students develop deeper conceptual understanding of the concepts and topics presented in lectures in the course in the standard fashion.

III. CONSIDER Approach and System

Design-based research (DBR) is an effective approach for both research and design of technology-based or technology-enhanced learning environments; see, for example, the paper by Wang and Hannafin [25] for a review of the topic. Roughly speaking, in DBR, the researchers start with an educational theory, design a system/intervention informed by the theory, try it in practice, and, based on the results, revise the system/intervention in an iterative cycle. This is the approach we have adopted in our work. As described earlier, socio-cognitive conflict among learners and how its resolution can drive the development of deep understanding among them provides the theoretical basis for the CONSIDER approach. We built a prototype system and over a few semesters of use in CS courses, on the basis of feedback from instructors and students, revised and refined it in an iterative cycle.

We will use an example from our Software Engineering (SE) course to illustrate the overall structure of a CONSIDER activity. This is a typical SE course, taken by juniors/seniors majoring in Computer Science. A main goal of the course is to help students recognize the importance of a systematic approach to understanding the overall domain in which the software system to be built is intended to operate, understand the problem that the system will help address, and the solution approach to be adopted in the system. Quite often, however, students want to jump straight into designing and coding the software system without going through a careful analysis of the domain, the problem in the context of the domain, etc. Indeed, frequently there is confusion between the domain problem and specific algorithmic or data-structure related problems that might be encountered when developing the software system. The problem below is intended to help tease out such misunderstandings.

Homework: Your team has been asked to build a campus wayfinding system to help visually impaired students on the campus. Five items identified during analysis are listed belw. Identify which category of analysis –that is, domain, problem, or solution– each element falls under. Briefly explain why.

1) A catalog of the types of building on a college campus;
2) The list of hard-to-find buildings on campus;
3) The range of visual and cognitive impairments that people suffer from;
4) Strategies by which people find their way in an unknown area, e.g., asking passers-by or identifying major streets.

Item (3) is especially interesting. Many students think it falls under the *problem* category. In fact, however, it is part of the *domain* as it provides information about the range of impairments people suffer from. The software system, after all, is *not* intended to solve the problem of visual impairments (e.g., by controlling an artificial eye to help the person see).

Different students come up with different answers and with different justifications. The standard approach is to have a discussion on the question in class, typically in the same class period as the one in which the graded homeworks are returned to the students. The class discussion helps some students, but others remain unclear about the distinction between the notions of domain, problem, and solution. In the CONSIDER approach, following the lectures on the topic, the instructor posts the problem on the CONSIDER web-app.[2] She also specifies –in addition to other items, see below– a deadline by which each student must submit his[3] answer. The problem may be similar to or the same as one above but we will assume there is only one question, item (3) above.[4]

Once the instructor has posted the homework, each student receives an email from the app asking him to log into the system and answer the question. The app will require the student to make a specific choice ("domain" or "problem" or "solution") and to include a brief *justification* as part of the

[2]In previous versions of the system, we had implemented it as an Android app since we felt students would find it easier to work with as most students use their smartphones frequently each day. But while it was true that students did indeed use their phones frequently, participating in a CONSIDER discussion was quite a challenge using a smartphone because of the small screen size, lack of keyboard, and various distractions that are all too commmon on such phones. Hence we have re-implemented the system as a web-app.

[3]In the interest of readability, we use female pronouns when referring to instructors and male pronouns for students.

[4]Later in the paper, we will argue that the structure of the problem as stated above is, in fact, preferable to one in which only item (3) is included.

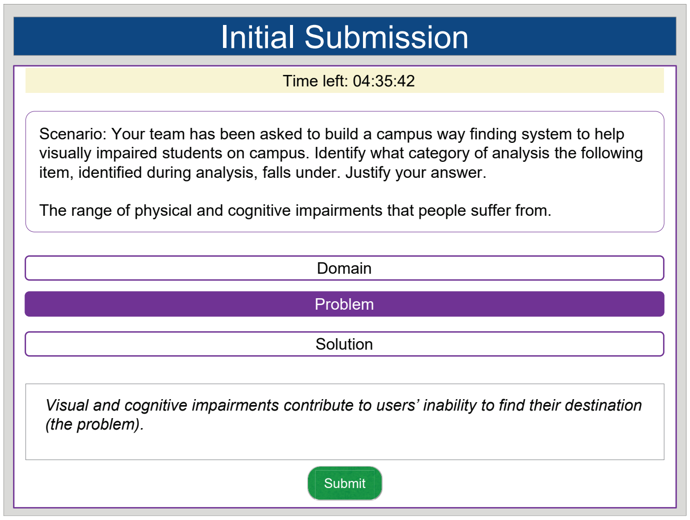

Fig. 1. Initial submission from a student (wrong answer)

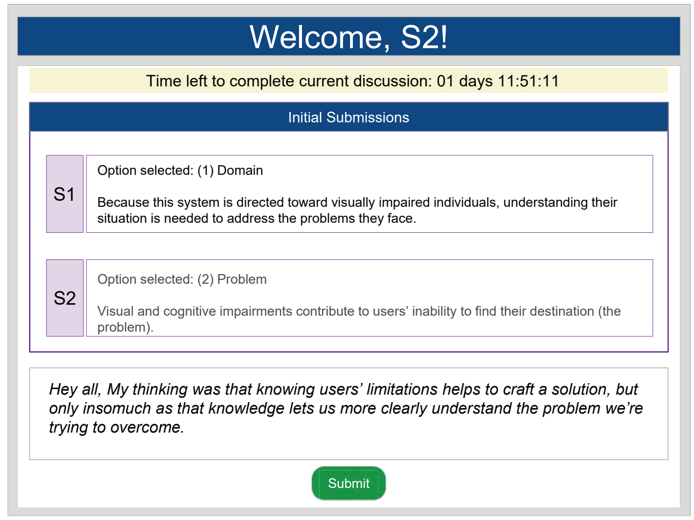

Fig. 2. Student S2's Round_1 submission

answer. We refer to this as the student's *initial submission*. Fig. 1 shows the initial submission made by one of the students; this student indeed has a misconception and thinks the specified item belongs to the *Problem* category. Two points should be noted. First, no groups have yet been formed, thus each initial submission is made by an individual student and reflects that student's (initial/current) conception of the problem. Second, after making the submission, the student is free to log in again as many times as he wishes (until the time of the deadline) and modify the submission in any manner; the system will retain, for each student, only the last version submitted by that student before the deadline.

Once the deadline expires, the system will (try to) automatically form groups of 4 or 5 students each with each group containing students who chose different answers. If most students make the same choices, the instructor will have to form the groups based on differences in the students' justifications. We will return to this issue later in the paper, but we note here that the app allows the instructor to specify a "buffer" period (typically several hours) between the deadline for the students' initial submissions and the start of the next phase, i.e., the discussion portion of the activity to provide the instructor adequate time to log into the app and form these groups "by hand" if necessary.[5]

When the discussion phase begins, the app will send another email to each student indicating that the discussion phase had started, so he should log in and start participating in the discussion.[6] Before considering how the discussion takes place, we should note that when the instructor creates the assignment on the app, she will also specify various important aspects of the discussion such as whether the discussion will be *forum-based* or *round-based*; whether it will be anonymous, so that students in each group will know each other as *S1, S2* etc. or students will see the identities of the other students in the group; the deadline for each phase of the activity; in the case of round-based discussions, this will include specifying the number of rounds and the deadline for each round.

For our current example, let us assume that the instructor has chosen the discussion to be anonymous and round-based. In a round-based discussion, each round will be of a fixed duration. During each round, each student is required to make one post before the deadline for the round expires; but, as in the case of the initial submission, the student may log in as many times as he chooses before the deadline and edit her post as she chooses; and, as in that case, only the most recent version of the post will be saved. In our experience, a duration of 24 hours per round seems ideal. It allows students sufficient time to correct any mistakes they might make when making a submission for a given round, and it also accounts for the varying time schedules of undergraduate students who often juggle school, work, and family commitments. Also in our experience, the appropriate number of rounds for typical homeworks in courses in computing at this level seems to be one or two. In any case, the app allows the instructor to tweak these parameters to suit the particular course and the nature of the homework and her own and the students' preferences.

Suppose the student whose initial submission is shown in Fig. 1 has been assigned (either automatically or by the instructor) to a particular group G and that this group has three other students. Since this is an anonymous discussion, the system will assign the labels S1, S2, S3, S4 to the four students. Let us assume that the system has assigned the label S2 to our particular student. We will label the rounds in the discussion, $Round_1$, $Round_2$, etc. For convenience, we will also refer to the initial submission round as $Round_0$.

When S2 logs in, once $Round_1$ begins, he is presented with the initial answers submitted by each student in G (including

[5]If the approach is to be adopted for large classes, possibly even MOOCs, it would not be possible for the instructor to have to form the groups by-hand. In the final section, we will consider how this problem might be addressed.

[6]The app also allows students to see, any time after the initial posting of the activity, the start-time and end-time for each phase of the activity so this email simply serves as a reminder.

himself), Fig. 2. For lack of space, only the initial answers of S1 and S2 (the current student) appear in the screenshot; S2 will be able to see the answers submitted by S3 and S4 by scrolling down as needed in the central window. Note that *S1* has submitted the correct answer. The hope is that when S2 reads this correct answer, he will resolve the resulting conflict by correcting his own misconception.

At the bottom of the screen is a window where S2 is expected to type in his post for the current round. Unfortunately, S2 did *not* understand the rationale behind S1's answer and, therefore, in his post for the current round, tries to justify his incorrect conception. A couple of points should be noted. The other students in the group, S1, S3, S4 may also be logged in at the same time and working on their $Round_1$ posts; some may have already made those posts; others may not yet have logged in for this rould. In none of these cases, this being a round-based discussion, will any of the students see the posts that any of the others in the group may have made for the current round. This allows students to work at their own pace and, since a student can log in again (multiple times, if he chooses) and edit his post for the current round; e.g., S2 might, before the end of this round, suddenly realize the mistake in his conception, log in again, and modify his answer.

Fig. 2 shows the style used in a previous version of CONSIDER. That version did not require the student to *specifically indicate* whether he agreed with or disagreed with the answers posted by the other students in the group. Although S2, judging from his post in the figure, has indeed read and considered S1's post, we were troubled by the possibility that some students may not be carefully reading or considering the posts of the other students and may be just repeating their previous positions. Hence we revised the app to include, next to the previous-round post of each student, three buttons reading "Agree", "Disagree", or "Neutral". S2 is now *required* to click one of these buttons to indicate that he agrees with all/most of the main points expressed in that particular previous-round post (by clicking the *green* "Agree" button), or he disagrees with one or more key points in that post (by clicking the *red* "Disagree" button), or he did not understand one or more key points in that post (by clicking the *blue* "Neutral" button). His selection is reflected in the app with some visual aid: the background of the author's alias changes to the respective color (green, red, or blue for agree, disagree, or neutral, respectively), and an icon thumbs-up, thumbs-down, or question-mark, respectively, appears under it. An example can be seen in Fig. 3.

Fig. 3 shows S2's $Round_2$ post in this version of the app. At this point, S2 has indeed understood the rationale for S1's position and recognized the correctness of that postion. Thus S2 is required to consider posts made by each student in the group G in the previous round and analyze its relation to his current position. S2 has to do this for his *own* post as well from that round. This is important because S2 may find the post(s) of one (or more) of the other students from the previous round so compelling that he changes his mind and no longer agrees with what he said in the previous round!

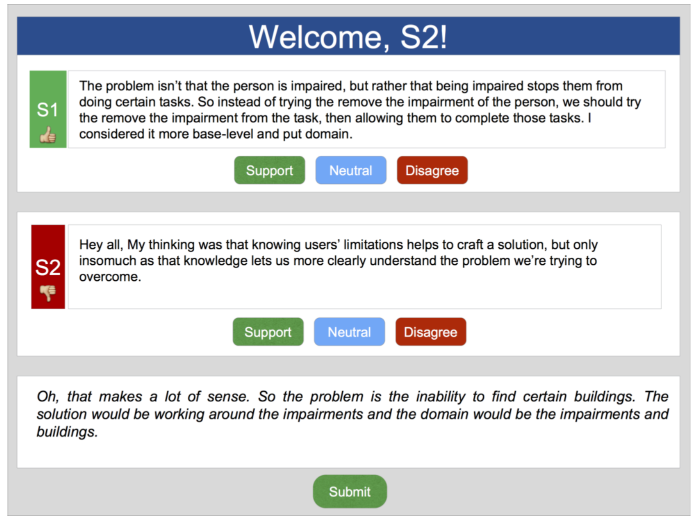

Fig. 3. Student S2's $Round_2$ submission

S2's current-round post which he types into the window at the bottom provides an explanation for why he no longer agrees with his previous position. Note the fundamental difference with a typical debate where such a change of position would be considered a defeat for S2 rather than, as is the case here, the *purpose* of the activity. The final submission round, whose details we omit, starts at the end of the last discussion round. Each student is required to, individually, submit his final answer to the question along with a brief justification relating it to the positions of the other students in his group.

Before concluding this section, we note one other point. One question that some students in the course raised during an informal class discussion on CONSIDER was, what happens if a given group is not able to arrive at a common conclusion? Indeed, just this question was also raised by one of the anonymous referees of the paper. The point is that arriving at a group consensus is *not* a goal of a CONSIDER activity; that may, indeed, happen but it is not a goal of the activity. Instead, the goal is to have each student in the group carefully analyze the opinions of each of the other students in the group and, if necessary, revise/refine his/her approach to the problem and to be able to explain/justify any changes so that each student in the group develops as deep an understanding of the topic as possible. We should also note that the situation may well be different in a professional project team, especially one with looming deadlines; in such a situation, the team may indeed have to reach a consensus on the approach to be adopted.

IV. Experimental Results

Section III focused on the *rounds-based* approach for the discussion phase of the activity. The alternative, as described earlier, is the *forum-based* approach in which each student posts as frequently or infrequently as he/she chooses and the post becomes available to the group as soon as it is made; and a key research question we were interested in was:

TABLE I
SUMMARY OF INITIAL AND FINAL SUBMISSION SCORES

Activity	Expt. condition	N	Submission	Median	Mean
$A1$	Round-based (H1)	11	Initial	2.000	2.091
			Final	2.500	2.409
	Forum-based (H2)	11	Initial	3.000	2.545
			Final	3.500	3.136
$A2$	Round-based (H2)	11	Initial	2.000	2.364
			Final	3.000	3.000
	Forum-based (H1)	14	Initial	2.500	2.429
			Final	3.000	2.929

> $R1$*: Is the rounds-based approach more effective than the forum-based approach in enabling students to develop conceptual understanding?*

We investigated this in a junior-level course on the concepts of programming languages, taken by most of our CS-majors. We divided the class of about 40 students randomly into two roughly equal halves, H1 and H2. Two activities, $A1$ and $A2$, related to topics discussed in the course, were assigned to all students, each activity being assigned in the days following the respective class discussions.[7] Students in H1 went through the $A1$ activity using the *rounds-based* approach and the $A2$ activity using the *forum-based* approach; those in H2 went through $A2$ using the *rounds-based* approach and $A2$ using the *forum-based* approach. Nearly all students in the class consented to having data from their work, after anonymization, to be used as part of our research study. But we have included in our analysis data from only those who posted both the initial and the final submission for the given activity. Due to various external reasons, like conflict with assignments from other classes, some students could not submit either the initial or the final answer, and were excluded from the analysis.

The course instructor, one of the current authors, developed a simple 4-point rubric to assess the correctness of the answers submitted by individual students in the initial and final submission for each activity. Without going into the details of the rubric (which depend on the specific technical programming language concepts that the activities were based on), a score of 1 meant that the student's answer was wrong and, moreover, did not mention any relevant interesting points; a score of 2 meant the answer was still wrong but the student offered an interesting explanation; etc.

Table I shows a summary of these measures across the two activities and experimental conditions. As explained above, for activity $A1$, the H1 half of the class participated in the *rounds-based* discussion, while the other half, H2, participated in the *forum-based* discussion. Each student individually submitted an answer to the question before and after the discussion. The course instructor then evaluated these answers on the 4-point scale discussed above (*Initial* and *Final* scores in the table, respectively). The last two columns of the table show the median and mean of these scores for the halves H1 and H2 under each experimental condition. The difference between the score for the initial submission of a given student for a given activity and the score for the same student's final submission was our measure of the activity's contribution to the student's understanding of the particular topic. We first analyzed whether the discussion interventions –round-based or forum-based– have any statistically significant effect on the students' understanding. Shapiro-Wilk normality test on each of the 8 metrics indicated that the data is not normally distributed ($p < .05$). Therefore, we used the Wilcoxon signed-rank test, a non-parametric equivalent of the one-tailed t-test for within-subject data. The analysis showed that the final submission scores were significantly higher than the corresponding initial submission scores in *all* the four conditions: $A1$ Round-based with group H1 ($N = 11, p < .05, r = -.86$), $A1$ Forum-based with group H2 ($N = 11, p < .05, r = -.77$), $A2$ Round-based with group H1 ($N = 14, p < .05, r = -.87$), and $A2$ Forum-based with group H2 ($N = 11$, $p < .05$, $r = -.77$). These results indicate that both discussion structures resulted in a significant improvement in students' understanding.

TABLE II
SUMMARY OF GAIN SCORES

Activity	Structure	Median gain	Mean gain
$A1$	Round-based	0.5000	0.7727
	Forum-based	0.5000	0.5909
$A2$	Round-based	0.5000	0.6364
	Forum-based	0.5000	0.5000

Turning to the research question mentioned above, our hypothesis was the rounds-based discussion will be more effective than the forum-based discussion. To evaluate this hypothesis, we performed a gain-score analysis by comparing the mean *gain* (difference between the final submission and initial submission score) under one condition (round-based) with the mean gain in learning in the other condition (forum-based) for a given activity. Table II summarizes the improvement in learning from the initial submission to the final submission for each activity, first in the round-based structure and then in the forum-based structure, observed in each condition. Once again Shapiro-Wilk normality test showed that the gain vectors are not normal either ($p < .05$), so we used the non-parametric Wilcoxon *rank sum* test, which is a nonparametric equivalent of the t-test for *between*-subject data. The analysis, however, showed that there was no significant difference in improvement in learning in the two conditions on either activity ($p > .05$).

This was rather surprising since we had expected that the *round-based* approach would be more effective in improving

[7] The topics in question are fairly standard for courses on programming language concepts at the junior level. $A1$ dealt with *subtype/inheritance-based polymorphism* in OO languages such as Java. $A2$ concerned the nature of variables in *functional* languages such as (pure) Lisp or Haskell.

student understanding than the much-less structured *forum-based* approach. Of course, this is only one experiment and we plan to repeat it but, nevertheless, the result was surprising. In the final section, we consider some possible explanations and our plans for investigating this further.

V. Discussion and Future Work

Assuming that further experimentation confirms the result reported in the previous section that round-based and forum-based discussions are equally effective, each being fairly effective in improving student understanding of the topic, why is that? Doesn't that contradict the experience of various authors we mentioned earlier, e.g., [7]? One possible reason is the fact that our activities were in *small groups* of students, typically four students each. A second possible factor might be the fact that students in our group (for both activities, in both conditions) were anonymous.

But we believe that the most important reason might be the fact that our groups were *heterogeneous*, i.e., they consisted of students with different understandings of the topic. This was, of course, the central point of Piaget's theory of learning driven by socio-cognitive conflict. It is this conflict that forces the student to consider and evaluate the alternative explanations on equal terms. We plan to investigate this in future work; the approach would be to have half of the class work on an activity with the groups in this half being formed based on conflicting approaches to the problem, these conflicts being identified on the basis of their initial submissions; and have the second half of the class work on the same activity but the groups in this half being formed at random. A comparison of the results across these two conditions should help shed light on the question.

Formation of conflicting groups, however, is not always an easy task. The particular software-engineering problem is perhaps an exception because the question of whether a given item belongs to the domain category, the problem category, or the solution category turns out to be challenging for many students encountering this concept. As a result, it is usually fairly straightforward to form such groups but this is not often the case. We have been exploring possible ways to help with this situation. One possibility is suggested by the original version of the software-engineering problem we considered; i.e., have the student, at least in the initial submission, consider which category *each* of the four elements listed near the start of Section 3, rather than just one of them as we did in our detailed discussion in that section. That would enable differences between the students to emerge more easily. Indeed, an approach such as this would be essential if the approach is to be applied to large courses including, possibly, MOOCs. For such courses, the possibility of having the instructor form the groups "by-hand" would be an unacceptable burden, at least as it concerns the amount of "buffer time" that would be needed between when the students make their initial submission and when the discussion portion of the activity begins. We plan to implement this facility in CONSIDER and study its effectiveness in helping form conflicting groups.

Before concluding, we mention one other important point. In the current version of CONSIDER, in their final submissions, students essentially address the same problem as they did in the initial submission and in the discussion in their group. This raises the important question: how can we be sure that what is happening is not simply that after the discussion, it is easy to see, even for students who started with little understanding, which of the answers proposed by the different students is correct and which incorrect? Indeed, one could argue that this is an especially significant problem for STEM disciplines because here, unlike in social science topics, there is generally one "correct answer"[8]. We adapted this approach since we had received informal feedback from the students to the effect that posing a different problem for the final submission would substantially increase the amount of work they were expected to complete for the course; and given all the other pressures on undergraduate students, this seemed a legitimate concern. Neverthelss, given the question that it raises, it would seem useful to have the student address, in the final submission, a *different* problem, but still one that is very much related to the topic of the original problem, rather than the same as that one. Such an approach will provide definitive evidence of how any particular activity does or does not contribute to the students' conceptual understanding. As for the students' concern regarding the amount of work they would have to put in, we will have to appropriately adjust other components of the course to address this.

References

[1] R. Driver, P. Newton, J. Osborne, "Establishing the norms of scientific argumentation in classrooms," *Science education*, vol. 84, no. 3, pp. 287–312, 2000.

[2] J. Osborne, S. Erduran, and S. Simon, "Enhancing the quality of argumentation in school science," *Journal of Research in Science Teaching*, vol. 41, no. 10, pp. 994–1020, 2004.

[3] N. Mirza and A. Perret-Clermont, *Argumentation and Education*. Springer, 2009.

[4] A. Weinberger and F. Fischer, "A framework to analyze argumentative knowledge construction in computer-supported collaborative learning," *Computers & education*, vol. 46, no. 1, pp. 71–95, 2006.

[5] K. Yeh and H. She, "On-line synchronous scientific argumentation learning: Nurturing students' argumentation ability and conceptual change in science context," *Computers & Education*, vol. 55, no. 2, pp. 586–602, 2010.

[6] A. Zohar and F. Nemet, "Fostering students' knowledge and argumentation skills through dilemmas in human genetics," *Journal of research in science teaching*, vol. 39, no. 1, pp. 35–62, 2002.

[7] J. Rick and M. Guzdial, "Situating CoWeb: A scholarship of application," *Int. J. of Computer Supported Collaborative Learning*, vol. 2, pp. 89–115, 2006.

[8] M. Cole, "Using wiki technology to support student engagement: Lessons from the trenches," *Computers & Education*, vol. 52, pp. 141–146, 2009.

[9] J. Piaget, *The early growth of logic in the child*. Routledge and Kegan Paul, 1964.

[10] C. Howe and A. Tolmie, "Productive interaction in the context of computer-supported collaborative learning in science," in *Learning with computers*. Routledge, 1999, pp. 24–46.

[8]This, of course, doesn't apply to such cases as *design problems* intended to build a system of some kind, such as in engineering projects and the like.

[11] S. Singer, N. Nielsen, H. Schweingruber *et al.*, *Discipline-based education research: understanding and improving learning in undergraduate science and engineering*. National Academies Press, 2012.

[12] C. Chinn and W. Brewer, "An empirical test of a taxonomy of responses to anomalous data in science," *Journal of Research in Science Teaching*, vol. 35, no. 6, pp. 623–654, 1998.

[13] L. Vygotsky, *Mind in society: The development of higher psychological processes*. Harvard University Press, 1978.

[14] W. Doise and G. Mugny, *The social development of the intellect*. Oxford: Pergamon, 1984.

[15] E. Nussbaum, "Argumentation in and student-centered learning environments," in *Theoretical foundations of learning environments*. Routledge, 2012, pp. 114–140.

[16] C. Crouch and E. Mazur, "Peer instruction: Ten years of experience and results," *American Journal of Physics*, vol. 69, no. 9, pp. 970–977, 2001.

[17] R. Dufresne, W. Gerace, W. Leonard, J. Mestre, and L. Wenk, "Classtalk: A classroom communication system for active learning," *Journal of computing in higher education*, vol. 7, no. 2, pp. 3–47, 1996.

[18] M. Scardamalia and C. Bereiter, "Technologies for knowledge-building discourse," *Comm. of the ACM*, vol. 36, no. 5, pp. 37–41, 1993.

[19] J. Larusson and R. Alterman, "Wikis to support the collaborative part of collaborative learning," *International Journal of Computer-Supported Collaborative Learning*, vol. 4, pp. 371–402, 2009.

[20] K. Leung and S. Chu, "Using wikis for collaborative learning: A case study of an undergraduate students' group project," in *Proc. of Int. Conf. on Knowledge Mgmt.*, 2009, pp. 1–14.

[21] T. Judd, G. Kennedy, and S. Cropper, "Using wikis for collaborative learning: Assessing collaboration through contribution," *Australasian Journal of Educational Technology*, vol. 26, no. 3, pp. 341–354, 2010.

[22] J. Andriessen, M. Baker, and D. Suthers, *Arguing to learn: Confronting cognitions in "CSCL" environments*. Kluwer, 2003.

[23] D. Jonassen and S. Land, *Theoretical foundations of learning environments*. Routledge, 2012.

[24] K. Hew and W. Cheung, *Student Participation in Online Discussions*. Springer, 2012.

[25] F. Wang and M. Hannafin, "Design-based research and technology-enhanced learning environments," *Educational technology research and development*, vol. 53, no. 4, pp. 5–23, 2005.