Characterizing Software Engineering Students' Discussions during Peer Instruction: Opportunities for Learning and Implications for Teaching*

TOM ADAWI
Chalmers University of Technology, Gothenburg, Sweden. E-mail: adawi@chalmers.se

HÅKAN BURDEN
Viktoria Swedish ICT, Gothenburg, Sweden. E-mail: hakan.burden@viktoria.se

DENNIS OLSSON
University of Gothenburg, Gothenburg, Sweden. E-mail: dennis.olsson.1984@gmail.com

RICKARD MATTIASSON
University of Gothenburg, Gothenburg, Sweden. E-mail: rickard.mattiasson@gmail.com

Peer instruction is a method for activating students during lectures, which has gained a considerable amount of attention in higher education due to claims of dramatic improvement in learning gains. The purpose of this qualitative research study is to investigate what types of discussions engineering students engage in during a peer instruction session and what learning possibilities that are enabled by these different types of discussions. We observed twelve students during six separate and simulated peer instruction sessions and the students were interviewed individually after the sessions. An analysis of the data revealed that the students engaged in three qualitatively different types of discussions: *affirmative* discussions, *motivating* discussions, and *argumentative* discussions. We characterize these different types of discussions in terms of the number of alternative answers the students discuss, the extent to which they draw on prior knowledge and experiences, as well as the fundamental difference between an explanation and an argument. A good opportunity for learning is opened up when students are aspiring to find the truth, not simply being satisfied with what they believe to be true. We conclude that students do not always engage in discussions that support their learning in the best way, and we discuss implications for using peer instruction as a teaching method.

Keywords: peer instruction; learning possibilities; argumentation; software designs; UML diagrams

1. Introduction

The process of understanding a problem and assessing the suitability of a solution involves the forming of a cognitive model of both the problem and the solution spaces [1]. This cognitive model evolves as the solution is implemented and validated, but also through the interactions with other stakeholders. Such interactions are important since they enable new interpretations and knowledge to be merged with the initial cognitive model. One way to challenge students to form and refine their cognitive models, or their conceptual understanding, is through *peer instruction* during lectures.

Mazur developed and popularized peer instruction as a method to activate students during lectures [2]. After a brief lecture (around 15 minutes), he gives the students a multiple-choice question. The students first answer the question individually (using clickers or flashcards), then discuss their answer with a fellow student (a peer), and finally answer the question again individually. Peer instruction has gained a considerable amount of attention due to claims of dramatic improvement in learning gains [3].

Crouch and Mazur found that very few students who had a correct answer before the peer discussion changed to an incorrect answer [4]. This implies that if a student with an incorrect answer is paired with a student with the correct answer, the student with the correct answer is often successful in convincing the one with the incorrect answer to change his/her answer to the correct one. This could mean that the student with the incorrect answer has learned something new based on his/her peer's arguments and therefore changed his/her answer. But it could also mean that the student with the right answer is just better at convincing the other student to change his/her answer based on other factors than strong arguments (for example, a student who is usually right or very confident in his/her answer). To distinguish between these two alternatives, Smith and colleagues designed an experiment where they, directly after the students had re-voted on the first question, asked them a similar, or *isomorphic*, question [5]. Not only did they find that the number of correct answers to question one increased after the discussion by an average of 20%, they also found that the number of correct answers to the isomorphic question was 21% higher

than the number of correct answers to the first question.

What the study by Smith and colleagues, which was published in *Science*, does not explain is what actually happens during the interaction between the students and how this influences their conceptual understanding of the subject. In other words, the students are treated as "black-boxes". In a response to the findings by Smith and colleagues, Knight and colleagues conducted a study where 83 student discussions were recorded, transcribed and analysed in terms of to what extent the students reasoned during peer instruction [6]. They found that the students' discussions involved the exchange of claims, reasoning based on evidence and warrants—what Toulmin [7] defines as statements linking claims to supporting evidence—and that this had a positive impact on their learning. To better understand the nature of the reasoning, the types of interaction that goes on between two students while they reason, we conducted a qualitative research study [8], seeking the answer to three research questions:

1. What type of reasoning can engineering students engage in during a peer instruction session?
2. What learning possibilities can these different types of reasoning enable?
3. What are the implications for using peer instruction as a teaching method?

In line with the qualitative nature of our study, the aim is neither to be exhaustive nor to count frequencies of reasoning types but to capture the various ways students can interact with each other while discussing multiple choice questions and the impact this has on the learning opportunities.

The rest of the article is organised as follows. In section 2, we describe the research setting, including the methods for data collection and data analysis. Section 3 contains the results, three different types of reasoning—affirmative, motivating and argumentative. In section 4, we discuss our findings in terms of opportunities for learning, impact on teaching and the relationship between reasoning and software design. We then conclude and propose future strands of research in relation to argumentative discussions and software designs in section 5.

2. Method

As the purpose of this study is to investigate how students discuss the multiple-choice questions during a peer instruction session and what possibilities for learning these discussions enable, a *qualitative* research approach was used, with a combination of data collection strategies [8–10].

Twelve third-year software engineering students volunteered to participate in the study. The students were put into six groups to simulate a peer instruction session. This session did not take place in a lecture hall surrounded by other students, but in a smaller room with one of the authors as an observer. The reason was that we wanted to be better able to monitor and record the students when reasoning and interacting and therefore chose a more controlled setting.

The students were divided into six pairs based on their own preference, just as they would sit next to their preferred peer in a lecture hall and chose whom to discuss with. There is no focus on the students' individual background or other contextual factors such as age, gender or ambition. For this study it was most important that they represent the same level of education, bachelor students in computer science related programmes, currently taking a class in model-driven software development [11]. Moreover, the questions were relevant to what the students had been taught during class in their current course.

The multiple-choice questions that the students discussed focused on software designs represented as, sometimes simplified, UML diagrams [12]. These questions were developed by Stikkolorum et al. [13]. For the purpose of this study we chose two sets of questions containing a pair of isomorphic questions each. One of the sets focused on the interpretation of a directed association and included the question presented in Fig. 1. This question was complemented by an isomorphic question where *reuse* was replaced by *change*. In the second set, the questions concerned the principle of high cohesion and low coupling [14]. For this purpose a question on how to organise the responsibilities of a modem, detailed by Stikkolorum et al. [13], was used together with an isomorphic question where the syntactic structure of the candidate solutions were the same but the domain instead was an audio system. We varied the ordering between the isomorphic questions from discussion to discussion. By changing the questions and generating more
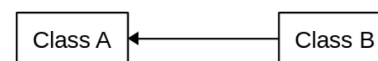
If we want to reuse Class B we:
(Check **one** box)

☐ 1. have to change Class A
☐ 2. also need to use Class A
☐ 3. do not need to use Class A
☐ 4. need to inherit from A

Fig. 1. An example of a multiple-choice question used in this study.

Fig. 2. The design used to investigate the peer discussions, where Q1ad is the second opportunity to answer Q1, and Q2 is isomorphic to Q1.

varied discussions, we gained a better understanding of what types of discussions that can occur.

Our sequence of posing questions replicates the one used in the study by Smith et al. [5]. The first question, Q1 (see Fig. 2), was answered individually without collaboration to test the students' individual conceptual understanding of software design. When both students had answered they were asked to discuss their answers. They were then asked to answer the first question again, Q1ad (ad is an abbreviation for "after discussion), with the opportunity to revise and change their answers. Finally, the students individually answered the isomorphic question, Q2, on the same topic but in a new context. The reason for the second question is to see if the discussions between Q1 and Q1ad had a positive or negative impact on the students' conceptual understanding.

The empirical data for this study consisted of: (1) the students' answers to the sequence of questions (collected on separate sheets of paper); (2) audio recordings of the peer discussions; and (3) recordings from semi-structured interviews with individual students directly after each session. We decided that the positive aspects of individual interviews outweighed those of group, or pair interviews, since we wanted to know the individuals thoughts regarding the discussion and its implications. Allowing for peer influence here would mean that the same phenomena that influence the individual's decision can have an impact on how a student chooses to answer the interview questions. This also gave us the possibility to contrast the answers from the interviews with what was said during the discussions.

The recordings of the peer discussions and the individual interviews were transcribed verbatim after the event. The transcripts were analysed using a *general inductive approach*. The aim of this kind of qualitative data analysis is to identify and describe patterns or themes that emerge from and cut across the data. During the data analysis, preliminary results were discussed among all the authors of this paper on a regular basis to strengthen the reliability of the results [9].

As described in the introduction, the study by Smith and colleagues uses the isomorphic question to evaluate what kind of learning the students have achieved through their discussions [5]. Our objective is instead to explore what type of discussions the students engage in. Initially, we also wanted to see if

there were any correlations between types of discussions and types of learning. While we found different types of discussions we could not map these onto different types of learning outcomes. There were two main reasons for this; what we, as teachers, see as an isomorphic question does not seem to be isomorphic from a student perspective and there were too few discussions with respect to types to cover all permutations in terms of order of the isomorphic questions. Including more student pairs into the study was not an option due to time constraints. So choosing between focusing on what happened during the discussions and the relationship between discussion and learning types, we decided to focus on how students reason during peer instruction. The different types of discussions that we identified are presented in the next section. We are, however, still interested in what opportunities for learning that these different types of discussions enable and what the implications are for using peer instruction as a teaching method, so these are topics we turn to in the discussion section.

3. Results

An analysis of the data resulted in the identification of three qualitatively different types of discussions. In order to capture the educationally critical aspects of these different types of discussions, we found it fruitful to characterize a certain type of discussion along three dimensions: *range*, *resource* and *reasoning*. Range denotes the extent to which the students discuss alternative answers; resource denotes the extent to which students draw on prior knowledge and experiences during the discussion; and reasoning denotes the level at which the students support their statements. For this last dimension we have drawn on the distinction between an *explanation* and an *argument*: "To offer an explanation of a fact is presuming it is true. An argument, in contrast, is an attempt to establish truth" [15]. Worth noting here is that Toulmin defines argumentation as making a claim, supporting the claim with evidence and ultimately supplying a warrant that links the claim to the evidence [7]. Knight et al. [6] follow Toulmin's definition and do not distinguish between explanatory and argumentative discussions.

In Table 1, we label the three types of discussions we identified— *affirmative*, *motivating* and *argumentative*—and characterize them along the three dimensions. What follows in this section is a descrip-

Table 1. Characterising the three types of discussions along three dimensions

Type	Range	Resource	Reasoning
Affirmative	One (or only few) candidate solution is discussed	No (or little) connection to prior experience and knowledge	Stating an answer
Motivating	Several candidate solutions are discussed	Prior experience and knowledge is used	Explaining an answer
Argumentative	All candidate solutions are discussed	Extensive use of prior experience and knowledge	Arguing for an answer

tion of the three types of discussions, supported with illustrative extracts from the peer discussions and the interviews.

3.1 Type 1—Affirmative discussions

The first type of discussion is characterized by a low number of alternatives being discussed as well as a low extent of experiential discussion. The students more or less simply state their answer with their peer agreeing. The discussion is over when the students have reached a consensus on what they believe to be the correct answer and they do not show any desire of discussing the question further. This type of discussion therefore tends to be much shorter than the other types: 1–2 minutes long, in comparison to 4–6 minutes for Type 2 and Type 3.

Half of the groups, three out of six, engaged in this type of discussion, and these students picked the same alternative, as illustrated in Table 2. Two groups picked a wrong (W) alternative and one group picked the right (R) alternative. The students did not know if they had picked the correct alternative or not, but they knew that they had picked the same alternative.

The following excerpt is an example of what the discussion looks like in the case of an affirmative discussion:

A1　The tricky part as I see it is the word "Reuse".
A2　Yes.
A1　What is meant by that really? If you should use class B in itself . . .
A2　Or if you should subclass it.
A1　Yes, exactly. Or change it . . .
A2　Or use it somewhere else.
A1　Mm, that's the most unclear [part of the question] since it's not a word we normally use. But as I interpret it, we want to use class B as it is. And since class B is defined as an inheritance of A then the answer, for me anyway, is that it must inherit from A.
A2　Mm, I had the same reasoning.

Here, the focus of the discussion in this pair, which we will call A1A2, is confined to support as the first student explains his/her choice of alternative and the other student supports this statement, clearly supporting each other's theories and explanations. There are no alternative interpretations of the answer. When both students are done comparing and confirming each other's answers, the discussion is treated as completed.

The excerpts below are from two different interviews with students from two different Type 1 discussions. The answers are characteristic for students engaging in Type 1 discussions and show that the students themselves used the discussion only to confirm their own theory. The question was: Did you find the discussion stimulating?

A2　It was good for making me more certain, I interpreted the arrow in the right way. So it's nice to have a discussion to... Maybe the discussion would have given more if it was a question where we had different answers in the beginning and we could discuss why one alternative would be better than the other. But still, one becomes more certain having a discussion and agreeing.

Student A3 answered the same question in the following way:

A3　It was quite brief; we had both picked the same answer so it was not exactly a discussion. More of agreeing on our choice. So we did not actually have any arguments. It was quite brief. I think that perhaps with such an easy question you don't really need a big discussion about it.

These excerpts support the classification of Type 1 discussions as narrow and limited to one or few alternatives being discussed. The students themselves even exhibit uncertainty as to whether their discussions are discussions at all, as they know they have the same answer and there is no one to question their theory. This shows that the students are satisfied with confirming their answer with each other and neglect to discuss alternative answers or principles. The first excerpt also shows that the

Table 2. Outcomes for groups engaging in affirmative discussions

Participants	Q1	Q1ad	Q2
A1 + A2	W + W (Fig. 1)	W + W (Fig. 1)	R + R
A3 + A4	W + W (Fig. 1)	W + W (Fig. 1)	R + R
B1 + B2	R + R	R + R	W + W (Fig. 1)

students express a confidence boost, from Q1 to Q1ad, by talking to their peer about the question and their answers.

There are occasions in all Type 1 discussions where one of the students tries to introduce more abstract explanations or arguments in the discussion. Here is one example:

B3   But I'm unsure of the second alternative, I want to use that one to kind of, but I don't know if it means the same thing, I don't know.
B4   Uh... [pause] But the . . . no.
B3   [Quick response, interrupting B4] But no, you don't need to. It's just that B is the same as A but something else as well.
B4   Mm.
B3   That's right. [Pause] Well, then I agree with you.
B4   Well, all right then.

This excerpt shows how one of the students tries to broaden the discussion. But once they begin to, the discussion comes to a halt and the question is not being taken any further, leading the discussion to an end. This example shows how Type 1 discussions are characterized by short discussions where alternative answers or principles do not seem necessary to discuss. Rather, it seems that the main objective is to find the right answer and to reach an agreement on that answer.

3.2 Type 2—Motivating discussions

The second type of discussion is far more extensive in time and content than Type 1. During Type 2 discussions, most of the candidate solutions are taken into account. The students also engage in a discussion that allows them to bring in a variety of resources, such as previous knowledge and experiences. They compare answers and explain why their choice is right and why the other alternatives are wrong before coming to a consensus and ending the discussion. However, the students also tend to agree with each other on the majority of the statements being made. The explanations are either new explanations or additions to the explanations by their peers. There are very few counter arguments that could provide an alternative perspective on the subject. One group in our study engaged in a Type 2 discussion and both students in this group picked the same incorrect alternative for Q1, see Table 3.

The following excerpt illustrates the difference between Type 1 and Type 2 discussions, where the students now continue to discuss after the initial stating and agreeing:

C2   So, we believe design B . . .

C1   Yes.
C2   What did you think?
C1   Uhm, well I went a bit on intuition. I dismissed [alternative] C direct generally.
C2   Mm, yes.
C1   Because it doesn't make sense to have one responsibility per class, a responsibility that tight.
C2   No, then it becomes too many classes.
C1   Yes, exactly. It would be . . . absurd, especially when some things, thinking of a responsibility as send and receive, are very related. Then I think it would become too much.

Here, there is an initial consensus but rather than just simply stating the explanation to the answer, like in Type 1 discussions, the students continue the discussion. C2 is asking for a motivation to the answer and thus the discussion is lead on to a level of explanation and argumentation. As another example of a Type 2 discussion, consider the following excerpt where the students are discussing the coupling and cohesion of the modem:

C2   And A is a bit . . . well everything in the same class, we have learned that you shouldn't have that.
C1   Yes, exactly. Well . . . I pondered a bit about it that . . . it wouldn't be that terrible anyway.
C2   No, it's a pretty small system actually.
C1   Yes, exactly. Because if one were to consider it [alternative A] to be a part of a bigger system if you would expand with more functions then I think B is much better than A in that case. Because A would be infinitely long almost.
C2   Well yes, you have to make some levelling there. You don't want too many distributions, not too few either.
C1   Yes, exactly.

This is an example of how the students take different alternatives into account. They also widen the discussion with hypothetical comparisons and bring in outside experience to the discussion. The two excerpts above are typical for Type 2 discussions in the sense that the students express the same opinion and view. There are no counter arguments in the discussion, making it a discussion without new perspectives.

The following excerpt from the interview with a student from a Type 2 discussion supports our characterisation of this type of discussion. The question was: Did you find the discussion stimulating?

C1   Yeah, yeah I think so. Discussing these types of programming patterns is very helpful to develop your understanding and often discussions have helped me from believing one practice is the best to believe another practice is the best, and I think

Table 3. Outcomes for groups engaging in motivating discussions

Participants	Q1	Q1ad	Q2
C1 + C2	W + W (see [4])	W + W (see [4])	R + R

that's pretty general thing . . . if you can come up with enough arguments you can basically convert anyone to your way of thinking. So I think such discussions are very helpful, especially when it comes to programming where there's no one specific solution that is correct, just one that is neater, I guess. 'Cause it's not so much about facts, it's more about . . . It's subjective really.

Students who engaged in Type 2 discussions viewed the discussion as an opportunity to explain different alternatives in order to reach a correct answer, and therefore also discussed most of the other alternatives to exclude them.

3.3 Type 3—Argumentative discussions

Here, the students take the discussion to a more abstract level. The students also use counter arguments as they try to reinforce their answers and prove their peer wrong. They use previous knowledge and external experiences to support their arguments and show that the principles are applicable outside of the context of the question and situation.

Two groups engaged in a Type 3 discussion and both pairs had different answers on Q1 before the discussion, see Table 4. But in one group, B3B4, a student changed his/her alternative from Q1 to Q1ad and in the other group, C3C4, neither of the students changed their answer from Q1 to Q1ad. In the individual interview it is clear that, even if none of the students in B3B4 changed their answer, one of the students, B3, reconsiders which alternatives might be true. The discussions of this type begin with argumentation to why they have chosen their respective answers.

As an example of how the argumentation plays out in a Type 3 discussion, consider the following excerpt where the students in group B3B4 are discussing the isomorphic question to Fig. 1:

B3 But I don't think you need to inherit from A. If you change something in class B then why would you need to inherit something from A, you already inherit from A, I think.

B4 Yes, well I don't need to inherit something new. But if I subclass something I must inherit what's in there. It depends on what it [pointing to arrow?] refers to, if it . . .

B3 Yes, exactly . . .

B4 If it refers to the fact that I must have that exact behaviour or... because class B inherits automatically.

B3 But it's not the arrow we investigate. That means

that . . . if you draw an arrow from B to A, then it means that we inherit it.

B4 Mm.

B3 But the question is if we change class B . . .

B4 Yes, that's right.

B3 So then it's clearly answer 3.

B4 Yes, ok. Precisely, it's that structured.

Here the students use arguments and not explanations since they try to convince each other what the implications of the arrow are and then choose an alternative instead of the other way around. This can clearly be seen by the fact that it is at the end they first mention which alternative they should choose. Note how B4 starts this segment of the discussion with a counter argument to what B3 believes regarding how the to interpret the association. Here, B4 presents a new interpretation that is in conflict with what B3 said in his/her argument. This excerpt also provides a glimpse of how the discussion is more abstract than the Type 1 discussion, when he students bring in the more abstract term subclass to reason around the association between classes A and B.

As an example on how counter arguments affect the individual learning process, consider the following excerpt from the interview with B4. The question was: What made you change your answer?

B4 The discussion we had really made me realize that . . . why B was a subtype of A is kind of logical when you see B to A. So you know that B knows about A and that's the logical way to program it. That convinced me that [alternative 3] was the right answer.

Here, B4 does admit to succumbing to B3's answer, as it made him see the question in a logical way. This shows that students can influence their peers to see the subject from another point of view by using counter arguments. As an illustration of how both counter argumentation works and how they use abstract arguments in Type 3 discussions, consider the following excerpt from the discussion between C3 and C4 where they are discussing the modem question from the perspective of coupling and cohesion:

C3 It's for the most general case I believe that [model] one is better.

C4 Yes, if we want to expand . . .

C3 Yes, but maybe it's unnecessary to do that by some . . . Yes well, if one were to see it again . . . then it feels like a better . . .

C4 But how are they linked there? Because later on we must somehow, depend on "connection" I guess?

Table 4. Outcomes for groups engaging in argumentative discussions

Participants	Q1	Q1ad	Q2
B3 + B4	R + W	W + W	W + W (Fig. 1)
C3 + C4	W + R (see [4])	W + R (see [4])	R + R

C3   Hmm, well yes [pause]. But then A should still be lying in there somehow. That the modem in control determines how it is transmitted and then . . .

C4   But it's more, you can't really come to that there . . . or can you? Because it's more a have relationship? [pointing to the class diagram] An arrow like that?

C3   Hmm, that's right.

C4   The modem doesn't have any public . . .

C3   Well maybe it is . . .

C4   It doesn't have any public methods whatsoever. [pause]

C4   If you were to code this it would be interesting, because then you would have needed, then you need some kind of interface towards the others.

C3   Mm, I believe I thought more of heritage so then you could access... It could just be that it's only have [relationship], just a separation of the code then.

Here, the students present different principles and try to draw on outside experiences to support their arguments. They use examples from other contexts than the question itself, which hints that the dialogue is being driven by an inquiry to understand each other's point of view. The students in this group seem to understand each other's point of view but do not share it due to a lack of convincing arguments. The following excerpt shows an example of how the students perceived the discussion. It's from the interview with C3, who is answering the question: What are your general thoughts about the discussion?

C3   It was good to have I guess, you get a new perspective on the question and though it did not change my final answer my choices were changed between the questions so I had a different mind-set but in the end I choose the same.

Here, the student claims that even though he/she did not change answer from Q1 to Q1ad he/she gained a new perspective through the discussion. C4, on the other hand, did not perceive the discussion with C3 as stimulating, but instead brings up an important aspect of the questions in the interview. Consider the following excerpt were C4 answers the question: Have you had any good experience of when questions to discuss in the classroom have led to good discussions?

C4   It probably depends on the subject. If it's an interesting subject then it's a good idea, but for the teachers it's hard to know how the students think about the subject, are they interested or are they just there for . . .

This student points out that some topics are just not interesting enough in themselves to create good discussions, and some students do not have any interest in learning the subject for that particular lecture.

4. Discussion

We have identified and characterized three qualitatively different types of discussions that students engage in during a peer instruction session. In this section, we turn to our second and third research question, and discuss what learning possibilities that each type of discussion enables as well as implications for teaching and reasoning about software designs.

4.1 Opportunities for learning

In Type 1 or affirmative discussions, the students agree with each other and quickly come to a conclusion. In these discussions there is no real argumentation taking place, only a short series of explanations as to why they have chosen the "correct" alternative and the discussion is over very quickly. The resources that the students bring into the discussion and can learn from are their own explanation regarding why their alternative is correct and their peer's explanation as to why the alternative is correct. To explain why the alternative is right, the students need to formulate their explanations in their own words. So even if they have the same explanation they may have two different formulations of it [16, p.22]. In the process of turning thoughts into words, the students need to structure their thoughts and they might see their own explanation in a new perspective. In general, when a student turns his/her thoughts into words it seems to make the thought clearer. It also seems to make the thought more open for self-critique. (This phenomenon can be observed in every classroom where the teacher only needs to ask the students to explain their thinking to make the students see and correct their own mistake). But we have in our study no data of any students engaging in Type 1 discussion changing a wrong answer to the correct one. This would have been an indicator of a student correcting a mistake in their explanation. The results show that students tend to stick to their alternative and are more comfortable with their choice after the discussion. This is an indicator that Type 1 discussions are incapable of challenging misconceptions. However, the learning possibilities the students have when engaging in this type of discussion is a broadening and strengthening of previous beliefs. This happens regardless if the answer is correct or not, and can therefore strengthen misconceptions.

Our results show that Type 2 or motivating discussions start the same way as Type 1 discussions, with acknowledging the other student's reasoning and agreeing with each other. But then the students continue to discuss the other alternatives, offer explanations as to why they are wrong and

some of the discussions contain references to more general principles. To try to explain why the other alternatives are wrong is good since there is as much value in knowing what is wrong and why, as there is in knowing what is right and why [15]. By widening the discussion to include more explanations and general principles the discussion offers better possibilities for learning. Firstly, the students have a greater chance to get familiar with arguments that they did not know before and to hear more arguments explained in different ways, an experience also discussed by Knight et al. [6]. This process of broadening and strengthening previous beliefs is the same as in Type 1 discussions. But since there are now more explanations in this type of discussion the students have more material to draw on to broaden previous beliefs. Secondly, having both more explanations and general principles makes it easier to construct their own principles, which help them transfer their learning to other contexts. In other words, by taking what is common in the explanations and make sure they do not contradict each other they can broaden their previous beliefs about the subject. This will make the subject easier to use outside the context of the question [17, p.68].

In the interviews, all students engaging in Type 1 and Type 2 discussions said that they kept their initial answers since their peer confirmed their ideas. The reason the students from these types of discussion kept their answer, and even gained more confidence in their answer, is that their choice was confirmed by another person. The desired reason to keep or change the answer should rather depend on the quality of the argument. The students excluded the other alternatives when they answered it individually but not in the discussion. There lies a danger in these types of discussions. When the students acquire new explanations to their false beliefs they also gain more confidence that it is a correct belief and therefore have less motivation to question it and correct their mistake. These discussions do not seem to aspire for a quest to find the true answer, but rather a quest to pick the "correct" answer and give an explanation as to why it is correct. And the correct answer in this case seems to be based on something different than the strength of the argument: the "correct" answer is found when the students come to a consensus.

Type 3 or argumentative discussions offer an opportunity for the students to see the question from another point of view and hence offers a possibility for conceptual change. What mainly distinguishes Type 3 discussions from Type 2 discussions is the use of counter arguments. There is a shift in focus here from explaining their answer to arguing which alternative is right. This shift is important for what learning possibilities the discus-

sion holds. Osborne, for example, points out that argumentation plays an important role in science education since it seems to be a core process in learning, thinking and creating new understanding. He also writes that arguments containing counter-arguments and rebuttals have the greatest value, "as they require the ability to compare, contrast and distinguish different lines of reasoning" [15, p. 464]. This means that the possibilities for learning are much better in Type 3 discussions. This type of discussion has the same richness in terms of content as Type 2 discussions and thus the same potential to broaden and strengthen previous beliefs, but the use of counter arguments also opens up for the possibility for the students to reassess previous beliefs and see the subject from a new perspective. With this follows the possibility to spot and correct misconceptions.

4.2 Implications for teaching

Our results validate that students do not always discuss the question in a way that helps them to learn [18–20]. Instead, reasoning seems to be a skill that has to be explicitly taught and used in practice by students [6, 20–22]. Teachers need to encourage students to discuss the questions in a way that includes argumentation and different perspectives on the problem. How this could be done more specifically is an area for another study, but we can offer some thoughts on what might work. One strategy, also used by Mazur, is to encourage students to discuss with a partner who has picked a different alternative. Another strategy is to encourage students to discuss all of the alternatives and try to defend them, instead of just choosing the one they think is right and coming up with an explanation as to why it is right. This requires that teachers explain the difference between an argument and an explanation.

We have seen that the discussion can actually strengthen the students' misconceptions. Teachers need to address this after the discussion. One way to do this, is to not only explain why the correct answer is correct, but also why the other alternatives are wrong. Another way is to do a different kind of assessment in order to identify and correct student misconceptions.

Illeris [17] emphasizes that intrinsic motivation, or wanting to learn, is also important. It was clear from the interviews that the students have experienced discussions where they did not discuss at all. If the discussions are going to be of good use for learning, there must at least be one student in the discussion who is there with an intention to use the discussion to learn. This can be triggered when the students have chosen different alternatives. But the problem with using multiple-choice questions is that

the teacher cannot force the students to answer the way he/she wants to, and even if they do, the discussion might not reach its full potential. In some cases it might be better to use more open questions with a greater number of answers, and maybe more than one correct one, to increase the learning potential of the discussion. What the teacher looses is the opportunity to choose which alternative answers the students should discuss, usually common misconceptions, and the opportunity to collect quantitative data. The teacher must, as always, balance the gains and losses against each other and choose the most effective strategy in that particular situation.

4.3 Abstract reasoning and software designs

One set of questions concerned the relationship between two classes and what would happen if class B, unidirectionally related to class A, should be changed or reused, respectively. For these questions the students were confused about the interpretation of the arrow, repeatedly taking it to mean inheritance. While the students use the appropriate terminology they use it for the wrong concepts. And the ensuing discussion only strengthened their misconception. An example of this is when student A2 claims that the discussion was fruitful since it confirmed that "I interpreted the arrow in the right way". This resonates with earlier critique of the UML in terms of size and complexity [23]—the students have difficulties in remembering the semantics of the different arrowheads.

The other two questions involved the notions of cohesion and coupling. Here the students in discussion C1C2 did not use the right terminology, an indication of low abstract reasoning, but still got the right answer after a motivating discussion. While the pair struggled with the terminology they still showed ability for abstract reasoning when discussing if the example in the question was large enough for the idea of high cohesion and low coupling to be applicable. The other student pair, C3C4, that reasoned around cohesion and coupling where also helped by their combined reasoning, this time from an argumentative discussion. Also C3 and C4 had problems finding the right terminology but demonstrated abstract reasoning skills when trying to relate the information conveyed by the models to source code. While our samples are too small for a quantitative analysis, the student answers imply that domain knowledge is important for assessing the suitability of candidate designs. In our case the students struggled with the question of coupling and cohesion for the modem but got it right for the audio system.

Kramer argues that students' capability of interpreting designs depends on their ability to reason about abstractions [24]. This is in contrast to the findings of Bennedsen and Caspersen who studied abstractions as an indicator for students' ability to learn software concepts [25]. Our own results point in both directions. While it seems that discussions containing a more abstract reasoning tend to help student pairs to get the right answer, the students in our study also found the problems concerning coupling and cohesion easier to reason about. This raises the question whether the nature of the problems enable or mitigate abstract reasoning.

5. Conclusions and future research

Based on previous research on student discussions, we designed a qualitative research study to answer three questions: (1) What types of discussions can engineering students engage in during a peer instruction session? (2) What learning possibilities can these different types of discussions enable? and (3) What are the implications for using peer instruction as a teaching method? Here, we briefly conclude in relation to each of these questions.

What type of discussions can engineering students engage in during a peer instruction session? We have identified and characterized three qualitatively different types of discussions: affirmative discussions, motivating discussions and argumentative discussions. The affirmative discussion (Type 1) is characterized by a stating of answers and quick agreement. The motivating discussion (Type 2) is more extensive in time and wider in the sense that more alternatives are discussed. Here, the students explain why they have picked their answer and explain why they have excluded the other alternatives. Argumentative discussions (Type 3) contain a more advanced level of support as they include counter arguments.

What learning possibilities can these different types of discussions enable? The learning possibilities vary for the three different types of discussions. During Type 1 discussions, students may broaden and fortify their previous beliefs. During Type 2 discussions, the learning possibilities are increased by the inclusion of principles and contextual references. Type 3 discussions offer an opportunity to see the question from a different point of view. Here, the use of counter arguments opens up for the possibility for the students to reassess their previous beliefs and hence to spot and correct misconceptions.

What are the implications for using peer instruction as a teaching method? Teachers need to encourage students to discuss the questions in a way that includes argumentation and different perspectives on the problem. One strategy is to encourage students to discuss with a peer who has picked a different alternative. In some cases it might be better

to use more open questions with greater number of answers, and maybe more than one correct one, to increase the learning potential of the discussion. In practice, this means that both the way teachers chose to teach reasoning and the nature of the problems students are asked to reason about will have an impact on the learning opportunities.

We still don't know how to ensure that peer discussions are fruitful. There is a need for a follow up study with a larger student group to explore when students engage in argumentative discussions, which seem to hold the best learning possibilities; a study that investigates circumstances that lead students to argue for their answers and what teachers can do to create such learning environments. Another promising line of enquiry is the concept of isomorphic questions and which problems students find more challenging in terms of UML notations and problem domains.

References

1. R. France and B. Rumpe, Model-driven Development of Complex Software: A Research Roadmap, *2007 Future of Software Engineering (FOSE '07)*, Washington, DC, USA, 2007, pp. 37–54.
2. E. Mazur, *Peer Instruction: A User's Manual Series in Educational Innovation*, Prentice Hall, Upper Saddle River, NJ, 1997.
3. E. Mazur, Farewell, Lecture?, *Science*, **323**, 2009, pp. 50–51.
4. C. H. Crouch and E. Mazur, Peer Instruction: Ten Years of Experience and Results, *American Journal of Physics*, **69**, 2001, pp. 970–977.
5. M. K. Smith, W. B. Wood, W. K. Adams, C. Wieman, J. K. Knight, N. Guild and T. T. Su, Why Peer Discussion Improves Student Performance on In-Class Concept Questions, *Science*, **323**, 2009, pp. 122–124.
6. J. K. Knight, S. B. Wise and K. M. Southard, Understanding Clicker Discussions: Student Reasoning and the Impact of Instructional Cues, *CBE Life Science Education*, **12**, 2013, pp. 645–654.
7. S. E. Toulmin, *The Uses of Argument*, Cambridge University Press, Cambridge, UK, 1958.
8. J. M. Case and G. Light, Emerging Methodologies in Engineering Education Research, *Journal of Engineering Education*, **100**(1), 2011, pp. 186–210.
9. C. B. Seaman, Qualitative methods in empirical studies of software engineering, *IEEE Transactions on Software Engineering*, **25**(4), 1999, pp. 557–572.
10. P. Runeson and M. Höst, Guidelines for conducting and reporting case study research in software engineering, *Empirical software engineering*, **14**(2), 2009, pp. 131–164.
11. H. Burden, R. Heldal and T. Siljamäki, Executable and Translatable UML—How Difficult Can It Be? In *APSEC 2011, 18th Asia-Pacific Software Engineering Conference*, Ho Chi Minh City, Vietnam, 2011.
12. M. Fowler, *UML distilled: a brief guide to the standard object modeling language*, Pearson Education, Inc, Boston, MA, USA, 2003.
13. D. R. Stikkolorum, C. Stevenson and M. R. V. Chaudron, Assessing Software Design Skills and their Relation with Reasoning Skills, *EduSymp@ MoDELS*, Miami, FL, USA, 2013.
14. W. P. Stevens, G. J. Myers and L. L. Constantine, Structured design, *IBM Systems Journal*, **13**(2), 1974, pp. 115–139.
15. J. Osborne, Arguing to Learn in Science: The Role of Collaborative, Critical Discourse, *Science*, **328**, 2010, pp. 463–466.
16. R. Säljö, *Lärande i praktiken: ett sociokulturellt perspektiv*, Norstedts, Stockholm, Sweden, 2010.
17. K. Illeris *Lärande*, Studentlitteratur, Lund, 2007.
18. D. Kuhn, *The Skills of Argument*, Cambridge University Press, Cambridge, UK, 1991.
19. D. Kuhn and W. Udell, The development of argument skills, *Child Development*, **74**, 2003, pp. 1245–1260.
20. A. Zohar and F. Nemet, Fostering students' knowledge and argumentation skills through dilemmas in human genetics, *Journal of Research in Science Teaching*, **39**, 2002, pp. 35–62.
21. K. L. McNeill, D. J. Lizotte, J. Krajcik and R. W. Marx, Supporting students' construction of scientific explanations by fading scaffolds in instructional materials, *Journal of Research in Science Teaching*, **15**, 2006, pp. 153–191.
22. J. Osborne, S. Erduran and S. Simon, Enhancing the quality of argumentation in school science, *Journal of Research in Science Teaching*, **41**, 2004, pp. 994–1020.
23. R. France, S. Ghosh, T. Dinh-Trong and A. Solberg, Model-driven development using UML 2.0: promises and pitfalls, *Computer*, **39**(2), IEEE, 2006, pp. 59–66.
24. J. Kramer, Is abstraction the key to computing? *Communications of the ACM*, **50**(4), April 2007, pp. 36–42.
25. J. Bennedsen and M. E. Caspersen, Abstraction ability as an indicator of success for learning computing science? In *Proceedings of the Fourth International Workshop on Computing Education Research, ICER'08*, ACM, New York, NY, 2008, pp. 15–26.

Tom Adawi is Professor in Engineering Education Research (EER) and is heading the Division of EER at Chalmers University of Technology. He is the Chair of the Nordic Network of Engineering Education Research (NNEER) and his research interests include students' understanding of threshold concepts in science and engineering, the interplay between and the development of conceptual/mathematical understanding, the scholarship of teaching and learning, as well as theoretical and methodological issues in EER.

Håkan Burden is a senior researcher at Viktoria Swedish ICT. He received his PhD in Computer Science in 2014 at the Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg. Current research focuses on facilitating innovation within ICT and embedded systems. Examples of on-going projects include developing business models and tools for enabling the participation of third-party developers in automotive software development, exploring the possibilities and challenges regarding team teaching as well as student-industry collaborations involving safe and connected transportation.

Dennis Olsson is a Master's student in Education at the University of Gothenburg, Sweden.

Rickard Mattiasson is a Master's student in Education at the University of Gothenburg, Sweden.