

Factors Influencing Quality of Team Discussion: Discourse Analysis in an Undergraduate Team-Based Learning Biology Course

Sarah M. Leupen,^{1*} Kerrie L. Kephart,² and Linda C. Hodges²

¹Department of Biological Sciences and ²Faculty Development Center, University of Maryland Baltimore County, Baltimore, MD 21250

ABSTRACT

Group activities as part of active-learning pedagogies are thought to be effective in promoting student learning in part because of the quality of discussion they engender in student teams. Not much is known, however, about which instructional factors are most important in achieving productive conversation or how these factors may differ among different collaborative pedagogies. We explored what provokes meaningful group discussions in a university physiology course taught using team-based learning (TBL). We were most interested in discussions that evoke explanations that go beyond statements of basic facts and into disciplinary reasoning. Using transcribed conversations of four randomly selected teams three times throughout the semester, we analyzed three distinct discursive phenomena—conceptual explanations, re-evaluations, and co-construction—that occurred in productive conversations. In this paper, we provide examples from student discussions showing the role of each of these elements in moving students toward conceptual understanding. These phenomena were more likely to occur in response to higher-order questions in Bloom's taxonomy. Preclass preparation and student accountability as part of TBL may be important factors in this finding. We share implications for practice based on our results.

INTRODUCTION

Cooperative and collaborative learning pedagogies have been shown to enhance student learning outcomes such as knowledge acquisition, problem-solving ability, and higher-level reasoning and to promote student engagement and persistence (Johnson *et al.*, 1998, 2006; Johnson and Johnson, 1999; Springer *et al.*, 1999). Such group learning approaches are thought to be effective in part because of the quality of discussion they stimulate among student teams—discussion that prompts students to think about and explain the course material to one another in ways that benefit both the students giving and the students receiving the explanations (e.g., Chi, 2009; Chi and Wylie, 2014). For example, studies in classes taught using peer instruction suggest that students benefit from discussion even if none of the discussants initially answers the question correctly (Smith *et al.*, 2009). Not much is known, however, about which instructional factors are most important in cultivating discussions that engender conceptual explanations. Nor is it understood how these factors may differ among different collaborative pedagogies.

Team-based learning (TBL) is a flipped-classroom collaborative pedagogy in which teams follow a specific sequence of activities throughout a unit (Michaelson, 2002). Each unit begins with an individual Readiness Assurance Test, a multiple-choice assessment in which students are held accountable for having gained specific learning objectives before class. Students then immediately retake the test as a team using a form of immediate feedback for each answer, most commonly the Immediate

Jennifer Knight, *Monitoring Editor*

Submitted Jun 11, 2019; Revised Dec 11, 2019;
Accepted Dec 24, 2019

CBE Life Sci Educ March 1, 2020 19:ar7

DOI:10.1187/cbe.19-06-0112

*Address correspondence to: Sarah M. Leupen
(leupen@umbc.edu).

© 2020 S. M. Leupen *et al.* CBE—Life Sciences Education © 2020 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

Feedback–Assessment Technique (IF-AT) scratch-off forms (Epstein Educational Enterprises, Inc.). Thus, at the end of this relatively brief process, all students have achieved basic learning objectives needed to wrestle with more complex questions in class.

The remainder of in-class time in the unit is usually spent having students apply their knowledge to answer complex, conceptual “application questions” on the course topic. To be successful in promoting student learning and discourage unrelated socializing, the process of asking and discussing application questions ideally meets four criteria: significant problem, same problem, specific choice, and simultaneous report. Problems should be as nontrivial and relevant to the real world as possible (significant problem), and teams should be asked to make a decision rather than answer an open-ended question (specific choice), promoting a more focused and productive discussion. Further, all teams should always be working on the same problem; otherwise, teams will be less motivated to be interested when others report. Finally, to avoid a collapse toward the apparent right answer, students should report their answers to application questions simultaneously. Simultaneous reporting may be accomplished through a variety of methods, from using clickers to a simple system of holding up voting cards on the instructor’s signal.

Although TBL has been demonstrated to improve content knowledge and retention (Haidet *et al.*, 2014; Kubitz, 2014; Swanson *et al.*, 2017), to our knowledge, no studies exist on how students engage with application questions in a TBL classroom. Work in other collaborative learning contexts, however, has been informative. Bierema *et al.* (2017) analyzed group discussions during modeling activities in an introductory biology course and found that team members commonly co-constructed their knowledge during modeling activities and engaged in sense-making interactions to build their understanding of the biological model. Similarly, Knight *et al.* (2013) found that the majority of student discussions of clicker questions in an upper-level biology class used evidence-based reasoning, and questions that requested or required student reasoning were more likely to produce it.

Theoretical Frameworks

We wanted to explore what provokes the sorts of explanations that we are most interested in seeing, that is, explanations that go beyond statements of basic facts and into disciplinary reasoning. Eliciting such reasoning is one of the major underlying motivations behind group learning pedagogies, as described by Repice *et al.* (2016) in their study of student discussions during peer-led team learning (PLTL). They noted that when solving problems in an introductory chemistry course that primarily involved calculations, data analysis, or model building, students’ statements that described pertinent concepts or real-world connections were the smallest proportion of their discourse. But these statements, which the authors called “conceptual explanations,” were the most important, because they involved students in making meaning.

How students engage during group learning can play an important role in whether they create conceptual explanations. Chi (2009) proposed a framework that links different levels of engagement in active learning with resulting student learning outcomes. The framework, ICAP (interactive, constructive, active, and passive), has been tested in subsequent research

studies (Chi and Wylie, 2014, and references therein). In this framework, passive behaviors (e.g., only paying attention) result in students receiving information, possibly storing it, but not integrating it into prior learning. *Active* engagement, as defined in this framework, entails receiving information and connecting it to prior knowledge through behaviors such as transcribing notes or rehearsing. In the ICAP framework, higher-order learning is triggered by what are called “constructive” or “interactive” behaviors that involve students in making inferences, either individually or collaboratively, respectively. Constructive actions such as explaining or debating ideas and posing or answering questions involve students in generating new understandings and making meaning. When students are interactive as well as constructive, taking turns and building on one another’s thoughts, they draw on the power of socially mediated learning to prod each other along paths in their thinking that they would otherwise not take.

A key factor in inciting such engagement in group learning activities is the design of questions to prompt discussion. The importance of question quality has been discussed by many other authors. Beatty *et al.* (2006) point out that the characteristics of good student response questions in class are very different from those of good homework or test questions, especially as part of what they called “question-driven instruction,” such as TBL, in which in-class questions are the engine of learning rather than a check on comprehension of a lecture segment. They recommend that instructors ask questions that force students into disciplinary thinking to arrive at an answer and avoid including distracting information in questions that could divert students from this goal. This advice echoes the common teaching suggestion to teach by leading students into a situation from which they can only escape by thinking. This principle emphasizes the importance of avoiding questions that students can answer by calculation or by looking up the answer without exploring concepts and testing their understanding. Wood *et al.* (2014) further demonstrated the importance of designing questions that activated students’ higher-level cognitive resources. They noted that a question’s wording, beyond just its topic, could push students toward conceptual thinking or, if poorly worded, allow them to resort to simple calculation, presumably without interactive modes of engagement.

Some authors have promoted specific question qualities as being instrumental to producing higher-level responses. Ertmer *et al.* (2011) found that focal questions, which require students to make a decision, were among the most effective question types in promoting responses at a high level of Bloom’s taxonomy in online discussions. Drawing on the ICAP framework, focal questions more likely encourage interactive debate. The kind of task that questions prompt students to undertake may also affect discussion quality. Repice *et al.* (2016) found in PLTL groups that a data-analysis prompt (as opposed to other kinds of prompts, especially those that focused primarily on calculations) guided students into a mode of discourse that led to conceptual explanations. This mode of discourse included students asking one another what the authors defined as “open questions,” which were those that tended to facilitate or draw out further elaboration of ideas rather than focus attention on a specific item of content or procedure. Given the importance of students talking through concepts to develop their understanding, these authors advocated for designing group and

problem-solving activities to involve students in asking each other open questions. In this reasoning, the question promotes deeper understanding indirectly by engendering interactive engagement that draws out conceptual explanations.

As students move from constructive to interactive engagement in groups, their conversation can extend from providing explanation to engaging in argument. Osborne and Patterson (2011) described explanation and argument as two different moves in discourse. An explanation gives a reason or accounts for something, whereas an argument probes whether an explanation is appropriate or improves on other options. Thus, questions that prompt students to engage interactively, re-evaluating one another's responses and weighing the merits of other possible choices, can promote better learning outcomes as theorized in the ICAP framework. One possible way to evaluate the ability of questions to achieve these outcomes is to categorize them according to Bloom's taxonomy of learning objectives, which has previously been successfully translated into the language of undergraduate biology (Crowe *et al.*, 2008).

Thus, our questions in this study were:

1. Did questions that demanded greater cognitive resources (i.e., questions posed at a higher level on Bloom's taxonomy) elicit a greater number of conceptual, discipline-specific (in this case, physiological) explanations than lower-level questions in students' small-group conversations?
2. How did the level of cognitive demand posed by a question relate to the level of interactive engagement among students in their responses? That is, did more cognitively demanding questions elicit co-construction of meaning more frequently among the student groups?
3. Under what circumstances did questions catalyze more re-evaluation among student teams—for example, more instances of argumentation such as debating, questioning explanations, asking new questions—potentially resulting in more interactive engagement?

METHODS

Context of the Study

This case study was conducted at a midsize public research university recently designated as a minority-serving institution. Human Physiology is an upper-level course taken by biology majors with an enrollment of 93. The class is taught in an active-learning classroom with large tables, screens and whiteboards accessible to each table, and no “front of the room” (see Supplemental Figure S1). In Fall semester 2017, students were divided into 15 teams of six (three teams had seven members) in such a manner as to evenly and transparently distribute among teams those students who had taken relevant additional course work (e.g., biochemistry, anatomy) or who had previously taken a TBL class. Teams worked together every class day throughout

the semester. The semester was divided into seven units, each of which began with the “readiness assurance” process described earlier. For the rest of the unit, class time was designed around short problems or questions, often clinical applications, that probed the most difficult learning goals for the unit. Some problems were open ended, but most followed the TBL “specific choice” format (Michaelson, 2002), in which student teams are asked to make a decision. Teams voted simultaneously on the answer to each problem or question following a countdown by the instructor, using a series of color-coded, lettered cards. Following voting, the instructor selected students to defend their teams' choices, calling on them at random by using a stack of cards with student names and photos, until enough students had provided responses so that the correct reasoning was clear, at which time the next application question was given.

Data Collection

Four of the 15 teams were chosen to have their conversations during class audio-recorded; the same four teams were recorded throughout the semester. They were selected on the basis of their location within the classroom in order to maximize their distance from one another in the room as well as their relative isolation from other teams, thus minimizing cross-recording of teams. This arbitrary selection resulted in a fairly representative cross-section of the class (see Table 1). Three times throughout the 14-week semester, each of the four selected teams had conversations recorded using a small digital Olympus audio recorder placed in the center of the team's table: one class session in week 3, one in week 8, and one in week 12. The audio recordings, which lasted ~75 minutes each (the length of a class period) were then transcribed. These transcripts provided the primary source of data for this study.

To supplement discourse analysis data, we also collected direct student learning outcome data. Because a high quality of student discussion in teams is thought to partially underlie the success of collaborative learning approaches in promoting student learning, we believed it was important to establish that students achieved learning outcomes both in class and across the semester. These secondary data sources included quizzes assessing conceptual understanding of unit concepts that were given at the beginning and end of each class period during which team conversations were recorded (Supplemental Files S1–S3), as well as a 20-question conceptual physiology test given to the class both on the first day of class and embedded into the final exam (Supplemental File S4). Each question on the conceptual physiology test assesses the achievement of an individual conceptual learning objective (from the list of objectives given to the students) from a different topic in the course (cardiovascular physiology, renal physiology, etc.) This test was written by the first author (S.M.L.) and is not an externally validated instrument. Changes in scores from pre- to posttest were evaluated

TABLE 1. Demographics of students in recorded teams and the class overall

	% Female	GPA at time of the study	%White	% Asian	%Black, Latinx +/-or multiracial
Recorded students (<i>n</i> = 22)	69	3.1	36	45	18 (4 of 22)
Entire class (<i>n</i> = 90)	66	3.2	30	41	26 (23 of 90)

using a repeated-measures *t* test. The study was classified by University of Maryland Baltimore County's Office of Research Protections and Compliance as exempt (IRB protocol Y18SL04023).

Data Analysis

The questions posed to the students during the three recorded class periods were categorized by each author into low (knowledge), medium (comprehension or simple application), or high cognitive demand level (complex application, analysis, or synthesis), following the guidelines for categorization of biological questions into Bloom's taxonomy described in Crowe *et al.* (2008). Discrepancies were resolved by consensus.

Following the principles of grounded theory (Glaser and Strauss, 1967; Strauss and Corbin, 1998), we conducted iterative readings of the transcripts to discern recurring patterns and functional types of contributions to the conversation. As with the categorization of questions, all three authors (S.M.L., K.L.K., and L.C.H.) qualitatively coded all 12 transcripts independently and then engaged in intensive group discussion to resolve discrepancies by consensus (Saldaña, 2015). Pseudonyms are used for student names in the transcript excerpts provided.

Our basic unit of analysis was the turn-at-talk, as defined in conversation analysis (Sacks *et al.*, 1974). Each turn, or parts within turns, could be coded as more than one functional type, as speakers' thinking might develop or change or their purposes might shift, sometimes multiple times within a single turn. Thus the function of their contribution within a single turn might change. Across all four teams whose conversations we analyzed, we discerned two functional types of contributions, "conceptual explanations" and "re-evaluations," and one other discursive phenomenon, "co-construction," that we found to be relevant to understanding how participants verbally grappled with the questions that were posed. Remaining uncoded talk in the transcripts included instances of students stating or asking about factual or procedural information; talking about the question at a surface, nonphysiological level; discussing another topic; or not talking to one another at all (see Excerpts 6 and 7).

While our use of the term "conceptual explanation" is similar to how Repice *et al.* (2016) use it to describe students making meaning and making sense around numbers, equations, and diagrams, our definition differs, mainly because of the differences in the disciplinary subject matter being discussed in our respective studies. We define conceptual explanations as stretches of talk in which one or more students in the group produced a hypothesis of physiological functioning that was not simply fact based but drew upon a conceptual understanding of the human body as a system. While not all of the conceptual explanations we observed in the data were based on accurate conceptual understandings, we argue that their main purpose within the conversation was to advance the team's understanding of some aspect of the problem posed in the question, and as such, these contributions were often crucial to the teams' development of a rationale for their answer to the question.

The other important contribution type was the "re-evaluation" of a previous stretch of talk, which consisted of talk in which the speaker expressed disagreement with a teammate, pointed out something another person had not thought of, or directly asked for further explanation. Our definition roughly encompasses several codes defined in Asterhan and Schwarz (2009), including challenge, opposition, rebuttal, and request-

ing information. Re-evaluation turns appeared to trigger conceptual explanations, further elaboration of previous conceptual explanations, or articulation of counter-explanations to previous explanations.

One discursive phenomenon we discerned in the data that we argue may be especially important to students' learning is their joint interpretation of questions and negotiation and creation of answers, a phenomenon we refer to as the "co-construction" of meaning and understanding. While in some sense, all social interaction, including all classroom talk around problem solving, is co-constructed (Jacoby and Ochs, 1995), for our purposes in this study, we coded as co-construction stretches of talk wherein two or more students' turns responded to and built upon one another to create an explanation of the phenomenon in question. This aligns with Chi's (2009) use of the term. We included in these stretches of co-constructed explanations turns in which the speaker responded by backchanneling, that is, turns such as "uh-huh" or "yeah" that serve a meta-conversational function. This backchanneling indicates the speaker's attentive listening to the participants who were more actively engaged in producing an explanation.

We then identified and counted how many students participated in conceptual explanation, re-evaluation, and co-construction for each question and resolved differences by consensus. We calculated the mean and SD of these counts for lower-level and higher-level questions (there were not enough medium-level questions to be included in the analysis). For each of the three coding categories, a Student's *t* test was used to test for statistical significance in modes of student discourse seen after lower-level versus higher-level questions.

RESULTS AND DISCUSSION

Examples of Discourse Components Occurring in Response to Question Prompts

In this section, we provide examples of student conversations in which the discursive phenomena described earlier occurred (for definitions and examples of discursive phenomena, see Table 2). An example of conceptual explanations as part of student discourse can be seen in the following discussion in which students were asked to explore the consequences of a specific congenital heart defect.

Question: A newborn baby needs surgery because she was born with an aorta that arises from the right ventricle and a pulmonary trunk that issues from the left ventricle, a condition called "transposition of the great vessels." Why does this need to be corrected?

- there would be too much blood delivered to the systemic circuit
- the blood pressure in the pulmonary blood vessels would be too high
- oxygenated blood would never reach the systemic circuit
- no blood would ever be delivered to the lungs

The correct answer is "c." The key insight is to realize that this defect produces two separate noninteracting circuits of blood. Typically, a few individuals realize that "b" is also a potential problem due to the greater pumping capacity of the left ventricle; this usually arises in whole-class discussion after voting.

TABLE 2. Definitions and examples of discursive phenomena identified in student conversations

Discursive phenomenon	Definition	Example
Conceptual explanation	Stretches of talk in which one or more students in the group produce a hypothesis of physiological functioning that is not simply fact based but draws upon a conceptual understanding of the human body as a system.	Matthew: I was thinking because it would be coming back up through the inferior vena cava. And since veins have valves as well, just like the heart, to prevent backflow, that it would kind of be like being accumulated there 'cause it's not pumping hard enough to get it like all the way back quick enough. (see Excerpt 3)
Re-evaluation	In reference to a previous stretch of talk, talk in which the speaker expresses disagreement with a teammate, points out something another person had not thought of, or directly asks for further explanation.	Brianna: [referring to Samira's earlier explanation of pulmonary edema as a buildup of fluid in the lungs] How would you get buildup of fluid? (see Excerpt 4)
Co-construction	Turns in which two or more speakers respond to and build upon one another's contributions to create an explanation of the phenomenon in question.	Jennifer: Wouldn't it have to do with oxygenation, though? Kristina: Because you would get the oxygenated blood from the lungs. Jennifer: But like if her aorta is from her left ventricle, right? Or the right ventricle, sorry. Kristina: Right. So the right ventricle and the right atrium aren't going to be able to pump as much as if it had been the left. 'Cause it's smaller. So then you would get less blood being pumped out to the body. (see Excerpt 2)

Excerpt 1. Conceptual Explanation

Codes: CE = conceptual explanation, R = re-evaluation, Co = co-construction

Codes		Transcript: Team 7
CE		Megan: Yeah, so, in th—in the right aorta, or right atrium, you have deoxygenated... Jasmine: Yeah.
CE		Megan: ...de-[laughs]-genated blood that's going to the aorta. I mean, it's going to the body which means you're just pumping useless blood... Jasmine: Useless blood. Megan: ...to the body. [Laughs.] Khaled: Mm-hm.
CE		Megan: And then here, you've got oxygenated blood going back to the lungs. Jasmine: Yeah. Megan: That's completely flip-flopped. Khaled: Yeah. [Megan laughs.] It's completely flip-flopped. Megan: So, "c." That's not "d" because blood would be delivered to the lungs from the left ventricle. So it's "c." Khaled: What is it? [Reading] "No blood would be ever delivered—" Jasmine: [Reading softly and slowly] "Pulmonary blood vessels would be too high." No.
CE	Co	Khaled: So in other words, the blood's still coming in at the right atrium and it's just going out of the left—right ventricle, 'cause the aorta being connected to the right ventricle it said, right? Oh, wait, pulmon—yeah, right ventricle. Aorta arises from the right ventricle. Right?
	Co	Megan: These two should be flipped.
CE	Co	Khaled: They should be flipped. Yeah. 'Cause the right ventricle sends it to the pulmonary trunk, but what's happening in this case is the blood is coming in the atrium, the right atrium, and it's just getting sent out from the right ventricle.
CE	Co	Megan: [Over Khaled] And then shooting to the aorta, yeah.
	Co	Khaled: Yeah, so nothing's really happening on this side, right?
	R	Jasmine: So c—I was about to say, so—
CE	Co	Megan: Ye—Well—yeah, because it's—If you start at that point...
CE	Co	Khaled: [Over Megan] You don't have blood coming back from the pulmonary trunk into the left atrium.

In all four recorded teams, a significant amount of conceptual explanation was seen in response to this question. In another team discussing the same question,

we can also see, in addition to conceptual explanation, many instances of re-evaluation in the team's discussion.

Excerpt 2. Conceptual Explanation and Re-evaluation

Codes: CE = conceptual explanation, R = re-evaluation, Co = co-construction

Codes		Transcript: Team 14
CE		Kristina: So I was thinking “c.” If you have this oxygenated blood pumping from this—um ...from this right side through the aorta, then you would have less blood being pumped to the body. Right? Because...
CE		Priyanka: Oh, wait, I think this would be [unintelligible]. So basically everything is—Oh, it’s like flipped.
	R	Co Jennifer: Wouldn’t it have to do with oxygenation, though?
CE		Co Kristina: Because you would get the oxygenated blood from the lungs.
	R	Co Jennifer: But like if her aorta is from her left ventricle, right? Or the right ventricle, sorry.
CE		Co Kristina: Right. So the right ventricle and the right atrium aren’t going to be able to pump as much as if it had been the left. ‘Cause it’s smaller. So then you would get less blood being pumped out to the body.
		[6-second pause]
		Jennifer: Hmm.
		[17-second pause]
		Co Nicole: So from the—from the body, it goes into the... right...
		Co Priyanka: Right.
		Co Nicole: Right.
		Co Priyanka: Right?
		Co Jennifer: Well, normally...
	R	Co Priyanka: In a normal...See, I’m just confused. If this—The pulmonary trunk issues from the left ventricle...
		[7-second pause]
CE		Co Nicole: So then would blood ever go into the left side? Because if it goes from the right aorta down to the right ventricle, to the aorta, to the body, then it would go back up into the right aorta.
	R	Co Kristina: An aorta is artery, so it’s not...
CE		Co Nicole: I mean atrium. [7-second pause] Right? ‘Cause like if you ... the blood would start in the right atrium, then it would go down to the right ventricle, and then if it says that the right ventricle takes it to the aorta, then it would go to the aorta, to the body, which then takes it back to the right atrium.
	R	Co Priyanka: Yeah, I just don’t understand what it means, your ... your pulmonary trunk is from the left ventricle. Like...
		Co Nicole: What do you mean?
		Co Priyanka: The pulmonary trunk issues from the left ventricle.
CE		Co Nicole: So instead of coming from the right ventricle...
		Co Priyanka: Yeah.
CE		Co Nicole: ...it would go off the left...
		Co Priyanka: ...just from the left.
		Co Nicole: ...and so it...
	R	Co Priyanka: So would it ever even go to the left?
		Co Nicole: Yeah, I don’t think so.
CE	R	Co Priyanka: But how could that be possible if—‘Cause the blood that goes to the right is not oxygenated, right? It has to go to the lungs to get oxygenated and then travels [unintelligible].

In general, teams found this question difficult and tended to struggle through it together. This interactive engagement was associated with co-construction of understanding for this question. Another example of the role of co-construction and re-evaluation in conceptual explanation comes from a question about the pathophysiological condition called left ventricular failure.

Question: A middle-aged woman is admitted to the coronary care unit with a diagnosis of left ventricular failure resulting from a myocardial infarction. Her skin is pale and cold, and moist sounds are heard over the lower regions of both lungs.

Why does left ventricular failure cause moist sounds to be heard in her lower lungs? [no answer choices; open-ended question]

The correct answer is that an inability of the left ventricle to eject an appropriate volume of blood causes a buildup of blood and therefore blood pressure “behind it” in the pulmonary circuit. Students typically envision the moist sounds as being those of the built-up blood itself (which is good progress), while in whole-class discussion, a few individuals will go further to identify the situation as one of pulmonary edema, in which the high pressure forces fluid out of the pulmonary capillaries into the interstitial space.

Excerpt 3. Re-evaluation and Co-construction

Codes: CE = conceptual explanation, R = re-evaluation, Co = co-construction

Codes		Transcript (Team 5)
CE		David: The only thing I can think of is like fluid is being sent there to try and like recuperate. Matthew: She has pneumonia? David: 'Cause she has pneumonia. Matthew: Bronchitis? David: In her lower lung. Matthew: Lower regions of both lungs. David: Well...
CE	Co	Matthew: I don't know. 'Cause blood's getting caught in the valve of the vena cava down here instead of making it all the way back up?
CE	Co	Amina: Yeah. So if it's the blood is not flowing to the ventricle, it stays in the lungs and that's what makes the—the sound. David: Varicose veins. Matthew: I don't know what else it would be. Moist sounds.
CE	Co	Amina: Or maybe the valve, you know, the valve—maybe that's the sound it makes, it's like the valve that is between the ventricle and the—the lungs.
	Co	Matthew: Uh-huh.
	Co	Amina: Maybe that's what's making the sound.
	R	Matthew: Yeah, but that's still in the heart. What would that have to do with the lower lung?
	Co	David: It's—yeah.
CE	Co	Amina: And the—I mean—it pumps the—the blood to the ventricle [<i>unintelligible</i>]. It opens and it closes.
	R	Matthew: Say again?
CE	Co	Amina: So the blood that flows from the lungs, oxygenated blood that comes from the lungs...
	Co	Matthew: Uh-huh.
CE	Co	Amina: ...to the ventricle.
	Co	Matthew: Goes to the...
	Co	Amina: Goes to the [<i>unintelligible</i>]. Like it opens and it closes.
	Co	Matthew: Mm-hm. That'd be a heart murmur, if the valve wasn't working right.
	Co	Amina: I guess so, yeah. [6-second pause]
	R	David: Why is it in the lower lung?
CE	Co	Matthew: I was thinking because it would be coming back up through the inferior vena cava. And since veins have valves as well, just like the heart, to prevent backflow, that it would kind of be like being accumulated there 'cause it's not pumping hard enough to get it like all the way back quick enough. I don't know. That's what I'm thinking.

The discussion in Excerpt 3 began weakly and erroneously, but the team nevertheless managed, through co-construction and repeatedly questioning one another, to make significant progress toward the correct answer. In

another team (Excerpt 4), the answer to this question was arrived at with less discussion, but still involved conceptual explanations and had a useful instance of re-evaluation.

Excerpt 4. Conceptual Explanation and One Instance of Re-evaluation

Codes: CE = conceptual explanation, R = re-evaluation, Co = co-construction

Codes		Transcript (Team 10)
CE	Co	Samira: Could it be like pulmonary edema where you have buildup of fluid in your lungs because you can't—properly dispose of ... oxygen?
	R	Alexis: But why would that—
	R	Brianna: [<i>Over Alexis</i>] How—how would you get buildup of fluid? [45-second pause]
CE	Co	Brianna: Maybe it's like—kind of like what you said, maybe it's just like building up in the lungs 'cause, if it's not being pumped out as fast from the ventricles, it's like filling up the ventricle and the atrium and like going ... to the lungs. Like there's just like a lot of-
CE	Co	Alexis: [<i>Over Brianna</i>] Because the rest of the heart would still work normally, right? So it would be ... the right ventricle would still be pumping blood to the lungs, but then if they left ventricle isn't pumping it to the rest of the body as fast, there would be like a backup of blood.

Notably, both of the questions discussed in the four excerpts presented above are clinical applications, the real-world context of which may be inherently motivating to students, which we discuss further later in the article. For some concepts or learning goals, a true real-world or clinical application is difficult to construct, but a similar effect may be achieved by asking students to determine the effect of a theoretical change to the normal state of the system. For example, when the class was studying action potentials in neurons, a question about how the action potential would change if ion concentration gradients were different produced a rich conversation (Excerpt 5) with multiple instances of conceptual discussion as well as re-evaluation.

Question: What would happen to the size and shape of the action potential if we doubled the concentration of potassium outside the cell?

The correct answer is that the repolarization phase of the action potential will be slower and not as negative (“less hyperpolarization,” as Kristina says). Further, the resting membrane

potential (“baseline” in the conversation below) will be relatively depolarized.

Team discussions in response to some other questions contained far less conceptual explanation, re-evaluation, and co-construction, and often resulted in shorter discussions followed by periods of silence or, in a few cases, off-topic conversations. An example is a question about the effect of a constriction of blood vessels on blood flow; Team 10’s discussion (Excerpt 6) is typical:

Question: During exercise, the diameter of the renal artery, to the kidney, decreases by half—say from 2 mm to 1 mm. How does the rate of blood flow through the renal artery change?

- a) increase by half
- b) decrease by half
- c) decrease by 4 times
- d) decrease by 16 times

The correct answer is “d”; blood flow is proportional to the fourth power of the blood vessel diameter.

Excerpt 5. Conceptual Explanation and Re-evaluation

Codes: CE = conceptual explanation, R = re-evaluation, Co = co-construction

Codes	Transcript (Team 14)
	Austin: Would it be about the baseline this time?
	Nicole: Mm-hm, I think so.
	[8-second pause]
CE	Austin: And potassium is trying to go out, but because there’s more potassium on the outside, it will be slower?
	Nicole: The—yeah, that’s wh—that’s what I would think.
	[9-second pause]
CE	Priyanka: So if there’s more potassium outside, when it’s leaving, at some point it’s going to be like, “OK, I should stop leaving.” So the one that’s trying to come back, won’t it come back faster? To baseline?
R	
CE	Kristina: Like in the depolarization, wouldn’t that be slower, because less potassium would—Like it’d be harder for the potassium to go out? Because there’s already potassium outside the cell? Isn’t that what we’re thinking?
R	
	Co Priyanka: Yeah.
	Co Kristina: I think what they’re talking about is the—uh...
CE	Priyanka: See, I just don’t like the whole baseline thing. So it’s like, it’ll leave and it’ll go through the point where it’s extremely negative, like goes up to negative 90 or something? And I’m trying to think like when—at some point it’ll re-enter...
R	
	Co Jennifer: Mm-hm.
	Co Priyanka: ...to go back at threshold, like the negative 70 sometime?
CE	Co Kristina: I don’t know if it would even do as much of the hyperpolarization because there’s gonna be—
CE	Priyanka: [Over Kristina] That—yeah, so wouldn’t that be ... so what if ... so if we’re saying—baseline is negative 70, if the hyperpolariz—if it’s not going down all the way to negative 90 and it’s just going down to like negative 80 or something, so won’t it come back to baseline?
	Co Kristina: I don’t know...
	Co Priyanka: [Over Kristina] I don’t—yeah.
	Co Kristina: ...because the entire depolarization might be slower.
	Co Priyanka: OK.
	Co Kristina: You know?
	Co Priyanka: I und... yeah, yeah.
	Co Kristina: I think, like...
	Co Priyanka: I think I’m just thinking about...
	Co Kristina: Slower depolarization but less hyperpolarization?
	Co Priyanka: Yeah, that’s what I’m a bit confused, like...
	Co Kristina: But...
	Co Nicole: I think it’s repolarization, so it would be...
	Co Kristina: Oh, alright, repolarization.

Excerpt 6. No Conceptual Explanation, Re-evaluation, or Co-construction

Codes: CE = conceptual explanation, R = re-evaluation, Co = co-construction

Code(s)	Transcript (Team 10)
	Samira: OK, so it goes from two, right? Goes from two ... [pause] so...
	Brianna: C?
	Samira: ...blood flow would go to sixteen ... It would decrease by sixteen times.
	Samira: Yeah.
	Shreya: How did you do that?
	Samira: 'Cause there's an equation. It said that blood flow is radius, uh, to the fourth.
	Shreya: OK.

Similar poor-quality discussion was seen in response to a question that asked students to calculate the cardiac output of each student in their team. Even though cardiac output is a central concept of cardiovascular physiology, and the students did not previously know how to calculate it, the question did not produce conceptual discussion in any team.

Question: What is the cardiac output of the person next to you, assuming a stroke volume of 70 ml/beat?

The correct result is achieved by multiplying 70 ml/beat times the person's heart rate (achieved by taking the pulse of one's teammate).

Excerpt 7. No Conceptual Explanation, Re-evaluation, or Co-construction

Codes: CE = conceptual explanation, R = re-evaluation, Co = co-construction

Code(s)	Transcript (Team 10)
	Samira: We're doing cardiac output, right?
	Brianna: Yeah.
	Samira: So it's just heart rate times stroke volume and they already gave us the stroke volume, so we just need to take...
	Alexis: [Unintelligible.]
	Samira: No, so we just need to take our...
	Shreya: Pulse.
	Samira: ...our heart rate, right?
	Michaela: Mm-hm.

Association of Team Discourse Components with Cognitive Level of Questions. The observed variability in frequency of the discourse elements we focused on in student discussions suggested that differences in characteristics of the question asked, such as its cognitive level, might be influencing these characteristics of team discussions. We coded each team discussion for quantitative analysis, and questions asked during class time were independently scored as to cognitive level, as described earlier, by each of the three authors; differences were resolved by consensus. Summary data and examples of questions in each category are provided in Table 3. Questions were rated as high, medium, or low cognitive level. There were not a sufficient number of medium-level questions to analyze (three total across the semester). Five questions were rated as "mixed";

these were multipart questions containing both higher- and lower-level components and were also discarded from the analysis, because they did not fall clearly in any category. Thus, we focused our quantitative analysis on the high- and low-cognitive-level categories. Differences in instances of conceptual explanations, re-evaluation, and co-construction between high- and low-cognitive-level questions are shown in Figure 2 later in this article. High-cognitive-level questions were associated with significantly more instances of conceptual explanation ($p = 0.0037$), re-evaluation ($p = 0.028$), and co-construction ($p = 0.0093$) than lower-cognitive-level questions. This relationship is illustrated in the excerpts. For example, the question prompts for the responses in Excerpts 1–5 were scored as higher level; the ensuing discussions showed that students had to approach the question conceptually to arrive at an answer. By contrast, questions that did not require higher-level reasoning to answer, even questions that were often missed by students on individual exams, did not appear to promote conceptual explanations in team discussions. For example, the questions in Excerpts 6 and 7 did not require significant discussion once one student provided the simple formula required to determine the answer.

Pre–Post Data. Students scored significantly higher on the physiology concept posttest taken on the last day of class than they did on the pretest taken on the first day of class ($75.8 \pm 17.3\%$ vs. $37.9 \pm 14.4\%$, $p < 0.0001$). This finding suggests that students in this class gained overall understanding of physiological concepts over the semester. In addition, pre–post concept-based quizzes given at the beginning and end of class days in which recordings occurred all showed significant gains from the beginning to end of class time; because students had already had some preclass preparation on the topic, this demonstrates an improvement on individual understanding as a result of class work. Figure 1 shows one such result: In September, students scored an average of 1.9/5 on the neurophysiology pretest and a 3.8/5 on the posttest 1 hour later (Figure 1, $p < 0.0001$). Significant gains across class time were also seen in October (cardiac physiology; pretest mean, 2.1/5; posttest mean, 4.1/5, $p < 0.0001$) and December (immune system physiology; pretest mean, 3.0/5; posttest mean, 4.3/5, $p < 0.005$). These data are consistent with an improvement in conceptual understanding of each day's topic as a result of class activities. Although class time was filled with conceptual questions posed to teams, the quiz questions did not repeat, nor were they generally isomorphic to, in-class questions.

The Role of Question Level in Discussion Quality. In our study, higher-level questions prompted more meaningful discussions than lower-level questions as evidenced by increased instances of conceptual explanation, re-evaluation, and co-construction of understanding in response to such questions. This finding is in contrast to Knight *et al.* (2016) and James and Willoughby (2011), who did not find a strong influence of Bloom's level on quality of student discussion. In these studies, students were just as likely to make arguments using reasoning in response to questions that were rated as low level on the Bloom scale. A possible explanation for the difference between our findings and those of these authors is the existence of built-in baseline student knowledge as part of the structure of TBL. Students participating in discussions in our study had

TABLE 3. Cognitive level of questions asked of student teams during the three recorded class periods

Cognitive level	Number of questions in this category	Examples
Low	9	What causes the resting membrane potential? What is the cardiac output of the person next to you, assuming a stroke volume of 70 ml/beat? During exercise, the diameter of the renal artery, which carries blood to the kidney, decreases by half, from 2 mm to 1 mm. How does the rate of blood flow through the renal artery change?
Medium	3	At this point in the action potential, there is one type of K ⁺ channel open (leak channel) and one type of Na ⁺ channel open (voltage-gated). Why, then, is there more Na ⁺ current and a net depolarization? If you wanted to make (and sell) a male contraceptive, should you block GnRH, LH, or FSH?
High	13	What will happen to the shape and size of the action potential if we double the concentration of sodium <u>inside</u> the cell? Why? A middle-aged woman is admitted to the coronary care unit with a diagnosis of left ventricular failure resulting from a myocardial infarction. Her skin is pale and cold, and moist sounds are heard over the lower regions of both lungs. Why does each of these symptoms occur? If you took a drug that blocked FSH, would ovulation still occur?

already completed the readiness assessment process, which 1) incentivized them to obtain specific baseline topical content outside of class, 2) engaged them in team discussions about this content, and 3) through the immediate feedback provided by the scratch-off forms, corrected any errors or gaps in comprehension of this basic material. If the lower-level questions posed during class tapped into prior knowledge gleaned from the readiness assessment process, then those questions may have been seen as less worthy of discussion. By the same token, this preparation was likely a necessary prerequisite for the amount and quality of conceptual discussion that occurred in response to higher-level questions.

Many of the questions used in the class sessions under study here could be coded at Bloom's level 3 and 4, application or analysis, and were by nature often framed in a real-world context, usually medical. Students may find such questions inherently interesting and are probably more motivated to expend the cognitive energy required to figure them out. Using a meaningful context is often recommended for those writing in-class questions in order to increase student initiative; for example, Anastasio and Ingram (2018, p. 3) concluded that "questions work best when they focus on something worth considering; that is, when they are nontrivial." Practitioners of TBL recom-

mend a similar approach through the important question-writing criterion of "significance," which encourages instructors to write in-class questions that use the real professional context of the material as much as possible.

The Role of Re-evaluation in Student Discussions. In addition to being associated with more conceptual explanation, higher-level questions were associated with a greater number of instances of re-evaluation (Figure 2). Re-evaluation may have facilitated the effect of question level on conceptual explanation, as we often saw that conceptual explanations occurred in direct response to a re-evaluation. This interpretation is suggested by Repice *et al.* (2016, p. 31), who observed the same relationship: "When students ask each other 'why' or 'how' they reached their answers or conclusions, a *conceptual explanation* often followed."

Alternatively, more difficult questions may primarily have promoted the confusion that led to re-evaluation statements ("See, I'm just confused"; "But how would that be possible?"; "But why would that cause...?") and independently provoked the required conceptual explanation to address it. Making students aware that their explanations are not clear to others can help them recognize their own misunderstanding. In addition, these questions may provide both the challenge and (especially when framed in a real-world or clinical context) motivation to pursue clarification when confusion arises. These kinds of

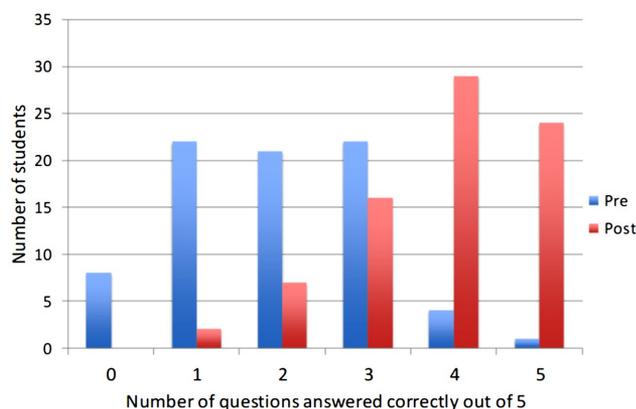


FIGURE 1. Performance of all students in the class on the five-question conceptual quiz on the day's topic given at the beginning and the end of the September recording day (see Supplemental File S1).

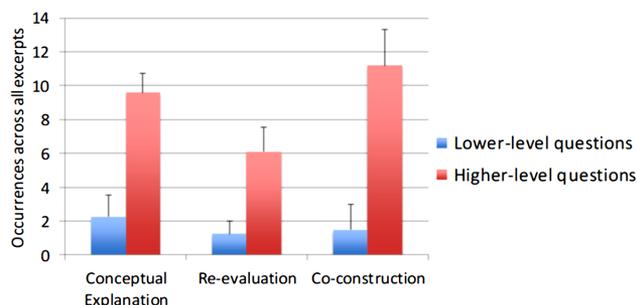


FIGURE 2. Average number of students per question (total among four recorded teams) participating in conceptual explanation, re-evaluation, and co-construction after questions of low and high cognitive level.

questions that prompt re-evaluation are pushing students into the ICAP interactive mode (Chi and Wylie, 2014). Students engaged in a back-and-forth discussion to address one another's concerns are moving beyond their own cognitive constraints and are benefiting from another student's perspective and discernment. Specifically, when confusion arises, individuals are stimulated either to justify what they have said so that it makes sense to someone else or to reconcile their prior knowledge with the new insights that emerge.

As with the importance of question quality, student preparation may have a role in promoting quality re-evaluation in team discussions. As noted earlier, students did not generally ask each other about the vocabulary of the question or basic definitions, perhaps because these had been previously provided by readiness assessment. Such prior domain knowledge appears to stimulate question-asking at a level that promotes conceptual understanding rather than primarily advancing basic factual knowledge (Chin and Osborne, 2008). More-developed understanding may also raise "vexing questions" that provide an affective component that motivates team members to sustain interest in the overall problem (Odden and Russ, 2019).

The Importance of Co-construction in Student Discussions. Smith *et al.* (2009) established that discussion among students improves understanding rather than simply spreading the influence of the most knowledgeable individuals. Smith *et al.* (2011) further established the importance of peer discussion after a clicker question for answering a follow-up, isomorphic question correctly, even for the strongest students, demonstrating the importance of student discussion for students of all skill levels. Versteeg *et al.* (2019) demonstrated a similar positive effect of peer discussion in a physiology class context and showed that student performance improved among individuals with an initially incorrect answer even after discussion with another student with an incorrect answer. They attributed this effect to the importance of co-construction in interactive engagement as described by Chi and Wylie (2014). Kulatunga *et al.* (2013) also found, in a general chemistry class taught using peer-led guided inquiry, that co-construction was common in student discourse and that students were more likely to elaborate on their reasoning when more than one student was involved in construction of the argument. In our study, we also found that co-construction of answers (as opposed to one student explaining the answer to the others in the group) was common in student discourse, particularly in response to higher-level questions (Figure 2). Our data further suggest that higher-level questions provoke more such interactive engagement, stimulating students to expand on their own understanding through elaboration and co-construction of meaning. As noted before, such interactive engagement has been found to produce better student learning outcomes than individual students constructing explanations for the group (Chi and Wylie, 2014).

CONCLUSIONS AND RECOMMENDATIONS FOR PRACTICE

When we teach, we want to stimulate passionate interactive discussion among our students, to have them be genuinely interested in the question and its answer. How can we generate this strong vested interest in the questions and answers to the questions?

Our first recommendation for practice, on the basis of our results and those of others, is to use questions that are themselves at a high Bloom's level/conceptual level and, further, that are rich in the context of the discipline. Instructors should create questions that can only be answered via a process of disciplinary thinking that the instructor hopes to promote.

The second recommendation from our study results is that instructors plan to promote the practice of students questioning one another. Enticing students to debate and disagree capitalizes on the social aspects of learning as described in the ICAP framework's interactive mode of engagement and enables student teams to advance beyond any individual's understanding. As discussed earlier, an important factor that contributes to productive disagreement is the use of higher-level questions. Another key element is presumably the quality of group functioning, which can influence the willingness of individuals to undertake the risk of disagreeing. Faculty can both set expectations and provide support for students' civil disagreement. For example, helping students set guidelines for group discourse at the beginning of class can both illustrate that disagreement and questioning are expected as well as provide rules for that form of engagement. Supplying students with sample language for respectful divergent dialogue can further ease student concerns.

Our third recommendation relates to the role of accountability in the quality of student discussion. Holding students accountable for their answers to questions posed in class may be an important contributing factor in producing high-quality discussion (Dallimore *et al.*, 2006; Knight *et al.*, 2016). This responsibility can be imposed at both the team and individual level. At the team level, we recommend making the teams' answers public through some manner of simultaneous report, such as voting with colored cards or electronic devices, or filling in an online data sheet that is revealed to all students at a specific time. In TBL classrooms such as the one in this study, most questions are reported out through simultaneous vote via colored letter cards. The pending prospect of, and countdown to, the unveiling of the group's answer choice provides some element of focus and motivation to the discussion. Such public display may harness the positive power of peer pressure to encourage students to prepare a higher-quality answer than they would if reporting out anonymously via clickers or polling software.

At the individual level, we can promote accountability via random call. Evidence suggests that students called on randomly become more comfortable speaking in class and are more likely to voluntarily answer questions in class (Dallimore *et al.*, 2012). Further, random call is cited by students as a major factor in promoting attendance and engagement in class (Broeckelman-Post *et al.*, 2016) and has been demonstrated to increase self-reported preparation for class (Dallimore *et al.*, 2006). The classroom under study involved a form of random call sometimes called "warm calling," in which students are only targeted to answer a question *after* they have had a chance to discuss that question within the team. This practice is designed to reduce student discomfort with random call, an issue highlighted by Cooper *et al.* (2018). In our study, individual random call (using a shuffled stack of cards with names and photos of students) was used multiple times per question and may have been a factor in promoting high-quality discussion.

Group random call may also be effective in promoting these ends. In an introductory biology course, Knight *et al.* (2016) found that a group random call condition was associated with increased exchanges of quality reasoning, more turns at speaking, and increased number of requests for information and feedback from teammates.

Our study built on and shared similarities to work in other contexts such as a chemistry PLTL course (Repice *et al.*, 2016) and biology courses taught using cooperative learning (Bierema *et al.*, 2017) or peer instruction (Knight *et al.*, 2013). Our findings, however, are limited to our experience with one specific course and one subject area over the course of only one semester in one institution and reflect only certain functions of student discourse. In addition, we monitored conversations of only a few arbitrarily selected groups of students, and although they shared many characteristics with students in the class as a whole (grade point average [GPA], gender, and race; see Table 1), their behaviors could have been specific to their group dynamics. We look forward to the application of our coding schemes by other researchers in other courses with a broader sample of students to determine the generalizability of our results and further delineate the role of question quality in the level of student engagement and discourse.

One factor contributing to quality of discussion that was not directly explored by our study, but is a promising area for future study, is the length of time students have spent working together as a team. Instructors often struggle with the question of whether student learning is best promoted by using permanent teams or by re-forming teams periodically. Although evidence is scant on this issue (Hodges, 2018), it is possible that the more time students spend together, the more they are able to, or practice, decentering behaviors, in which they take on the perspective of another individual or context besides their own (Moon *et al.*, 2017). In addition, a study in a peer instruction classroom found that students became more expert-like in their thinking if groups were permanent (Zhang *et al.*, 2017). One advantage of the use of permanent teams, as encouraged in TBL, is that students can become more accustomed to their groups' processes and seemingly have the time to progress through Tuckman's phases of group process: forming, storming, norming, and performing (Tuckman, 1965).

We conclude by recommending that instructors strongly consider recording team conversations from time to time, even if they do not plan to make a formal study of them. Many lessons can be learned from what confuses students, what they already understand, and how they approach problems. For example, during the neurobiology class period, the instructor made an unplanned comment suggesting that team members who had explained the reasoning behind the previous question should be quiet and allow other team members to explain their reasoning for the current, isomorphic, question. This instruction resulted in immediate and long-term changes in participation in at least one team, a result that would have been invisible to the instructor had the team discussions not been recorded. We can learn a lot by listening to our students.

ACKNOWLEDGMENTS

We thank Cynthia Wagner for assistance with the design of the study and determining the coding scheme and Sarah Swatski for assistance with figure preparation.

REFERENCES

- Anastasio, D., & Ingram, E. (2018). *Better questions: A learning opportunity* (IDEA paper #71). Manhattan, KS: IDEA Center.
- Asterhan, C. S., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cognitive Science*, 33(3), 374–400.
- Beatty, I., Gerace, W., Leonard, W., & Dufresne, R. (2006). Designing effective questions for classroom response system teaching. *American Journal of Physics*, 74(1), 31–39. <https://doi.org/10.1119/1.2121753>
- Bierema, A. M., Schwarz, C. V., & Stoltzfus, J. R. (2017). Engaging undergraduate biology students in scientific modeling: Analysis of group interactions, sense-making, and justification. *CBE—Life Sciences Education*, 16(4), ar68. <https://doi.org/10.1187/cbe.17-01-0023>
- Broeckelman-Post, M., Johnson, A., & Schwebach, J. R. (2016). Calling on students using notecards: Engagement and countering communication anxiety in a large lecture. *Journal of College Science Teaching*, 45(5), 27–33.
- Chi, M. T. H. (2009). Active-Constructive-Interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1, 73–105. <http://doi.org/10.1111/j.1756-8765.2008.01005.x>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <http://doi.org/10.1080/00461520.2014.965823>
- Chin, C., & Osborne, J. (2008). Students' questions: A potential resource for teaching and learning science. *Studies in Science Education*, 44(1), 1–39. <https://doi.org/10.1080/03057260701828101>
- Cooper, K. M., Downing, V. R., & Brownell, S. E. (2018). The influence of active learning practices on student anxiety in large-enrollment college science classrooms. *International Journal of STEM Education*, 5, ar23. <https://doi.org/10.1186/s40594-018-0123-6>
- Crowe, A., Dirks, C., & Wenderoth, M. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education*, 7(4), 386–381. <https://doi.org/10.1187/cbe.08-05-0024>
- Dallimore, E. J., Hertenstein, J. H., & Platt, M. B. (2006). Nonvoluntary class participation in graduate discussion courses: Effects of grading and cold calling. *Journal of Management Education*, 30, 354–377. <https://doi.org/10.1177/1052562905277031>
- Dallimore, E. J., Hertenstein, J. H., & Platt, M. B. (2012). Impact of cold-calling on student voluntary participation. *Journal of Management Education*, 37(3), 305–341. <https://doi.org/10.1177/1052562912446067>
- Ertmer, P., Sadaf, A., & Ertmer, D. (2011). Designing effective question prompts to facilitate critical thinking in online discussions. *Design Principles and Practices: An International Journal*, 5(4), 1–28. <https://doi.org/10.18848/1833-1874/CGP/v05i04/38121>
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Haidet, P., Kubitz, K., & McCormack, W. (2014). Analysis of the team-based learning literature: TBL comes of age. *Journal on Excellence in College Teaching*, 25, 303–333.
- Hodges, L. C. (2018). Contemporary issues in group learning in undergraduate science classrooms: A perspective from student engagement. *CBE—Life Sciences Education*, 17(2), es3. <https://doi.org/10.1187/cbe.17-11-0239>
- Jacoby, S., & Ochs, E. (1995). Co-construction: An introduction. *Research on Language and Social Interaction*, 28(3), 171–183.
- James, M. C., & Willoughby, S. (2011). Listening to student conversations during clicker questions: What you have not heard might surprise you! *American Journal of Physics*, 79, 123–132. <https://doi.org/10.1119/1.3488097>
- Johnson, D. W., & Johnson, R. T. (1999). *Learning together and alone: Cooperative, competitive, and individualistic learning* (5th ed.). Boston: Allyn & Bacon.
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college: What evidence is there that it works? *Change*, 30(4), 26–35. <https://doi.org/10.1080/00091389809602629>
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (2006). *Active learning: Cooperation in the university classroom* (3rd ed.). Edina, MN: Interaction.
- Knight, J. K., Wise, S. B., & Sieke, S. (2016). Group random call can positively affect student in-class clicker discussions. *CBE—Life Sciences Education*, 15(4), ar56. <https://doi.org/10.1187/cbe.16-02-0109>

- Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: Student reasoning and the impact of instructional cues. *CBE—Life Sciences Education*, 12(4), 645–654. <https://doi.org/10.1187/cbe.13-05-0090>
- Kubitz, K. (2014). The evidence, please. In Sibley, J., & Ostafichuk, P. (Eds.), *Getting started in team-based learning* (pp. 45–61). Sterling, VA: Stylus.
- Kulatunga, U., Moog, R., & Lewis, J. (2013). Argumentation and participation patterns in general chemistry peer-led sessions. *Journal of Research in Science Teaching*, 50(10), 1207–1231. <https://doi.org/10.1002/tea.21107>
- Michaelson, L. (2002). Getting started with team-based learning. In Michaelson, L., Knight, A. B., & Fink, L. D. (Eds.), *Team-based learning: A transformative use of small groups* (pp. 27–51). New York: Praeger.
- Moon, A., Stanford, C., Cole, R., & Towns, M. (2017). Decentering: A characteristic of effective student-student discourse in inquiry-oriented physical chemistry classrooms. *Journal of Chemical Education*, 94(7), 829–836. <https://doi.org/10.1021/acs.jchemed.6b00856>
- Odden, T. O., & Russ, R. (2019). Vexing questions that sustain sensemaking. *International Journal of Science Education*, 41(8), 1052–1070. <https://doi.org/10.1080/09500693.2019.1589655>
- Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, 95(4), 627–638. <https://doi.org/10.1002/sce.20438>
- Repice, M. D., Sawyer, R. K., Hogrebe, M. C., Brown, P. L., Luesse, S. B., Gealy, D. J., & Frey, R. (2016). Talking through the problems: A study of discourse in peer-led small groups. *Chemistry Education Research and Practice*, 17, 555–568. <https://doi.org/10.1039/c5rp00154d>
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.
- Saldaña, J. (2015). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323, 122–124. <https://doi.org/10.1126/science.1165919>
- Smith, M. K., Wood, W. B., Krauter, K., & Knight, J. K. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE—Life Sciences Education*, 10(1), 55–63. <https://doi.org/10.1187/cbe.10-08-0101>
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21–51.
- Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.
- Swanson, E. A., McCulley, L. V., Osman, D. J., Scammacca Lewis, N., & Solis, M. (2017). The effect of team-based learning on content knowledge: A meta-analysis. *Active Learning in Higher Education*, 20(1), 1–12. <https://doi.org/10.1177/1469787417731201>
- Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin*, 63(6), 384–399. <https://doi.org/10.1037/h0022100>
- Versteeg, M., van Blankenstein, F., Putter, H., & Steendijk, P. (2019). Peer instruction improves comprehension and transfer of physiological concepts: A randomized comparison with self-explanation. *Advances in Health Sciences Education*, 24(1), 151–165. <http://doi.org/10.1007/s10459-018-9858-6>
- Wood, A. K., Galloway, R., Hardy, J., & Sinclair, C. (2014). Analyzing learning during peer instruction dialogues: A resource activation framework. *Physical Review Special Topics—Physics Education Research*, 10(2), 020107. <https://doi.org/10.1103/PhysRevSTPER.10.020107>
- Zhang, P., Ding, L., & Mazur, E. (2017). Peer Instruction in introductory physics: A method to bring about positive changes in students' attitudes and beliefs. *Physical Review Special Topics—Physics Education Research*, 13(1), 010104. <https://doi.org/10.1103/PhysRevPhysEducRes.13.010104>