

"Oh, that makes sense": Social Metacognition in Small-Group Problem Solving

Stephanie M. Halmo,[†] Emily K. Bremers,[‡] Sammantha Fuller,[‡] and Julie Dangremond Stanton^{†*}

[†]Department of Cellular Biology and [‡]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602

ABSTRACT

Stronger metacognition, or awareness and regulation of thinking, is related to higher academic achievement. Most metacognition research has focused at the level of the individual learner. However, a few studies have shown that students working in small groups can stimulate metacognition in one another, leading to improved learning. Given the increased adoption of interactive group work in life science classrooms, there is a need to study the role of social metacognition, or the awareness and regulation of the thinking of others, in this context. Guided by the frameworks of social metacognition and evidence-based reasoning, we asked: 1) What metacognitive utterances (words, phrases, statements, or questions) do students use during small-group problem solving in an upper-division biology course? 2) Which metacognitive utterances are associated with small groups sharing higher-quality reasoning in an upper-division biology classroom? We used discourse analysis to examine transcripts from two groups of three students during breakout sessions. By coding for metacognition, we identified seven types of metacognitive utterances. By coding for reasoning, we uncovered four categories of metacognitive utterances associated with higher-quality reasoning. We offer suggestions for life science educators interested in promoting social metacognition during small-group problem solving.

INTRODUCTION

As researchers investigate ways to support life science instructors' use of interactive learning in their classes (Wilson *et al.*, 2018), there is a parallel need to uncover processes that help students fully benefit from these increasing opportunities. Successful interactive group work includes collaboration, or engagement in a coordinated effort to reach a shared goal. Social metacognition, or awareness and regulation of the thinking of others, can increase effective student collaboration during interactive group work (e.g., Kim and Lim, 2018). To support life science students' use of social metacognition during group work, we need to characterize social metacognition in the context of the life sciences. Then we can use our understanding of social metacognition in the life sciences to provide guidance, such as prompts to pose during group work, that helps students fully benefit from opportunities to collaborate with their peers.

In this study, we characterize the unprompted social metacognition life science undergraduates use when they work in small groups to solve problems. We investigate the aspects of their social metacognition that are associated with higher-quality reasoning, which we define as reasoning that is correct, backed by evidence, and generated by more than one individual. Our analysis of student conversation during small-group problem solving draws upon several guiding frameworks. In the following sections, we present relevant background information on the frameworks we use to guide our study of interactive group work, social metacognition, and reasoning.

Stephanie Gardner, *Monitoring Editor*

Submitted Jan 18, 2022; Revised May 13, 2022; Accepted Jul 6, 2022

CBE Life Sci Educ September 1, 2022 21:ar58
DOI:10.1187/cbe.22-01-0009

*Address correspondence to: Julie Dangremond Stanton (stantonj@uga.edu).

© 2022 S. M. Halmo *et al.* CBE—Life Sciences Education © 2022 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 4.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/4.0>).

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

Increasing Adoption of Interactive Group Work

Interactive group work is increasingly being adopted in college science classrooms (Wilson *et al.*, 2018). Having students work in small groups to solve problems helps students develop essential skills, like collaboration, which are valued in the sciences (Kuhn, 2015; National Research Council [NRC], 2015). The adoption of group work also aligns with the view that knowledge construction is a socially shared activity, rather than an individual one. Social cognitive theory, or the idea that learning occurs in a social context from which it cannot be separated (Vygotsky, 1978; Bandura, 1986), forms the foundation behind the promotion of interactive group work.

One framework for studying interactive group work is the ICAP framework, which hypothesizes that learning improves as students' cognitive engagement progresses from passive to active to constructive to interactive, with the deepest level of understanding occurring in the interactive mode (Chi and Wylie, 2014). The interactive mode occurs when students take frequent turns in dialogue with one another by interjecting to ask questions, make clarifications, and explain ideas (Chi and Wylie, 2014). Through this exchange of dialogue, students are able to infer new knowledge from prior knowledge in an iterative and cooperative manner as they take conversational turns. When students work in groups to solve problems, they can employ the social practice of conversation or discourse (Cameron, 2001; Rogers, 2004) to co-construct knowledge in the interactive mode (Chi and Wylie, 2014). As more science instructors implement group work in their courses, we need to better understand how social learning contexts impact important aspects of learning, like metacognition.

Metacognition in Social Learning Contexts

Metacognition, or the awareness and control of thinking for the purpose of learning, is linked to higher academic achievement and can be engaged at the individual or social level. Metacognition is composed of two components: metacognitive knowledge and metacognitive regulation (Schraw and Moshman, 1995). Metacognitive knowledge consists of what one knows about their own thinking and what they know about strategies for learning. Metacognitive regulation consists of the actions one takes to learn, including *planning* strategy use for future learning, *monitoring* understanding and the effectiveness of strategy use during learning, and *evaluating* plans and adjusting strategies based on past learning (Schraw and Moshman, 1995). Metacognition gained prominence in cognitive science and education over the last 50 years because of its relationship to enhanced individual learning (Tanner, 2012; Stanton *et al.*, 2021). For example, students with stronger metacognitive skills learn more and perform better than peers who are less metacognitive (e.g., Wang *et al.*, 1990).

Most research on metacognition has focused at the level of the individual learner. Metacognition was initially conceptualized as an individual process, because discussions on learning were influenced by Piaget's individual-based theory of cognitive development (Brown, 1978). Since then, some researchers have conceived of metacognition more broadly as people's thoughts about their own thinking *and* the thinking of others (Jost *et al.*, 1998). In essence, metacognition includes both individual and social components. Individual metacognition is one's awareness and regulation of one's own thinking for the

purpose of learning, and social metacognition is awareness and regulation of *other's* thinking for the purpose of learning (Stanton *et al.*, 2021). While the theoretical boundaries between individual and social metacognition are clear, distinguishing between the two in practice can be challenging. For example, during small group work, it can be difficult to know whether a student's spoken metacognition is directed inward versus outward (i.e., a reflection of individual vs. social metacognition). However, when a student shares their metacognition in this way, it could potentially stimulate metacognition in another group member. For this reason, we operationally define social metacognition as metacognition that is shared verbally during collaborative work.

Social metacognition, also known as socially shared metacognition, has been explored in just a few disciplines, such as mathematics (Goos *et al.*, 2002; Smith and Mancy, 2018), physics (Lippmann Kung and Linder, 2007; Van De Bogart *et al.*, 2017), and the learning sciences (Siegel, 2012; De Backer *et al.*, 2015, 2020). Social metacognition researchers have focused on identifying the metacognitive "utterances," or words, phrases, statements, or questions, students use during small-group problem solving. Metacognitive utterances are identified through discourse analysis, which is the investigation of socially situated language (Cameron, 2001; Rogers, 2004). From these foundational discourse analyses in other disciplines, a conceptual framework of social metacognition emerged. Social metacognition can happen when students share or disclose their ideas to peers, invite their peers to evaluate their ideas, or evaluate ideas shared by peers (Goos *et al.*, 2002). Social metacognition also occurs when students enact, modify, or assess their peers' strategies for problem solving (Van De Bogart *et al.*, 2017).

Research on social metacognition in other disciplines has shown that students working in small groups can stimulate metacognitive processes in one another, leading to improved learning. For example, researchers found that one variation in social metacognitive dialogue, which they called "interrogative" (i.e., evoked by a thought-provoking trigger and generally followed by elaborative reactions), was positively related to college students' individual performance on a learning sciences knowledge test (De Backer *et al.*, 2020). Middle school students who came up with a correct solution as a group had higher levels of metacognitive interactions or made more metacognitive utterances during group problem solving (Artz and Armour-Thomas, 1992). Support for this finding was provided by a comparison of successful versus unsuccessful problem solving in a high school math class. Successful problem solving (i.e., working together as a group to come to a correct solution on a math problem) involved students assessing one another's ideas, correcting incorrect ideas, and endorsing correct ideas, while during unsuccessful problem solving, students lacked critical engagement with one another's thinking (Goos *et al.*, 2002).

Although metacognitive utterances have been identified in a few disciplinary contexts, the metacognitive utterances that college students use during small-group problem solving in the life sciences has yet to be documented. Social cognitive theory posits that learning is socially situated, meaning it is specific to the context and social environment in which it is embedded (Bandura, 1986). This means learning is not easily transferable from one context to another. For example, the nature of metacognition that occurs in a high school calculus

class using problem-based learning could differ from that which occurs in a college biology class using process-oriented guided inquiry learning (POGIL). Additionally, just because a student can use metacognition in their calculus class does not necessarily mean they will employ the same metacognition in their biology class. Given that the metacognition students use can differ based on context, defining the metacognitive utterances life science majors use during small-group problem solving is an important first step for understanding social metacognition in the life sciences. We can then use this understanding to provide guidance to students as they work together in groups.

Social metacognition has also been linked to reasoning. In physics labs, metacognitive utterances impacted learning behavior by helping students transition from logistical to reasoning behavior. For example, a metacognitive utterance helped students transition from recording data (logistical behavior) to assessing their experimental design (reasoning behavior). The action the group takes after metacognitive utterances seems to be what matters most for successful problem solving in physics labs (Lippmann Kung and Linder, 2007). Research in grade school mathematics classrooms indicated a positive association between metacognitive talk and transactive talk, or reasoning that operates on the reasoning of another. Results from this study suggest that metacognitive talk is more likely to be preceded or followed by reasoning (Smith and Mancy, 2018). These promising results indicate that social metacognition is associated with improved reasoning in other disciplinary contexts.

Reasoning in Social Learning Contexts

The skill or practice of scientific reasoning is a valued outcome of science education and a focus of major science education reform efforts (NRC, 2007; American Association for the Advancement of Science, 2011). Scientific reasoning reflects the disciplinary practices of scientists and can create a more scientifically literate society. Scientific reasoning is the process of constructing an explanation for observed phenomena or constructing an argument that justifies a claim. Scientific reasoning skills include identifying patterns in data, making inferences, resolving uncertainty, coordinating theory with evidence, and constructing evidence-based explanations of phenomena and arguments that justify their validity (Osborne, 2010).

One important framework for reasoning is Toulmin's argument pattern. Toulmin described an argument as the relationship between a claim, the information that supports the claim, and an explanation for why the claim flows logically or causally from the information (Toulmin, 2003). Toulmin's argument pattern framework is domain general in nature and therefore does not include an assessment of whether an argument is coherent or accurate. Many studies have used Toulmin's argument pattern to guide analysis of scientific reasoning in group discourse (Osborne *et al.*, 2004; Sampson and Clark, 2008; Knight *et al.*, 2013, 2015; Paine and Knight, 2020). A key adaptation of Toulmin's argument pattern is the evidence-based reasoning framework (Brown *et al.*, 2010).

The evidence-based reasoning framework is "intended to help researchers and practitioners identify the presence and form of scientific argumentation in student work and classroom discourse" (Brown *et al.*, 2010, p. 134). The evidence-based reasoning framework draws distinctions between the

component parts of scientific reasoning by identifying claims, premises, rules, evidence, and data (Supplemental Figure 1). A "claim" is a statement about a specific outcome phrased as either a prediction, observation, or conclusion. A "premise" is a statement about the circumstances or input that results in the output described by the claim. A "rule" is a statement describing a general relationship or principle that links the premise to the claim. "Evidence" is a statement about an observed relationship, and "data" are reports of discrete observations. Together, rules, evidence, and data can be considered forms of backing (Furtak *et al.*, 2010). For example, as someone is driving, they may think, "This traffic light is yellow. Yellow lights quickly turn to red, so I will slow down." In this example, the statement "This traffic light is yellow" is a premise, "Yellow lights quickly turn to red" is backing (specifically, a rule), and "I will slow down" is a claim. The evidence-based reasoning framework suggests that more sophisticated scientific reasoning occurs when students make a claim supported by backing (Brown *et al.*, 2010; Furtak *et al.*, 2010). Reasoning quality is relative and likely occurs on a qualitative continuum from lower to higher quality. One of our goals was to characterize this reasoning quality continuum. In this study, we use the evidence-based reasoning framework as a starting point to identify instances of complete reasoning during student discourse or conversation. We expand the continuum so that higher-quality reasoning also includes the consideration of correctness and the transactive nature of the reasoning (i.e., whether the reasoning was generated by more than one individual).

Research Questions

With the increasing adoption of interactive group work in undergraduate life science classrooms, there is a need to study the role of social metacognition, or metacognition that occurs out loud in a social learning context, and its relationship to reasoning. To address this gap, we used discourse analysis of small-group problem solving from an upper-division biology course to address the following qualitative research questions:

1. What metacognitive utterances do students use during small-group problem solving in an upper-division biology course?
2. Which metacognitive utterances are associated with small groups sharing higher-quality reasoning in an upper-division biology classroom?

METHODS

Context and Data Collection

This study was conducted at a large, public, research-intensive university in the southeastern United States. Participants were recruited from an upper-level cell biology course taken by life science majors in 2018. Average enrollment in this course was ~80 students per section. The course consisted of an interactive lecture 3 days a week and a smaller breakout session once per week (~40 students per breakout session). The breakout sessions were held in a SCALE-UP classroom designed to facilitate group work with multiple monitors and round tables where students sat in groups of three (Beichner *et al.*, 2007). During weekly breakout sessions, students worked in small groups of three to solve problem sets in-person using a pen-and-paper format.

The problem sets were designed using guided-inquiry principles (Moog *et al.*, 2006). The problems scaffolded student learning about a cell biology concept and asked students to analyze relevant published scientific data. The first problem set covered import of proteins into the nucleus (“nuclear import”), and the second problem set focused on transcription. The breakout session problem sets were formative assessments and were not letter graded but were highly aligned to the learning objectives and the exams in the course. Approximately 40% of the exam points covered material from the breakout sessions. In lieu of being letter graded, a graduate teaching assistant provided written feedback to the groups on their completed problem sets, similar to the feedback that would be provided if the problem set was an exam.

To form groups during the breakout sessions, participants were allowed to pick their own group of three or they could opt to be randomly assigned to a group. The groups of three did not change during the course of the study. In each group, there were three roles: manager, presenter, and recorder (Stanton and Dye, 2017). The recorder was responsible for writing and turning in the group’s answers for a participation grade and feedback. The presenter was responsible for sharing group results during the whole-class discussion at the end of the breakout session and for sharing group solutions on dry-erase boards during the breakout session. The manager was responsible for keeping the group on task during the allotted time. Group roles were randomly assigned at the start of each breakout session and rotated week to week. The study was classified by the University of Georgia’s Institutional Review Board as exempt (STUDY00006457).

Four groups of three students each agreed to be audio-recorded during two consecutive breakout sessions. Each participant was compensated \$20 for participation in the study, and all participants provided written consent. To accurately record individuals in a group setting, each group member was individually microphoned using 8W-1KU UHF Octo Receiver System equipment (Nady Systems, Inc.). After the audio recordings were collected, the individual recording tracks were synced and merged into one recording per group per breakout session using Steinberg Cubase software.

The audio recordings were professionally transcribed (Rev.com), and the transcripts were checked to ensure accuracy before analysis. The following transcription conventions were used: 1) speaker turns were arranged in vertical format, with all speaker turns arranged in a single column one above another to reflect the equal status of each speaker as students; 2) utterances ending in a sharp rising intonation were considered questions and were signified with a question mark (?); 3) a single dash (-) following a word was used to indicate interrupted, truncated, or cut-off words or phrases; 4) an ellipsis (...) was used to indicate pauses in speech or when a speaker trailed off; and 5) when available, researchers provided interpretations of nonspecific pronouns (e.g., this and that), which are indicated in brackets ([]) (Du Bois *et al.*, 1992; Edwards, 2005). We did not use transcription conventions to signify overlapping stretches of speech, but these instances were coded. Participant names were changed to pseudonyms in the transcripts. The transcripts and accompanying audio serve as the primary data for this study.

Here, we report on data from two of the four groups during two consecutive breakout sessions. The data from the third group were excluded from analysis because one group member was absent the day of the second recording. The data from the fourth group were excluded from further analysis because the group spent a significant amount of time looking through their notes and reading them to one another rather than discussing their thoughts about the problem set. Our rationale for this decision was that we could not study social metacognition if it was not evident. We will refer to the two groups analyzed in this study as Group A and Group B. The groups attended separate breakout sessions on the same afternoons. Group A consisted of three women: Bella, Catherine, and Michelle, who elected to work together. Group B consisted of one woman and two men: Molly, Adam, and Oscar, who were assigned to work together. The group roles of the participants for each problem set can be found in Supplemental Table 1. Our sample size, while small, is in alignment with sample sizes for foundational discourse analysis (Cameron, 2001; Rogers, 2004). For example, one study on social metacognition involved the analysis of one pair of students working on a single physics laboratory problem (Van De Bogart *et al.*, 2015).

Timeline Creation and Analysis of Silence

Each transcript was analyzed to create a timeline of each breakout session. Start and end times of on-task work, which we defined as students directly working on the problem set, and off-task work, which we defined as students discussing ideas unrelated to the problem set, were recorded. Next, two researchers (S.M.H. and E.K.B.) listened to all transcripts and recorded the start and end time of all silences equal to or greater than 5 seconds in length. The duration of the silences was summed for each transcript. Using this information, a percentage of time spent in silence was calculated for each transcript as follows:

$$\% \text{ Time in silence} = \frac{\text{Total time in silence}}{\text{Total time on task}} \times 100$$

Qualitative Discourse Analysis

Transcripts of student discourse were analyzed by a team of researchers using MaxQDA 2020. Our basic unit of analysis was an utterance. An utterance is either a word, phrase, statement, or question that an individual or group of students makes while collaborating. A single line of speech from a single student could contain multiple utterances. For example, one student could say, “Yeah, I don’t know. What do you think?” and this could be broken into three separate utterances with “Yeah” as a single word utterance, “I don’t know” as a phrase utterance, and “What do you think?” as a question utterance. Alternatively, multiple lines of speech from multiple students could compose a single utterance. For instance, when one student interrupts another to complete the other’s thought, the combined statement could be considered an utterance composed by two individuals. Our qualitative discourse analysis of these utterances occurred in multiple, iterative cycles.

First-Cycle Coding. First-cycle coding began with open, initial coding of all the transcripts (Saldaña, 2021). The goal of our initial coding process was to begin identifying the utterances from our data set that related to social metacognition and

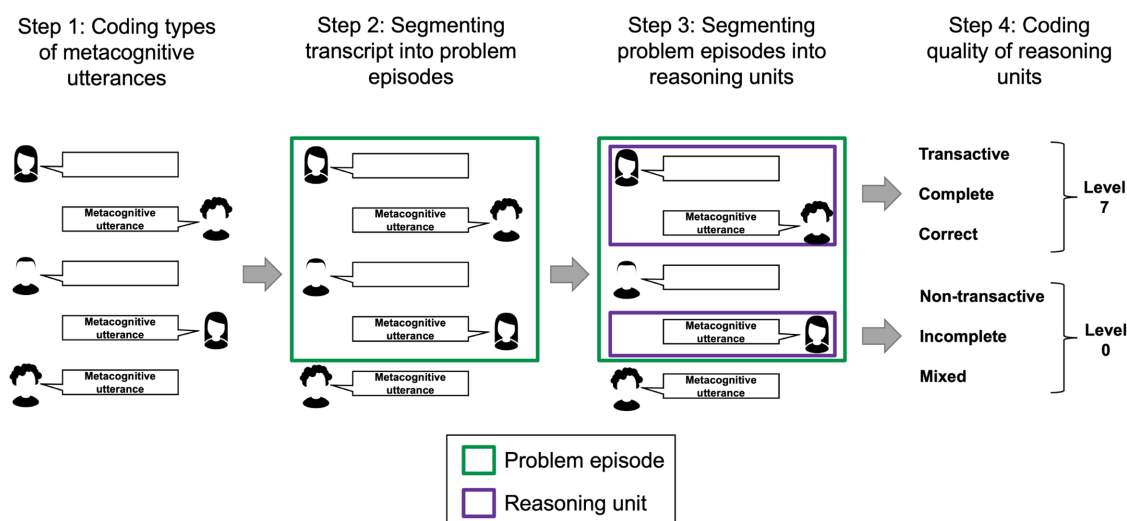


FIGURE 1. Schematic of dual-coding process. In step 1, types of metacognitive utterances throughout each transcript were coded. As a simplified example, the image under step 1 depicts five lines of a single transcript, three of which contain a metacognitive utterance. Metacognitive utterances are words, phrases, statements, or questions students made that were related to their awareness and control of thinking for the purposes of learning. In step 2, each transcript was segmented into problem episodes. Problem episodes consisted of sections of the transcript in which students were solving problems that dealt with cell biology data, e.g., “Explain the results in the YRC panel.” The image under step 2 shows that the first four lines of the example transcript make up a single problem episode (green box). In step 3, the problem episodes were further segmented into reasoning units. Reasoning units consisted of a chunk of discourse in which a student or students were discussing a single collection of connected ideas. The image under step 3 shows that the first two lines of the example transcript make up one reasoning unit, and the fourth line is a separate reasoning unit (purple boxes). In step 4, each reasoning unit was assigned a reasoning code (Supplemental Table 2). The image under step 4 shows the first reasoning unit was assigned with the level 7 code, and the second reasoning unit was assigned with the level 0 code (Supplemental Table 2). Attribution for images: Profile and Profile Woman by mikicon from the Noun Project.

reasoning. Two coding schemes were developed to investigate our research questions. The first codebook was developed to capture metacognitive utterances from student interactions. The second codebook was developed to capture the reasoning quality present in the discussion. This dual-coding methodology meant utterances could be double coded as both metacognitive and as a part of reasoning. Data were first coded as metacognitive utterances and then coded for reasoning quality (Figure 1). We discuss the development of each codebook in the following sections.

Social Metacognition. All four authors coded the transcripts for social metacognition, wrote analytic memos, and then met to discuss emergent ideas. Deductive codes originated from prior work on social metacognition: self-disclosure, feedback request, and other monitoring (Goos *et al.*, 2002). We developed inductive, or emergent, codes based on the utterances present in our data and our knowledge of metacognition as a construct. We refined these codes through discussion, listening to the audio, and careful consideration of which codes aligned with our research questions. Two researchers (S.M.H. and E.K.B.) then coded all four transcripts individually and subsequently met as a team to discuss how each researcher applied the codes. These discussions led us to add, remove, or redefine our existing codes, further refining the codebook. Through these iterations, our codebook stabilized. We then revisited segments of the data that were selected for reasoning analysis and coded them to consensus with the stabilized social metacognition codebook until all discrepancies were resolved. Attribute coding, or the

notation of participant characteristics (Saldaña, 2021), was also employed to take note of the group roles during the first-cycle coding for social metacognition. For example, in the nuclear import transcript for Group A, every line from Michelle was coded as a line from the manager, which was her assigned group role that day.

Reasoning. A reliable, systematic methodology for 1) identifying reasoning and 2) assessing its quality was needed. Once sections of the transcripts related to reasoning about scientific data were identified in initial coding, the transcripts were broken into problem episodes. Problem episodes consisted of all utterances in a transcript in which students discussed an answer to a specific problem in the problem set (Figure 2, green boxes). For example, there was one problem episode from Group A for problem 2 in the transcription problem set. Given the emphasis on backing (rules, evidence, and data) in the evidence-based reasoning framework (Supplemental Figure 1), we selected only those problems that required students to analyze a diagram or data figure (questions 2 and 4 for the nuclear import problem set and questions 2, 3, and 5 for the transcription problem set).

Problem episodes were further parsed into reasoning units. Reasoning units are defined as a conversational chunk of discourse in which a student or students are discussing a single collection of connected ideas. Two researchers (S.M.H. and E.K.B.) then evaluated each reasoning unit against a list of a priori codes consisting of structural reasoning components (premise, claim, backing) derived from the evidence-based reasoning framework (Brown *et al.*, 2010; Furtak *et al.*, 2010) and

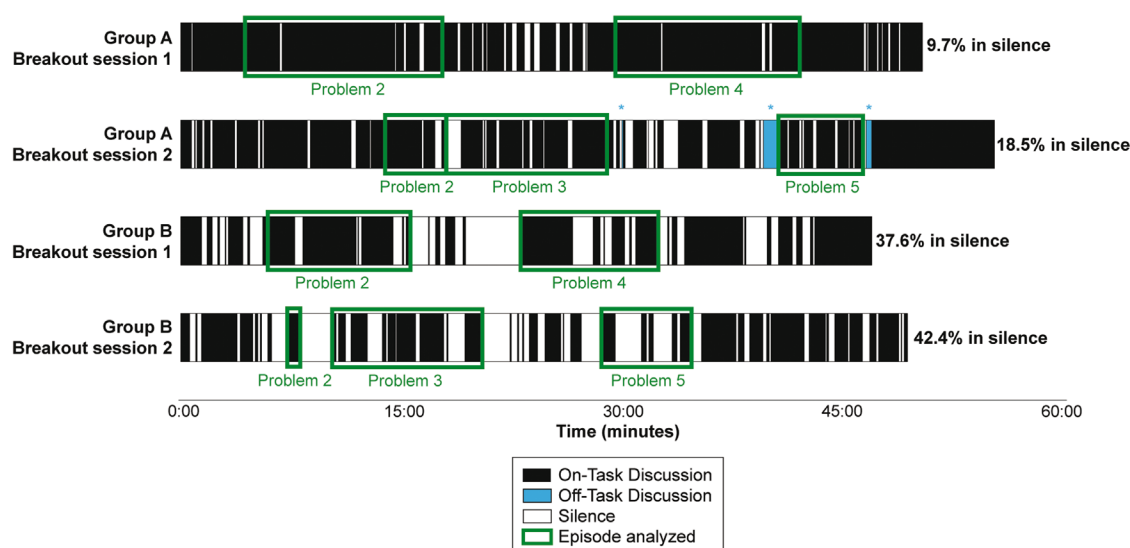


FIGURE 2. Timelines of small group work. Group work timelines are represented as bar graphs with time (in minutes) on the x-axis. The timelines show how each group spent each breakout session. White sections of the timeline indicate when groups were silent. Black sections of the timeline indicate when groups were discussing the problem set. Blue sections of the timeline indicate when groups were discussing ideas unrelated to the problem set. Problem episodes selected for reasoning analysis are indicated on the timelines in green boxes. Groups did not finish the problem set in the same amount of time. For example, Group B finished the problem set for breakout session 1 early, whereas Group A did not complete the problem set in the allotted class time for breakout session 2.

correct and incorrect scientific ideas. The evidence-based reasoning framework does not account for accuracy of scientific information, but it was critical in our analysis of the data to consider accuracy (correct vs. incorrect scientific ideas) of the structural reasoning components. Another researcher (J.D.S.) provided insight into what counted as backing and correct scientific ideas throughout this process because of their expertise in the course content and context. We undertook this part of first-cycle coding together and discussed all discrepancies until consensus was reached.

Second-Cycle Coding. Second-cycle coding began with establishing reasoning quality codes using our first-cycle codes for reasoning. The purpose was to assess the quality of reasoning that was occurring across three dimensions: a reasoning unit's 1) transactive nature, 2) completeness, and 3) correctness. First, reasoning units were either transactive, meaning two or more students participated in reasoning and one of those students clarified, elaborated, or justified the reasoning of another student(s); or they were nontransactive, in which individual students share their reasoning but their reasoning is not operated on by another student (Kruger, 1993). Second, the reasoning units were also either complete or incomplete. Complete reasoning units were defined by the presence of at least one, clear claim, a premise, and some form of backing that connected the premise to the claim through data, evidence, or a rule (Brown *et al.*, 2010; Furtak *et al.*, 2010). Incomplete reasoning units were defined as having a claim and/or a premise but lacked a form of backing such as data, evidence, or a rule. Third, the reasoning units were either correct or mixed. A correct reasoning unit could either solely consist of correct scientific ideas, or it could contain an incorrect scientific idea, as long as that incorrect idea was ultimately corrected within the rea-

soning unit. In contrast, a mixed reasoning unit contained both correct and incorrect scientific ideas, but the incorrect ideas were never corrected. We chose the word "mixed" instead of "incorrect," because no reasoning units were wholly incorrect.

Combining these three binary parameters resulted in eight codes that were ordered by first prioritizing transactive over nontransactive behavior (Supplemental Table 2). This decision to view transactive behavior as more beneficial than nontransactive behavior aligns with the continuum outlined in the ICAP framework (Chi and Wylie, 2014) and the view that exchanges of reasoning are of higher quality (Knight *et al.*, 2013). The ordering of the reasoning codes also reflects our decision to prioritize completeness of reasoning units over correctness. We made this choice, because it may be easier for instructors and students to correct ideas rather than to push students to reason with backing. This prioritization and ordering resulted in the view that transactive, complete, and correct reasoning units represent higher-quality reasoning, and nontransactive, incomplete, and/or mixed reasoning units represent lower-quality reasoning. These mutually exclusive codes were then applied to predefined reasoning units in selected problem episodes of the transcripts (Figure 1).

To identify the metacognitive utterances that co-occurred with higher-quality reasoning, we examined the results from our dual-coding process, both metacognitive utterance types and reasoning codes, together in the last phase of the second cycle. We relied on pattern coding as our selected second-cycle coding method to uncover categories in our dual-coded data (Saldana, 2021). Pattern coding is a way to group the results from first-cycle coding into larger categories. Specifically, we gathered all reasoning units with higher-level codes (level 6 and level 7) and looked for patterns among the metacognitive utterances in these higher-quality reasoning units. We investigated

both level 6 and level 7 reasoning units during pattern coding, because these reasoning units were both transactive and complete. The only difference was that level 6 reasoning units contained mixed ideas. This process was facilitated by complex code querying in MaxQDA 2020. In addition to first- and second-cycle coding, we also relied on our analytic memos about the data to inform our coding decisions (Saldaña, 2021).

RESULTS

We first present an overview of how the two groups spent their time during the breakout sessions to provide the reader with important context for the analysis that follows. Next, we demonstrate the results from our dual-coding scheme. To address our first research question (What metacognitive utterances do students use during small-group problem solving in an upper-division biology course?), we define the metacognitive utterances students used during small-group problem solving. Then we provide an analysis of reasoning that occurred during small-group problem solving, which is required to answer our second research question. Finally, by tying our two coding schemes together, we address our second research question (Which metacognitive utterances are associated with small groups sharing higher-quality reasoning in an upper-division biology classroom?) by presenting the metacognitive utterances that were associated with higher-quality reasoning in our data set.

How Students Spent Their Time during Group Work

Students spent the majority of the breakout sessions working directly on the problem set. Little to no time was spent off-task discussing ideas unrelated to the problem sets. Group A was rarely off-task and Group B was never off-task (Figure 2). Group B spent more of their time in silence. On average, Group B spent 40% of their working time in silence, whereas Group A spent 14% of their working time in silence. Despite spending a larger percentage of their time in silence, Group B completed the problem sets faster than Group A. When Group B finished early during the first breakout session, they spent the remainder of the group work time getting to know one another, because they had never met. In contrast, the members of Group A knew one another and had previously met. Overall, both groups were on-task and focused on the problem sets during the breakout sessions but approached their work together differently.

What Metacognitive Utterances Do Students Use during Small-Group Problem Solving?

We identified several types of metacognitive utterances that upper-division biology students used during group work throughout the breakout sessions analyzed in this study. Metacognitive utterances are words, phrases, statements, or questions students made that were related to their awareness and control of thinking for the purposes of learning. The metacognitive utterances we identified included “planning,” “statements to monitor understanding,” “corrections of another student,” “questions to monitor understanding,” “requests for information,” “evaluations of self,” and “evaluations of others,” which can all be mapped to the individual metacognitive regulation skills of planning, monitoring, and evaluating (Table 1). While metacognitive utterances related to planning may play an important role in small-group problem solving, they did not

directly impact reasoning in this data set and thus were not investigated further.

Metacognitive Utterances Related to Monitoring. The individual metacognitive regulation skill of monitoring involves assessing one’s understanding of concepts while learning (Stanton *et al.*, 2021). The metacognitive utterances related to monitoring that we identified in our data took the form of both statements (statements to monitor understanding and corrections of another student) and questions (questions to monitor understanding and requests for information) (Table 1). These metacognitive utterances are related to monitoring, because they involve assessing either one’s own or a group member’s understanding of concepts. For example, statements to monitor understanding included assessments of one’s own conceptual understanding through self-corrections or self-explanations, whereas corrections of another student involved assessments of a group member’s conceptual understanding.

Students also used questions to monitor understanding to assess their knowledge of concepts. These questions helped them clarify their own understanding or their group members’ understanding of a concept. These questions were closed in nature, meaning they could be answered with a simple one-word response like “yes,” “no,” or “correct.” Questions of this type included follow-up questions, questions to make sure group members were following along, or requests for confirmation on a shared idea (Table 1). In contrast to the closed nature of questions to monitor understanding, we also found evidence of students using more open-ended questions to request access to their group mates’ thinking (requests for information). Requesting access to a group mates’ thinking is related to metacognition, because one must first be aware of the thinking of others in order to act on it or regulate it. Students would make requests for information when they asked group mates to disclose their knowledge and information about a concept beyond a simple yes or no question. These requests for information only occurred when a student was asking for information they had not already supplied themselves and often centered around an interrogative word such as “who,” “what,” “when,” “where,” “why,” or “how” (Table 1).

Metacognitive Utterances Related to Evaluating. The individual metacognitive regulation skill of evaluating involves appraising one’s plan or approach for learning (Stanton *et al.*, 2021). Two of the metacognitive utterances we identified, evaluations of self and evaluations of others, are related to evaluating, because they involve appraising either one’s own or a group member’s thinking or approach and whether it is effective or relevant to the problem. Evaluations of others took the form of both statements and questions (Table 1). When in statement form, evaluations of others were matter-of-fact critiques that appraised the group’s current solution to a problem in the problem set, like Oscar’s statement “I don’t know how you can even say that there’s DNA present. If it wasn’t immunoprecipitated, it would’ve been washed away.” Audio does not reveal how Oscar’s group mates felt about this critique, nor does Oscar’s statement explicitly invite his group members to engage with his critique. In contrast, evaluations of others that were in question form requested some sort of engagement with the appraisal from the group. Evaluations of others posed as questions either

TABLE 1. Types of metacognitive utterances in small-group problem solving

Individual metacognitive regulation skill	Metacognitive utterance	Description	Example
Planning	Planning	When a student talked about how they wanted to go about completing the problem set in terms of time allocation or resource use	<i>Okay, let's see. How many questions are there total? One, two, three, four, five, and we need to be done by 3:00. So, 50 minutes I guess we could spend around like 10 minutes on each.</i>
Monitoring	Statements to monitor understanding	When a student shared what they were thinking out loud to make sure they understood a concept without prompting	<i>Oh. That makes sense. Because in the nucleus, we have Ran-GTP, so in the nucleus, it's in this form.</i>
	Corrections of another student	When a student directly corrected a perceived incorrect statement from another group member	Student 1: <i>Okay. So RanGAP is found in the nucleus-</i> Student 2: <i>In the cytoplasm.</i> Student 1: <i>Yeah, sorry. Cytoplasm.</i>
	Questions to monitor understanding	When a student asked their group member(s) a yes or no question to clarify an idea or approach	<i>Do y'all understand?</i>
	Requests for information	When a student asked their group member(s) to share knowledge about an idea or approach beyond a yes or no question	<i>Okay. But how does that affect the concentration gradients?</i>
Evaluating	Evaluation of self	When a student assessed their own thinking, approach, or solution and whether or not it was effective or relevant	<i>So, yeah, I shouldn't make assumptions. I should stay within the realm of the question.</i>
	Evaluation of others	When a student assessed the thinking, approach, or solution shared by a group member and whether or not it was effective or relevant	<i>Does it answer the question why?</i>

invite pauses for clarification, reorient the group back to the problem set, or challenge an idea, all of which can more directly impact reasoning. For example, Michelle's inquiry, "Does it answer the question why?," reoriented the group back to the question posed in the problem set.

The frequency of the metacognitive utterance types we identified by individual participant is provided in Table 2. The most frequently used metacognitive utterance types were questions to monitor understanding, evaluation of others, statements to monitor understanding, and requests for information.

Reasoning in Small-Group Problem Solving

To answer our second research question regarding the metacognitive utterances associated with higher-quality reasoning, we needed to first analyze the quality of reasoning that occurred when participants solved problems in small groups. Guided by the evidence-based reasoning framework, we identified reasoning components in our data and rated them using the reasoning coding scheme described in the *Methods* (Supplemental Table 2). Overall, the reasoning displayed by both groups in the problem episodes analyzed (green boxes in Figure 2) was high in quality (Supplemental Table 2). Group A had nearly double the number of reasoning units compared with Group B (Supplemental Table 2), which aligns with the finding that Group A spent more time talking compared with Group B (Figure 2). In this section, we use examples from participant discourse to illustrate the reasoning coding scheme (Supplemental Table 2) by presenting an example of lower-quality reasoning and then an example of higher-quality reasoning. For each example, we share an excerpt of group discourse with a line-by-line analysis of the discourse dynamics

and then offer an analysis of the reasoning units using the coding scheme (Russ et al., 2008).

An Example of Lower-Quality Reasoning. In the segment of discourse presented in Figure 3, Group A is solving problem 2b in the transcription problem set. They are discussing why it is important to control fragment length in a chromatin immunoprecipitation (ChIP) protocol. The discourse in Figure 3 is one conversation composed of two reasoning units (purple boxes), both of which are examples of lower-quality reasoning. Lower-quality reasoning units are nontransactive, incomplete, and/or mixed, meaning students are not operating on one another's reasoning, they do not provide backing to support linking their premises to a claim, and/or incorrect ideas are present.

Reasoning Unit Analysis. In Figure 3, there are two reasoning units (purple boxes). The first reasoning unit is nontransactive, because no group member directly acts on the reasoning Catherine shares in line 1. On the surface, it may appear as though Michelle and Bella are providing Catherine with confirmation on her reasoning in lines 2 and 3; however, they are not elaborating on the content of Catherine's reasoning. Therefore, this reasoning unit is considered nontransactive. In terms of structure, Catherine shares a claim about the regulatory proteins breaking or splitting in half. She also shares two premises about 1) the fragments being too small and 2) the goal of ChIP as figuring out "where the proteins bind to DNA." Given that Catherine's reasoning lacks backing in the form of data, evidence, or a rule, her reasoning in this first unit is incomplete. Additionally, Catherine's ideas about the proteins breaking or splitting in half are incorrect. Thus, this first reasoning unit is also mixed.

TABLE 2. Frequency of metacognitive utterances by individual group members^a

Individual metacognitive regulation skill	Metacognitive utterance	Group A			Group B			Total utterances by type
		Bella	Catherine	Michelle	Adam	Molly	Oscar	
Planning	Planning	4.3%	13.2%	8.0%	0.0%	2.8%	0.0%	15
Monitoring	Statements to monitor understanding	17.4%	11.3%	16.0%	31.3%	19.4%	23.5%	46
	Corrections of another student	1.4%	7.5%	4.0%	6.3%	16.7%	11.8%	18
	Questions to monitor understanding	43.5%	37.7%	24.0%	12.5%	25.0%	11.8%	77
	Requests for information	20.3%	0.0%	10.0%	0.0%	5.6%	11.8%	25
Evaluating	Evaluation of self	1.4%	1.9%	4.0%	0.0%	13.9%	2.9%	10
	Evaluation of others	11.6%	28.3%	34.0%	50.0%	16.7%	38.2%	67
	Total utterances by individual	69	53	50	16	36	34	

^aThe percentages were calculated based on the number of each utterance type made by an individual divided by the individual's total number of metacognitive utterances. The cell with each individual's most frequently used type of metacognitive utterance is highlighted in dark purple. The cell with each individual's second most frequently used type of metacognitive utterance is highlighted in light purple.

Taken together, the first reasoning unit in this excerpt of discourse was coded as the lowest-quality reasoning (level 0: non-transactive, incomplete, mixed reasoning).

The next reasoning unit in Figure 3 begins with Bella asking Catherine to repeat what she said previously and ends with Bella's decision about what to write down for this problem and to move on to the next problem in the set. At the start of this unit, Bella is interested in what Catherine said earlier and asks Catherine to repeat herself. Catherine reshapes her incorrect claim, and Bella presents her own incorrect claim. Catherine extends her own reasoning, and Michelle agrees with it. Bella asserts her own incorrect claim again. Catherine agrees with Bella in words, yet continues to extend the incorrect claim that she and Michelle agree on. Bella does not write down Catherine and Michelle's incorrect claim, but moves the group on to the next problem. Despite the fact that Bella initially elicits reasoning from Catherine, she does not seem to be acting on the reasoning shared by her peers. Michelle acts on Catherine's reasoning once, and Catherine appears to act on Bella's reasoning twice by confirming and agreeing with her ideas, albeit superficially. Because they are acting on one another's reasoning, this second reasoning unit is considered transactive. Given the presence of incorrect ideas that go uncorrected by peers and that Group A is solely sharing premises and claims in this section, this unit is considered incomplete and mixed (level 4: transactive, incomplete, mixed reasoning). This is an improvement from Catherine sharing her siloed reasoning in the first unit, but there were missed opportunities to constructively reason as a group (Figure 3).

An Example of Higher-Quality Reasoning. In the segment of discourse presented in Figure 4, Group B is solving problem 4a in the nuclear import problem set. They are considering fluorescence resonance energy transfer (FRET) data from a published figure and are tasked with explaining the results in one of the figure's panels. The discourse in Figure 4 is an example of higher-quality reasoning. Higher-quality reasoning units are transactive, complete, and correct, meaning students are operating on one another's reasoning, they provide backing to link their premises to a claim, and no incorrect ideas are present, or if they are, they are ultimately corrected.

Reasoning Unit Analysis. In the first reasoning unit shown in Figure 4, Oscar shares his idea that something in the figure only

makes sense if Ran-GTP is bound. Molly verbalizes that she was thinking the same thing as Oscar and explains what should be occurring if Ran-GTP is not bound in terms of fluorescence output for FRET, and she ties that to what the observed ratio should then be. Adam questions Molly's reasoning, pondering if both the yellow fluorescent protein (YFP) and cyan fluorescent protein (CFP) molecules should fluoresce when Ran-GTP is not bound. This question triggers Molly to share more of her reasoning, including an accurate description of how FRET works. Oscar then clarifies her idea by stating that you would get *more* YFP, not necessarily only YFP emission when FRET works. Molly agrees with Oscar, then Oscar ties their co-constructed prediction to the schematic representation of FRET provided in the problem. Molly then connects their reasoning to the observation that the cytoplasm is green in color, which corresponds to a ratio value greater than one. Group B's reasoning in this unit is transactive in nature, because they are exchanging reasoning. They are displaying complete reasoning by 1) sharing what they know about Ran-GTP concentrations from the prompt (premises) and 2) providing backing for their claims in the form of color observations (data) and ratio relationships (evidence). Their reasoning is also correct (level 7: transactive, complete, correct reasoning).

In the second reasoning unit shown in Figure 4, Group B follows a similar structure for their reasoning displayed in the first reasoning unit, but this time for what is happening in the nucleus. Molly shares her conclusion about YFP and CFP intensity in the nucleus. Oscar elaborates on her conclusion by adding in his prior knowledge about the location and concentration of Ran-GTP in the nucleus, and Adam brings in the observation that the nucleus should then be very dark in color. Group B's reasoning is transactive in nature, because they are acting on and building upon one another's ideas to co-construct shared reasoning. As Group B discusses this problem, they consistently provide backing for their claims in the form of data (color observations) or evidence (ratio relationships) and premises, including what they know about Ran-GTP concentrations and relevant pieces of information shared in the problem prompt. For this reason, their reasoning is complete. Additionally, their reasoning is also correct (level 7: transactive, complete, correct reasoning).

Both reasoning units in Figure 4 were coded as transactive, complete, and correct, which represent the highest-quality reasoning code (level 7: transactive, complete, correct reasoning). As is evident in the example discourse, students in our study

Reasoning Unit	Line	Discourse (conversation)	Analysis
Level 0 – Non-transactive, incomplete, mixed	1	<i>Catherine: Well, isn't it important so they don't break these regulatory proteins or like ... so, you're trying to figure out where the proteins bind to DNA. If they're too short, it may split it in half, and you may have an overabundance or a non-representative number, is that something that could happen?</i>	Catherine shares her reasoning about why it is important to control fragment length in a ChIP protocol. She phrases her reasoning in the form of a question to seek confirmation from her group members. Her reasoning consists of an incorrect claim and premises.
	2	<i>Michelle: Yeah, she also said something about PCR.</i>	Michelle seemingly provides confirmation for Catherine's reasoning, but instead of acting directly on Catherine's reasoning, she redirects by suggesting the answer has something to do with PCR.
	3	<i>Bella: Yeah. I wrote the steps is, first, you cross-link. Then, you fragment. Then, you immunoprecipitate. What does protein B bind to. After you cross-link, you have to add antibody to protein B, and then you immunoprecipitate everything that's bound to protein B. Then, you add a secondary antibody that brings down sequences bound to protein B, and then you're supposed to reverse the cross-links, or else it won't be able to add the primer, because DNA is bound by protein. Then, you amplify by PCR. If the DNA is amplified by PCR, then that indicates that your protein was bound to that sequence. Then, ChIP just basically identifies regions of DNA where a protein is bound. But for fragment length I don't know. What were you saying, Catherine?</i>	Bella shares the steps in a ChIP protocol that she has written in her notes from class. She then acknowledges that she does not know why it is important to control fragment length in a ChIP protocol and requests Catherine to share her reasoning again.
Level 4 – Transactive, incomplete, mixed	4	<i>Catherine: I was saying that we don't want the fragments to be ... so, in this step, you're breaking the DNA into small fragments around 300 nucleotides. You don't want it too small, or else it's not gonna be representative, and it will split them in half and show a higher signal.</i>	Catherine responds to Bella's request and reshapes her initial reasoning from line 1.
	5	<i>Bella: I was gonna say, maybe if you break it too small, the protein won't be able to bind.</i>	Bella offers an alternative incorrect claim.
	6	<i>Catherine: That, too. Yeah, and if it's too big, you're not really determining where exactly the location is or where exactly they're binding.</i>	Catherine confirms Bella's alternative incorrect claim, and offers an additional premise and claim. This time her claim is correct.
	7	<i>Michelle: Yeah, I think it's so you can better pinpoint where protein binding is happening.</i>	Michelle confirms Catherine's correct claim by restating it.
	8	<i>Bella: So, controlling fragment length is necessary so you can ensure that the protein is able to properly bind to that DNA sequence.</i>	As recorder for the group, Bella attempts to establish consensus by restating the reasoning she originally shared in line 5, which Catherine confirmed in line 6. Notably, she has not engaged with the new reasoning shared by Catherine and Michelle in lines 6 and 7.
	9	<i>Catherine: Yeah, and also so you can figure out exactly where it's binding.</i>	Catherine again reconfirms Bella's incorrect reasoning and tries to add the correct reasoning that she originally shared in line 6.
	10	<i>Bella: I just wrote, "So the protein's able to properly bind to the DNA sequence."</i>	Bella reads out loud what she wrote down for the problem. Her reasoning is incorrect. It does not incorporate the correct reasoning shared by Catherine that was confirmed earlier by Michelle.
	11	<i>Catherine: Yeah.</i>	Catherine confirms Bella's written solution for the problem and they decide to move on to the next problem in the problem set.
	12	<i>Bella: Okay.</i>	
	13	<i>Catherine: Okay, we can move onto the next one.</i>	

FIGURE 3. An example of lower-quality reasoning. An excerpt of Group A's discourse or conversation while solving problem 2 in the transcription problem set. The excerpt consists of two lower-quality reasoning units, outlined in purple boxes. Lower-quality reasoning units are nontransactive, incomplete, and mixed, meaning that students are not operating on one another's reasoning, they do not provide backing to support linking their premises to a claim, and incorrect ideas are present. Each line corresponds to a speaker turn. Line-by-line analysis of the discourse dynamics is provided.

often do not start with the data in their discussion of this problem. Rather, students start by sharing their conclusions and what they know about the problem before drilling down to the

data that support their conclusions. In both reasoning units within this example of discourse, Group B takes this conclusion-first approach to answering problem 4a.

Reasoning Unit	Line	Discourse (conversation)	Analysis
Level 7 – Transactive, complete, correct	1	Oscar: <i>This only makes sense if it is bound to Ran-GTP.</i>	Oscar shares his interpretation that the figure only makes sense (a claim) if a certain requirement is met (a premise).
	2	Molly: <i>That's what I was thinking too. I was like, what? 'Cause it should ... since it's not bound, FRET should work just fine, and you should only get emission from the yellow fluorescent protein, which means that the ratio that we're observing here should be greater than one.</i>	Molly agrees with Oscar, expresses her initial confusion with the figure, and expands on Oscar's thinking by providing a more explicit and correct claim and premise.
	3	Adam: <i>Wait, why would you not get both? Don't both of these fluoresce if it's not bound?</i>	Adam then questions Molly's shared reasoning, and asks for an explanation while sharing an alternative interpretation (counterclaim) that is incorrect.
	4	Molly: <i>Well, if it's not bound, then they're so close together that if you essentially shine the wavelength of light necessary for this one, it's going to transfer that energy to the yellow fluorescent protein. So, you'll only see emission of ... the emission spectrum for yellow.</i>	Molly meets Adam's request by providing a correct explanation that supports her reasoning that she shared in line 2.
	5	Oscar: <i>Or more ...</i>	Oscar extends the claim Molly makes at the end of the previous line (line 4).
	6	Molly: <i>Yeah.</i>	Molly confirms Oscar's clarification.
	7	Oscar: <i>So, you would expect high on this if it's in this form.</i>	Oscar summarizes Molly's reasoning (claim and premise).
	8	Molly: <i>Technically, green is halfway between 0.77 and 1.5, which would be greater than 1, but I don't know, I feel like that's kind of like not a great defense.</i>	Molly clarifies what Oscar means when he says "high on this" in the previous line by providing evidence from the figure by saying where the color green falls on the scale and makes an evaluation of her own thinking.
	9	Oscar: <i>Yeah.</i>	Oscar and Adam agree.
	10	Adam: <i>Yeah.</i>	
Level 7 – Transactive, complete, correct	11	Molly: <i>Then, the entire nucleus, the intensity of CFP is much higher than the intensity of YFP.</i>	Molly extends her thinking from the previous reasoning unit to reason about the nucleus portion of the figure. She provides a correct claim.
	12	Oscar: <i>Oh. That makes sense. Because in the nucleus, we have Ran-GTP, so in the nucleus, it's in this form.</i>	Oscar agrees with Molly's claim and clarifies her reasoning by adding a correct premise ("in the nucleus, we have Ran-GTP") and connecting her claim to the schematic provided.
	13	Molly: <i>Okay.</i>	Molly and Adam indicate they are following.
	14	Adam: <i>Yeah.</i>	
	15	Oscar: <i>Which means you get more CFP alone, which brings down the-</i>	Oscar continues his line of thinking by restating Molly's claim from line 11.
	16	Adam: <i>So it should be-</i>	Adam finishes Oscar's thought by connecting Oscar's idea that more CFP would bring down the ratio to data in the figure by indicating the color they are seeing should be very dark.
	17	Molly: <i>Okay.</i>	
	18	Adam: <i>... very dark.</i>	

FIGURE 4. An example of higher-quality reasoning. An excerpt of Group B's discourse or conversation while solving problem 4a in the nuclear import problem set. The excerpt consists of two higher-quality reasoning units, outlined in purple boxes. Higher-quality reasoning units are transactive, complete, and correct, meaning that students are operating on one another's reasoning, they provide backing to support linking their premises to a claim, and no incorrect ideas are present, or if they are, they are ultimately corrected. Each line corresponds to a speaker turn. Line-by-line analysis of the discourse dynamics is provided.

Metacognitive Utterances Associated with Higher-Quality Reasoning

To address our second research question (Which metacognitive utterances are associated with small groups sharing higher-quality reasoning in an upper-division biology classroom?) we investigated the overlap and interplay between our two coding schemes. We were particularly interested in the metacognitive utterances situated within higher-quality reasoning units (level 7: transactive, complete, correct reasoning). Four categories emerged from this analysis. The metacognitive utterances that were associated with higher-quality reasoning units either

included 1) evaluative questioning, 2) requesting and receiving evaluations, 3) requesting and receiving explanations, or 4) elaborating on another's ideas (Table 3). Some higher-quality reasoning units contained more than one of these metacognitive utterance categories (Supplemental Figure 2). In the following subsections, we highlight illustrative examples of each metacognitive utterance category in transactive, complete, and correct (level 7) reasoning units.

Evaluative Questioning. In this study, a key distinguishing feature of higher-quality reasoning was evaluative questioning. We

TABLE 3. Categories of metacognitive utterances associated with higher-quality reasoning

Individual metacognitive regulation skill	Category	Description	Types of metacognitive utterances involved	Example
Evaluating	Evaluative questioning	When students would question whether or not the solution answered the problem or when students would challenge one peer's reasoning with alternative reasoning	Question to monitor understanding <i>or</i> Requests for information <i>doubled coded as an</i> Evaluation of others	Figure 5 Supplemental Figure 3
	Requesting and receiving evaluations	When a closed-ended question was met with an evaluation or a correction	Question to monitor understanding <i>followed by an</i> Evaluation of others <i>or a</i> Correction of another student	Figure 6 Supplemental Figure 2
Monitoring	Requesting and receiving explanations	When an open-ended question was met with an explanation beyond a single-word answer	Requests for information <i>followed by an explanation</i>	Figure 7 Supplemental Figure 2
	Elaborating on another's reasoning	When students would explain or build on a group member's reasoning beyond a single-word answer without prompting	Statements to monitor understanding	Figure 8

define evaluative questioning as an appraisal of an approach or thinking formed as a question. For example, Michelle's question, "Does it answer the question why?," is considered evaluative questioning, because it is an appraisal of the group's solution to a problem and is posed as a question to her group members. In contrast, Bella's question, "Can you explain that?," seeks an explanation from her group mate Catherine but is not an evaluation of Catherine's answer or approach. In essence, evaluative questioning occurred in our data set when metacognitive utterances that were questions included within them an evaluation, or when a question to monitor understanding or a request for information overlapped with an evaluation of others (Table 3). Evaluative questioning was found only in level 6 and level 7 reasoning units.

Evaluative questioning took two forms. First, evaluative questioning occurred when students would question whether or not the constructed reasoning answered the question asked in the problem set. An example of this type of evaluative questioning can be seen in Supplemental Figure 3, when Oscar asked, "But does that have to do with the gradient?," in line 3. Another example of this type of evaluative question came from Michelle when she asked, "Does it answer the question why?" (Table 1). Her group was trying to answer a problem that asked why Ran-GDP is concentrated in the cytoplasm. Before Michelle's question, Bella and Catherine came up with a solution consisting of the correct rule that Ran-GTP has a high affinity for importin and exportin. This solution, while composed of correct backing, did not completely address the problem asked or link the correct backing to a claim. With her evaluative question, Michelle raises the concern that their solution might not fully answer the problem. Her question got Catherine to reflect and admit that their solution did not answer *why* Ran-GDP is concentrated in the cytoplasm. This led the group to discuss the role of regulatory proteins (e.g., GAPs and GEFs). Ultimately, the group established consensus around the correct idea that Ran-GDP is concentrated in the cytoplasm because RanGAP,

which activates the GTPase activity of Ran-GTP, is found in the cytoplasm. Michelle's evaluative question redirected and refocused her group to the problem that was posed.

The second form of evaluative questioning occurred when students would challenge a part of one peer's reasoning with alternative reasoning. An example of this type of evaluative questioning can be seen in Figure 5. In this example, Oscar provided his reasoning about how the Ran-GTP binding domain of importin beta will behave as a control in a FRET experiment that his group was discussing, which led Adam to ask an evaluative question in line 8 (Figure 5). Adam's evaluative question focused on Oscar's premise that half of Ran-GTP will bind to importin beta by suggesting an alternative. Adam asked, "Wouldn't you expect almost all of it to bind, because it looks like the same as this one?" In essence, Adam's evaluative question challenges Oscar's reasoning with alternative reasoning (Figure 5). An additional example of this type of evaluative question came from Adam when he asked, "Wait, why would you not get both? Don't both of these fluoresce if it's not bound?" (Figure 4, line 3). Before this question, Molly shared an explanation. Adam then asked for clarification and challenged her explanation by offering an alternative interpretation of what might be happening in the experimental figure. Adam's evaluative question pushed Molly to explain her reasoning in greater detail for the group.

Requesting and Receiving Evaluations. Another distinguishing feature of higher-quality reasoning in this study was requesting and receiving evaluations. Requesting and receiving evaluations occurred when a closed-ended question to monitor understanding was followed by an evaluation of others or a correction of another student (Table 3). An example of requesting and receiving an evaluation is seen when Adam asked for confirmation on an idea and was corrected by Oscar in lines 7 and 8 of Figure 6. Adam shared his reasoning posed as a question: "So, wouldn't it be like the GDP

Reasoning Unit	Line	Discourse (conversation)	Analysis
Level 7 – Transactive, complete, correct	1	Oscar: <i>It's going to act as a ... It's going to be competing with RBD of the YRC probe.</i>	Oscar shares his premise that the Ran-GTP binding domain of importin beta will act as a competitor to the Ran-binding domain (RBD) of the probe in the experiment.
	2	Molly: <i>Oh, so we're adding ... Okay, yeah.</i>	Molly indicates she is following Oscar's reasoning.
	3	Oscar: <i>So, there's just extra competition for binding Ran-GTP.</i>	Oscar restates his premise that he shared in line 1.
	4	Adam: <i>Okay. So, you have the [inaudible] bind the GTP-</i>	Adam starts to share his reasoning
	5	Oscar: <i>So, half of Ran-GTP will bind to the-</i>	But Oscar interrupts Adam to continue sharing a more detailed version of his premise.
	6	Molly: <i>Oh, okay.</i>	Molly indicates she is following Oscar's reasoning.
	7	Oscar: <i>... Importin beta, and half will bind to this, which is going to effectively reduce this.</i>	Oscar finishes extending his reasoning by sharing a premise ("half of Ran-GTP will bind to the importin beta and half will bind to this [the YRC probe]") and a claim ("which is going to effectively reduce this").
	8	Adam: Wouldn't you expect almost all of it to bind, because it looks like the same as this one?	Adam acts on Oscar's reasoning by asking an evaluative question . His question critiques Oscar's premise by suggesting a correct alternative (that all the Ran-GTP will bind to the competitor) backed by evidence ("it looks the same as this one").
	9	Oscar: <i>Yeah.</i>	Oscar appears to agree with Adam's critique.
	10	Adam: <i>I don't know if Importin binds faster or something.</i>	Adam qualifies his alternative premise by stating he does not know the nature of the competition.
	11	Oscar: <i>We just know that it's in competition-</i>	Oscar clarifies that they only know it is a competitor and Adam agrees.
	12	Adam: <i>Yeah.</i>	
	13	Oscar: <i>... with the RBD that they used.</i>	

FIGURE 5. Evaluative questioning in student discourse that challenges reasoning. A higher-quality (level 7) reasoning unit from Group B, outlined in purple. Adam's evaluative question is in bolded font. Parts of the discourse and analysis shown in gray text are provided as context for the reader to emphasize what occurs after the evaluative question (black text).

grabs it in here and then sends it through here and then it's converted to GTP in here?" By doing so he sought confirmation on his idea from his group members. His question to monitor his own understanding opened the floor for feedback from his group members and gave Oscar the opportunity to correct Adam's thinking. This resulted in Adam accepting Oscar's correction.

In higher-quality reasoning units, not every question to monitor understanding was met with an evaluation or a correction. In other words, asking a question to monitor understanding did not always guarantee a response such as that shown in Figure 6. Some questions to monitor understanding that sought an evaluation or confirmation were not addressed verbally by the group. We acknowledge the possibility of students receiving simple nonverbal answers (like the shake or nod of a head) from group members that cannot be detected via audio recordings. Alternatively, this may suggest that, when students asked for feedback in this way, they were sometimes ignored. We speculate that some questions to monitor understanding went unmet because the group might have thought the individual was talking to themselves. Additionally, questions to monitor understanding were also found in lower-quality reasoning units. However, in lower-quality reasoning units, questions to monitor understanding were met with simple one-word confirmations or with another question rather than more elaborate evaluations or corrections. How the group responds to requests for evaluation seems to be important.

Requesting and Receiving Explanations. Another distinguishing feature of higher-quality reasoning in this study was requesting and receiving explanations. Similar to requesting and receiving an evaluation, a request is made by one student and then met by one or more members of the group, but the nature of the request is slightly different. Requests that sought confirmation resulted in evaluations, whereas requests that were open-ended questions resulted in explanations. The nature of the request dictated the response from the group.

Requesting and receiving explanations occurred when an open-ended request for information was met with an explanation composed of more than a single-word answer (Table 3). For example, consider the reasoning unit in Figure 7 that begins with a request for information from Bella, "Can you explain that? I'm very confused" (Figure 7, line 1). Bella directly asked for an explanation from her group members and received one. Interestingly, every request for information was met with a response in our study. No request for information went unmet. This suggests that, when students in our study asked for help this way, they were never ignored. While requesting and receiving explanations were more common in higher-quality reasoning units, requesting and receiving explanations were occasionally found in lower-quality reasoning units. However, in lower-quality reasoning units, requests for information were met with explanations that often included incorrect or mixed ideas that were never corrected. We underscore that how these requests are met appears to be important.

Reasoning Unit	Line	Discourse (conversation)	Analysis
Level 7 – Transactive, complete, correct	1	Adam: Yeah. Due to the low concentration of Ran-GDP in the nucleus ...	Adam, the recorder, is saying out loud what he is writing down for the problem.
	2	Molly: And I would say high concentration of Ran-GTP.	Molly suggests an addition to Adam's written solution.
	3	Adam: Yeah. High ... the nucleus ... Ran-GDP ...	Adam agrees and continues saying out loud what he is writing, including the incorrect premise that Ran-GDP concentration is high in the nucleus.
	4	Molly: Well, if we're talking in the nucleus, it would be like Ran-GTP-	Molly corrects Adam's incorrect idea with the correct premise
	5	Oscar: Ran-GTP binds to Importin and releases the cargo.	Oscar finishes the correction that Molly began by adding a claim.
	6	Molly: Yeah, and induces ... or like, the exchange, the hydrolysis ... No, the exchange of nucleotides promotes like a conformational change, that releases the cargo protein. So, that only happens though because Ran-GTP has a high affinity for Importin, and Importin is bound to the cargo protein.	Molly agrees with Oscar's claim and builds on his reasoning by adding backing in the form of rules ("Ran-GTP has a high affinity for importin").
	7	Adam: Okay. So, wouldn't it be like the GDP grabs it in here and then sends it through here and then it's converted to GTP in here?	Adam asks a question to monitor his understanding that seeks confirmation on his shared idea.
	8	Oscar: No, that's more how export ... That is only a two-cargo system. This is three. So, don't get this confused with this.	Adam's request for confirmation is met with an evaluation from Oscar when he corrects Adam's thinking and suggests where he might be getting confused.
	9	Adam: Okay. I think that's what I was doing. Okay.	Adam agrees with Oscar's appraisal of his thinking.
	10	Molly: Yeah, also, it's flipped. The bottom is the cytoplasm and the top is the nucleus, so ...	Molly agrees and notes that the figure labeling may have contributed to Adam's confusion.

FIGURE 6. Requesting and receiving evaluations in student discourse. A higher-quality (level 7) reasoning unit from Group B, outlined in purple. Adam's request is in bolded font in line 7 and Oscar's evaluation is in bold in line 8. Parts of the discourse and analysis shown in gray text are provided as context for the reader.

Elaborating on Another's Reasoning. Elaborating on another student's reasoning was another category for metacognitive utterances associated with higher-quality reasoning units. Elaborating on another's reasoning occurred when students would self-explain a group member's reasoning or elaborate on a group member's reasoning beyond a simple "yeah" or "okay" (Table 3). These self-explanations and elaborations were unprompted statements to monitor understanding. To illustrate this category of elaborating on another's reasoning, Figure 8 shows an excerpt of discourse between Michelle and Bella as they formed a conclusion about a data figure on the occupancy of histone acetylation for a region of a yeast chromosome. In the first part of this excerpt (lines 1–13), Michelle and Bella co-constructed their prior knowledge about the role of acetylation and methylation of nucleosomes in regard to transcription. Bella then restated the question in the problem set ("Okay, so what can you conclude?"), and Michelle responded with her reasoning by providing evidence about a relationship between one gene in the figure and the level of acetylation ("At [gene 1], we've got a lot of acetylation") and a claim ("so that means that that's a gene that's being regularly transcribed"). Bella then elaborated on Michelle's reasoning by providing data ("Oh, yeah, because at [gene 1] is when it shoots up") and clarifying the evidence Michelle stated by bringing in the idea that the increase in acetylation is occurring at the start of the gene ("So, at the advent of [gene 1], we see dramatic increase in acetylation"). Michelle then provided unprompted confirmation of

Bella's idea by defining the start of the gene as the promoter region based on what she remembers from class ("Yeah, because she said the little arrow thing means promoter, so we've got a promoter right there. So, right there, it's getting going").

As seen in Figure 8, some unprompted elaboration found in the higher-quality reasoning units was composed of "yes, and/because/so" statements (lines 16 and 17). However, "yes, and/because/so" statements were not always indicative of higher-quality reasoning. These elaborative statements to monitor understanding, while more common in higher-quality reasoning units, were also found in lower-quality reasoning units. In some instances of lower-quality reasoning, these elaborative cues masked disagreement. Take for instance, Group A's use of elaborative cues in Figure 3 (lines 6 and 9). Catherine's use of "yeah, and" statements may have been perceived as agreement by her group members even when, based on the content of the conversation, it was clear to the research team that the group was not in complete agreement about their reasoning (Figure 3). The appearance of agreement and elaboration could be misleading for students working in small groups in real time.

In summary, metacognitive utterances that 1) stimulated reflection about the solution or presented alternative reasoning, 2) provided evaluations when requested, 3) provided explanations when requested, or 4) elaborated on another's reasoning in an unprompted manner emerged as critical aspects of higher-quality reasoning in the data analyzed for this study.

Reasoning Unit	Line	Discourse (conversation)	Analysis
Level 7 – Transactive, complete, correct	1	Bella: Can you explain that? I'm very confused.	Bella makes a request for information by asking her group members to explain their reasoning.
	2	Catherine: Yeah. Okay. So, look at this. FRET, when FRET is working, you're going to have only the emission of this one yellow fluorescent protein, because that's what's happening here. Even though this is the one excited, it transfers that so that only green light is emitted. But if they don't associate, then you get this color.	Bella's request for information is met with an explanation from Catherine.
	3	Bella: What happens to the green color?	Bella asks a follow-up question.
	4	Catherine: I don't know if that means it-	Catherine starts to answer.
	5	Michelle: So, for which part are you ...	Michelle interrupts and starts to ask a question
	6	Catherine: Oh, if you're doing one certain color, then that means you'll only get this ... Oh, sorry, one certain excitation, you only get this kind of emission, because that's not specific to this green, but when they associate ...	Catherine tries to answer Bella's question from line 3.
	7	Michelle: Yeah, like when they're together, if you hit it with what excites the blue, because they're close to each other, I don't know, the vibration will go across, whew, or whatever, and then the green light will be emitted, but if the green were just by itself and you hit it with that purple, nothing would happen. It's because it's hugging the blue one, it's able to get excited when the blue is excited.	Michelle builds on Catherine's explanation.
	8	Catherine: Okay, so 1.5, this yellowish color means that there's a lot of YFP being emitted, meaning that the R binding domain isn't ...	Catherine uses their explanation to link evidence (ratio) and data (color) to a claim ("there's a lot of YFP being emitted")

FIGURE 7. Requesting and receiving explanations in student discourse. A higher-quality (level 7) reasoning unit from Group A, outlined in purple. Bella's request is in bolded font in line 1, and Catherine's explanation is in bold in line 2. Parts of the discourse and analysis shown in gray text are provided as context for the reader.

DISCUSSION

Analysis of discourse during small-group problem solving in an upper-division biology course revealed seven types of metacognitive utterances (Table 1) and four categories of metacognitive utterances that were associated with higher-quality reasoning in our study (Table 3). These findings have not been described in this context before and begin to define social metacognition that occurs in the life sciences. We situate our findings from this unique context among broader findings on social metacognition, outline implications for life science instructors based on our data, and suggest future directions for research on social metacognition in the life sciences.

Social Metacognition and Reasoning in the Life Sciences

Our findings build on prior research on social metacognition from mathematics, physics, and the learning sciences and contribute the first exploration of social metacognition in the life sciences. Prior work conceptualized metacognitive utterances during small-group problem solving in secondary mathematics courses broadly as "new ideas," or occurrences when new information was recognized or alternative approaches were shared, and "assessments," or occurrences when the execution, appropriateness, or accuracy of a strategy, solution, or knowledge was appraised (Goos *et al.*, 2002). Other researchers have used the individual metacognitive regulation skills of planning, monitoring, and evaluating to categorize metacognitive utterances from small group work during content analysis (Kim and Lim, 2018). Our rich descriptions of seven types of metacognitive

utterances in an upper-division biology course moves the field forward by 1) exploring social metacognition in a new context and 2) conceptualizing metacognitive utterances in social settings beyond new ideas and assessments (Goos *et al.*, 2002; Van De Bogart *et al.*, 2017) and the three broad individual metacognitive regulation skills of planning, monitoring, and evaluating (Lippmann Kung and Linder, 2007; Siegel, 2012; De Backer *et al.*, 2015; Kim and Lim, 2018). We propose alignment of the metacognitive utterance types and categories we found in our data set to the individual metacognitive regulation skills of planning, monitoring, and evaluating in order to begin to bridge individual metacognitive theory to the social metacognition framework (Tables 1 and 3). "How are individual and social metacognition related?" remains an open question in the field. More research is needed to determine additional criteria that make metacognition social and not individual.

Seven types of metacognitive utterances emerged in our data (Table 1). For the metacognitive utterances that were questions, the nature of the question determined the type of response it received. For example, open-ended questions like requests for information elicited more elaborate explanations from group members. In contrast, a closed request for feedback, like a question to monitor understanding ("Is my idea right?"), often elicited single-word responses ("Yeah."). Although this is the first study of social metacognition in the life sciences, this finding aligns with prior research on group work. In a study on student discourse in a life science course, open questions using the words "how" and "why" led to more conceptual explanations

Reasoning Unit	Line	Discourse (conversation)	Analysis
Level 7 – Transactive, complete, correct	1	Michelle: Okay. So, acetylation, does that-	Michelle asks for confirmation on her premise that acetylation encourages transcription.
	2	Bella: That makes it more-	
	3	Michelle: encourages transcription?	
	4	Bella: Yeah.	Bella confirms Michelle's premise
	5	Michelle: It like opens it up?	Michelle asks for confirmation again.
	6	Bella: Yeah.	Bella reconfirms Michelle's premise.
	7	Michelle: Okay.	Michelle accepts Bella's confirmation.
	8	Bella: We can write, like, "Acetylation encourages ..."	Bella, who is the recorder for the group, suggests what to write.
	9	Michelle: "...transcription, like opens up DNA." I'm just like, if [crosstalk].	Michelle completes Bella's statement about what to write.
	10	Bella: Or it makes the DNA more accessible.	Bella rewords the premise.
	11	Michelle: Yeah, if nothing else, I'm just writing it as a reminder to myself to [inaudible]. Yeah, it makes DNA more accessible. That's good.	Michelle agrees and evaluates Bella's suggested rewording of the premise.
	12	Bella: And methylation was the opposite.	Bella extends the premise to methylation.
	13	Michelle: Mm-hmm (affirmative).	Michelle agrees.
	14	Bella: Okay, so what can you conclude?	Bella restates the question asked in the problem set.
	15	Michelle: At [gene 1], we've got a lot of acetylation, so that means that that's a gene that's regularly being transcribed.	Michelle offers her reasoning by providing evidence and a claim.
	16	Bella: Wait I thought at, Oh, yeah, because at [gene 1] is when it shoots up. So, at the advent of [gene 1], we see dramatic increase in acetylation.	Bella elaborates on Michelle's reasoning by providing data and clarifying the evidence Michelle shared by making a statement to monitor understanding.
	17	Michelle: Yeah, because she said the little arrow thing means promoter, so we've got a promoter right there. So, right there, it's getting going.	Michelle provides unprompted confirmation to Bella's elaboration by adding a further premise and evidence to support their co-constructed reasoning.

FIGURE 8. Elaborating on another's reasoning in student discourse. A higher-quality (level 7) reasoning unit from Group A, outlined in purple. Bella's and Michelle's elaborations are in bolded font in lines 16 and 17. Parts of the discourse and analysis shown in gray text are provided as context for the reader.

from peers (Repice et al., 2016). In another study, peer learning assistants' use of open-ended prompting questions and reasoning requests during clicker discussions encouraged students to share their thinking and elicited student reasoning (Knight et al., 2015).

The four categories of metacognitive utterances that were associated with higher-quality reasoning in our study (Table 3) extend the ways that a student's thinking can become the subject of discussion (Goos et al., 2002). Our categories of evaluative questioning and requesting and receiving an explanation build on what Goos et al. (2002) called a "partner's challenge" or when one student (Student A) asks another student (Student B) to operate on the second student's (Student B's) thinking in order to clarify their meaning. Evaluative questioning extends this idea of a partner's challenge to also include when students question whether or not their co-constructed reasoning or solution answered the question asked in the problem set. Both types of evaluative questioning that we identified in our study and requesting an explanation elicited further reasoning from group members. Our category of requesting and receiving evaluations also builds on what Goos et al. (2002) called an "invitation" or when one student (Student A) asks another student (Student B) to operate on the first student's (Student A's) thinking in order to receive feedback. Requesting and receiving evaluations extends this idea of invitation and suggests that invitations are particularly powerful when met. Our category of elaborating on

another's thinking is aligned to what Goos et al. (2002) called "spontaneously, partner-initiated other-monitoring," in which one student (Student A) operates on another student's (Student B's) thinking in an unprompted manner in order to provide unrequested feedback. Unique to our investigation of higher-quality reasoning, we found that spontaneous other-monitoring often appeared as statements of agreement or elaboration, that is, "yes, and/because/so," rather than corrections. In fact, we did not see evidence of unprompted corrections of another student in our analysis of higher-quality reasoning. Every correction in our data set was invited or requested by a group member (Figure 6).

In research on reasoning and argumentation, the moments when students disagree appears to be critically important (Kuhn, 1991). In fact, some reasoning and argumentation frameworks rank discussions with counterclaims, disagreements, and rebuttals as more sophisticated (Osborne et al., 2004). Disagreements were present in some but not all of our higher-quality reasoning units. A few disagreements appeared overtly as direct corrections of another student ("No, ..."), but more often disagreements and counterclaims were present in our data set in the form of subtle evaluations of the group's solution through evaluative questioning (see Adam's and Oscar's questions in Figures 4 and 5 and Supplemental Figure 3). The phrasing of critiques and counterclaims as questions might be a way to be polite, soften the blow, or save face in a

TABLE 4. Prompts that may promote social metacognition during small-group problem solving

To encourage students to...	Have students consider...
respectfully challenge their group member's thinking	asking their group, "Does this answer the question asked?" or "What alternative ideas do we have?"
invite their group members to evaluate their thinking	asking, "What about my idea does not make sense?" after they share an idea.
ask for explanations	asking their group, "Can you explain that to me?"
elaborate on one another's ideas	using "I agree/disagree because..." or "That makes sense/doesn't make sense because..." statements after their group members share an idea.

group setting, because outright disagreement with one's peers can be seen as socially undesirable. Group members might feel more open to discussion or comfortable with the possibility of having an incorrect idea when an alternative idea is presented as a question. On the other hand, phrasing critiques and counterclaims tentatively might cause some contributions to be overlooked if not stated directly or assertively (see the exchange in Figure 3), especially depending on the group dynamic (i.e., if there is a more dominating group member present).

Implications for Instructors

We found that metacognitive utterances that 1) stimulated reflection about the answer or presented alternative reasoning, 2) provided evaluations when requested, 3) provided explanations when requested, or 4) elaborated on another's reasoning in an unprompted manner were associated with higher-quality reasoning. Students are likely to need structured guidance on how to be socially metacognitive (Chiu and Kuo, 2009; Stanton *et al.*, 2021) and how to reason (Knight *et al.*, 2013; Paine and Knight, 2020). Our rich, qualitative work begins to provide the foundational knowledge needed to develop this guidance.

Scripts or prompts for social metacognition could be provided to students working in small groups (Miller and Hadwin, 2015; Kim and Lim, 2018). Other researchers, particularly in the realm of computer supported collaborative learning, have identified and used scripting tools to structure and sequence collaborative online interactions like small-group problem solving (Miller and Hadwin, 2015; Kim and Lim, 2018). Scripting for social metacognition in a life science classroom could involve providing prompts for students to use during group work and then modeling when and why to use these prompts. Based on our data, we suggest several possible prompts in Table 4. Incorporating these prompts during small group work could encourage students to practice aspects of social metacognition that were associated with higher-quality reasoning in our study. The effectiveness of these prompts in promoting social metacognition and reasoning is unknown and will be tested by our lab in a future study.

Another way to possibly facilitate social metacognition during small-group problem solving is through the use and modification of group roles. Other researchers suggest that students' natural role choices are not necessarily optimal and that consideration of group roles can improve group discussions (Paine and Knight, 2020). In our study, students took on a defined group role as either recorder, presenter, or manager which are common group roles for a POGIL-style classroom (Moog *et al.*, 2006). To facilitate social metacognition, these group roles could be expanded and tested in the following ways. First, we do not suggest expanding the recorder role,

because it was already the most demanding group role in our study. Second, the presenter role could be expanded to include the role of prompter. The role of the prompter would be to encourage the group's use of scripting prompts throughout the group work (Table 4). The prompter could also be tasked with ensuring any questions asked by a member of the group are answered, because we found that requests and invitations were particularly powerful when met. Finally, the manager role could be expanded to include the role of moderator. The role of the moderator would be to encourage active listening.

A particularly interesting form of active listening to consider for the expanded role of moderator is apophatic listening (Dobson, 2014; Samuelsson and Ness, 2019). Apophatic listening is a process in which learners start by being quiet to give space for a speaker to share their ideas. Learners use this silence to temporarily suspend their expectations and reflect on the speaker's ideas before actively participating in the conversation. After actively listening to the speaker, learners then ask follow-up questions and interpret what the speaker shared. The listener then shares their alternative understanding with the speaker, and the listener and speaker work together to collectively create a shared and mutual understanding (Samuelsson and Ness, 2019). Adding the role of moderator to encourage apophatic listening could ensure all group members are heard. In our study, apophatic listening may have been more present in Group B's discourse because of the greater proportion of silence and turn-taking during group work compared with Group A's overlapping talk (Figure 2).

Structured guidance on how to reason is also needed in college life science classrooms (Paine and Knight, 2020). Another study of upper-division biology students showed that instructional cueing for reasoning led to higher-quality reasoning during clicker question discussions (Knight *et al.*, 2013). In alignment with these previous findings from a different population of upper-division biology students (Knight *et al.*, 2013), the majority of the reasoning units analyzed in our work were higher in quality (Supplemental Table 2). A combination of factors, including the prior educational experiences of the studied population, the nature of the task itself (Zagallo *et al.*, 2016), and the course expectations around sharing reasoning, may have been sufficient to elicit higher-quality reasoning in our study. Notably, students in our study were not explicitly taught to reason using the evidence-based reasoning framework (Brown *et al.*, 2010), and the most common approach to reasoning in our data set was a conclusion-first or claim-first approach (Figure 4). It might be helpful to teach students to reason during small-group problem solving with a data-first approach (Supplemental Figure 1) so their claims flow logically from backing (data, evidence, rules).

Future Directions for Research

Our initial investigation of social metacognition during small-group problem solving in a life science context reveals a hypothesis for future testing as well as interesting areas for further research. Based on our results, we hypothesize that certain categories of metacognitive utterances shared by students lead to higher-quality reasoning from their groups. In other words, reasoning is improved when students employ specific categories of metacognitive utterances during group work. If this is true, then we predict that training students to be socially metacognitive through the use of specific prompts will lead to higher-quality reasoning and group performance. For example, we predict that incorporating evaluative questioning and requesting and receiving explanations will transition students to higher-quality, more transactive, complete, and correct reasoning. We are testing this hypothesis using scripting interventions that structure and sequence group collaboration as others have done in the past (Miller and Hadwin, 2015). Alternatively, the inverse may be true, and higher-quality reasoning may lead to certain types of metacognitive utterances being shared. Our foundational work uncovered a possible association between metacognitive utterances and higher-quality reasoning, but the strength and directionality of the relationship needs to be further investigated. Additionally, our findings should be validated in other contexts within the life sciences. For example, the social metacognitive utterances we identified in our work may vary from those students use in a organismal biology class or in courses that use group work in a synchronous online format (Zheng *et al.*, 2019).

Other areas for future research of social metacognition and reasoning quality in the life sciences include the role of silence, cultural variations in collaboration, and group diversity. We found that Group B was more successful in their problem solving and spent more of their on-task time in silence compared with Group A (Figure 2). This silence was often, but not always, in between problems in the problem set, suggesting it could be representative of group members silently reading, writing, or thinking. Capturing video data in future work can help us understand what is occurring during those silent pauses. The idea of silent time during group work may be counterintuitive to the notion that effective collaboration involves constant conversation. In fact, our findings suggest it may be beneficial for all group members to have silent time to process and reflect on their own thoughts and the thoughts of others. This finding could have a direct impact on tools intended to measure active learning using sound (Owens *et al.*, 2017). However, it is important to note that this result may simply capture group differences in conversational style or cultural variations in the way collaboration is viewed (Alcalá *et al.*, 2018). For instance, the talking over one another that resulted in minimal silences in Group A's dialogue may show the group's enthusiasm rather than interruption.

Cultural aspects of collaboration and the impact of group diversity on social metacognition and reasoning quality should be considered in future work (Carter and Phillips, 2017; Alcalá *et al.*, 2018). For example, how does a more surface-level homogenous group (e.g., a group of all women) compare to a more surface-level diverse group (e.g., a group of men and women) in terms of social metacognition and reasoning quality? In other research, dissenting group members who were in

the social majority found surface-level diverse groups to be more accepting than surface-level homogenous groups (Phillips and Loyd, 2006). The two groups in our study were mostly surface-level homogeneous, which may have impacted how students perceived and shared potential disagreements. While there was high surface-level homogeneity among group members in our study, we are unable to comment on the deep, cognitive-level diversity (i.e., differences in knowledge, skills, and experiences) in the groups aside from the frequencies of metacognitive utterances by participant (Table 2). Our work does not reveal the impact that personality, prior knowledge, group role, and preference for group work may have on individual frequency of metacognitive utterance use. Individual and group differences will be important to consider when investigating social metacognition and reasoning quality in future studies.

Limitations

Our study was designed to qualitatively explore social metacognition and reasoning during small-group problem solving in an upper-division biology course through discourse analysis. Our sample size, while small, is in line with sample sizes traditional for foundational discourse analysis (Cameron, 2001; Rogers, 2004). This type of analysis is time and labor intensive. For example, on average, each transcript we analyzed contained approximately 700 coded segments. We do not aim to make generalizations from our study. Rather, our goal is to present a thick description of the metacognitive utterances and reasoning from two groups over the course of two consecutive breakout sessions in an upper-division biology course. Deep and rich analysis of this type is warranted to expand our understanding of emerging areas of research in the life sciences, like social metacognition.

Social metacognition is context dependent and discipline specific. Our analysis and results from an upper-division biology course may not reflect the nature of social metacognition in other contexts, like an introductory biology course that uses team-based learning. Not all utterances that students made were metacognitive in nature. For example, the question "Alright, so what should I write?" and statements like "ChIP is genome-wide" were not coded as being metacognitive in nature. These examples have been taken out of context, and we caution readers that the context around these utterances, as in all discourse analysis, was important and critical in our coding decisions. For this reason, it is important to present excerpts of discourse in the results, when possible, rather than single quotes.

We recognize that language is just one system that people use to create meaning. We relied on audio recordings in our analysis of discourse during small-group problem solving. In future studies, we can use video recordings to investigate how participants use gestures, objects, and technology. This would allow us to explore other nonverbal systems that interact with language during small-group problem solving. Additionally, the evidence-based reasoning framework we used breaks reasoning into its component parts and focuses on structure without focusing on the holistic nature of the reasoning or whether it is scientifically accurate (Brown *et al.*, 2010; Furtak *et al.*, 2010). This is a limitation of the reasoning framework we used, and one that we attempted to address by accounting for correctness in our reasoning coding scheme.

CONCLUSION

Social metacognition is important for undergraduates to fully benefit from increasing opportunities for small-group work in life science courses. This work begins to define metacognitive utterances shared by undergraduate students during small-group problem solving in life science courses and suggests a relationship between metacognitive utterances and higher-quality reasoning, which is a valued outcome in the life sciences. Our data suggest it may be important to provide life science undergraduates with guidance on how to be socially metacognitive through the use of scripted prompts and modified group roles during small-group work.

ACKNOWLEDGMENTS

We are grateful to members of the Biology Education Research Group at UGA for their feedback on our work. We thank Dr. Paula Lemons and Dr. Mariel Pfeifer for their feedback on an earlier version of this article and the reviewers for their critical insights. We also acknowledge Julio Cordero for his help with preliminary data analysis. This material is based on work supported by the National Science Foundation under grant no. 1942318. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

- Alcalá, L., Rogoff, B., & López Fraire, A. (2018). Sophisticated collaboration is common among Mexican-heritage US children. *Proceedings of the National Academy of Sciences USA*, 115(45), 11377–11384. <https://doi.org/10.1073/pnas.1805707115>
- American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC.
- Artz, A. F., & Armour-Thomas, E. (1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. *Cognition and Instruction*, 9(2), 137–175. https://doi.org/10.1207/s1532690xci0902_3
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Beichner, R. J., Saul, J. M., Abbott, D. S., Morse, J. J., Deardorff, D., Allain, R. J., ... & Risley, J. S. (2007). The Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP) project. *Research-based Reform of University Physics*, 1(1), 2–39.
- Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In Glaser, R. (Ed.), *Advances in instructional psychology* (Vol. 1, pp. 77–165). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Brown, N. J. S., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The evidence-based reasoning framework: Assessing scientific reasoning. *Educational Assessment*, 15(3–4), 123–141. <https://doi.org/10.1080/10627197.2010.530551>
- Cameron, D. (2001). *Working with spoken discourse*. Thousand Oaks, CA: Sage.
- Carter, A. B., & Phillips, K. W. (2017). The double-edged sword of diversity: Toward a dual pathway model. *Social and Personality Psychology Compass*, 11(5), e12313. <https://doi.org/https://doi.org/10.1111/spc3.12313>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Chiu, M. M., & Kuo, S. W. (2009). From metacognition to social metacognition: Similarities, differences, and learning. *Journal of Education Research*, 3(4), 321–338.
- De Backer, L., Van Keer, H., & Valcke, M. (2015). Exploring evolutions in reciprocal peer tutoring groups' socially shared metacognitive regulation and identifying its metacognitive correlates. *Learning and Instruction*, 38, 63–78. <https://doi.org/https://doi.org/10.1016/j.learninstruc.2015.04.001>
- De Backer, L., Van Keer, H., & Valcke, M. (2020). Variations in socially shared metacognitive regulation and their relation with university students' performance. *Metacognition and Learning*, 15(2), 233–259. <https://doi.org/10.1007/s11409-020-09229-5>
- Dobson, A. (2014). *Listening for democracy: Recognition, representation, reconciliation*. Oxford, UK: Oxford University Press.
- Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., & Paolino, D. (1992). *Santa Barbara papers in linguistics*, Vol. 4, *Discourse transcription*. Santa Barbara, CA: University of Santa Barbara.
- Edwards, J. A. (2005). The transcription of discourse. In *The handbook of discourse analysis* (pp. 321–348). Malden, MA: Blackwell Publishers. <https://doi.org/https://doi.org/10.1002/9780470753460.ch18>
- Furtak, E. M., Hardy, I., Beinbrech, C., Shavelson, R. J., & Shemwell, J. T. (2010). A framework for analyzing evidence-based reasoning in science classroom discourse. *Educational Assessment*, 15(3–4), 175–196. <https://doi.org/10.1080/10627197.2010.530553>
- Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics*, 49(2), 193–223. <https://doi.org/10.1023/A:1016209010120>
- Jost, J. T., Kruglanski, A. W., & Nelson, T. O. (1998). Social metacognition: An expansionist review. *Personality and Social Psychology Review*, 2(2), 137–154. https://doi.org/10.1207/s15327957pspr0202_6
- Kim, D., & Lim, C. (2018). Promoting socially shared metacognitive regulation in collaborative project-based learning: A framework for the design of structured guidance. *Teaching in Higher Education*, 23(2), 194–211. <https://doi.org/10.1080/13562517.2017.1379484>
- Knight, J. K., Wise, S. B., Rentsch, J., & Furtak, E. M. (2015). Cues matter: Learning assistants influence introductory biology student interactions during clicker-question discussions. *CBE—Life Sciences Education*, 14(4), ar41. <https://doi.org/10.1187/cbe.15-04-0093>
- 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>
- Kruger, A. C. (1993). Peer collaboration: Conflict, cooperation, or both? *Social Development*, 2(3), 165–182. <https://doi.org/10.1111/j.1467-9507.1993.tb00012.x>
- Kuhn, D. (1991). *The skills of argument*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511571350>
- Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, 44(1), 46–53. <https://doi.org/10.3102/0013189x15569530>
- Lippmann Kung, R., & Linder, C. (2007). Metacognitive activity in the physics student laboratory: Is increased metacognition necessarily better? *Metacognition and Learning*, 2(1), 41–56. <https://doi.org/10.1007/s11409-007-9006-9>
- Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior*, 52, 573–588. <https://doi.org/https://doi.org/10.1016/j.chb.2015.01.050>
- Moog, R. S., Spencer, J. N., & Straumanis, A. R. (2006). Process-oriented guided inquiry learning: POGIL and the POGIL project. *Metropolitan Universities*, 17(4), 41–52.
- National Research Council (NRC). (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press. <https://doi.org/doi:10.17226/11625>
- NRC. (2015). *Enhancing the effectiveness of team science*. Washington, DC: National Academies Press. <https://doi.org/10.17226/19007>
- Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328(5977), 463–466. <https://doi.org/10.1126/science.1183944>
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020. <https://doi.org/https://doi.org/10.1002/tea.20035>
- Owens, M. T., Seidel, S. B., Wong, M., Bejines, T. E., Lietz, S., Perez, J. R., ... & Tanner, K. D. (2017). Classroom sound can be used to classify teaching practices in college science courses. *Proceedings of the National Academy of Sciences USA*, 114(12), 3085. <https://doi.org/10.1073/pnas.1618693114>

- Paine, A. R., & Knight, J. K. (2020). Student behaviors and interactions influence group discussions in an introductory biology lab setting. *CBE—Life Sciences Education*, 19(4), ar58. <https://doi.org/10.1187/cbe.20-03-0054>
- Phillips, K. W., & Loyd, D. L. (2006). When surface and deep-level diversity collide: The effects on dissenting group members. *Organizational Behavior and Human Decision Processes*, 99(2), 143–160. <https://doi.org/10.1016/j.obhdp.2005.12.001>
- Repice, M. D., Keith Sawyer, R., Hogrebe, M. C., Brown, P. L., Luesse, S. B., Gealy, D. J., & Frey, R. F. (2016). Talking through the problems: A study of discourse in peer-led small groups. *Chemistry Education Research and Practice*, 17(3), 555–568. <https://doi.org/10.1039/C5RP00154D>
- Rogers, R. (2004). *An introduction to critical discourse analysis in education*. New York, NY: Routledge.
- Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, 92(3), 499–525. <https://doi.org/https://doi.org/10.1002/sce.20264>
- Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). Thousand Oaks, CA: Sage.
- Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, 92(3), 447–472. <https://doi.org/https://doi.org/10.1002/sce.20276>
- Samuelsson, M., & Ness, I. J. (2019). Apophatic listening. *Democracy and Education*, 27(1), 6.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351–371. <https://doi.org/10.1007/BF02212307>
- Siegel, M. A. (2012). Filling in the distance between us: Group metacognition during problem solving in a secondary education course. *Journal of Science Education and Technology*, 21(3), 325–341. <https://doi.org/10.1007/s10956-011-9326-z>
- Smith, J. M., & Mancy, R. (2018). Exploring the relationship between metacognitive and collaborative talk during group mathematical problem-solving—what do we mean by collaborative metacognition? *Research in Mathematics Education*, 20(1), 14–36. <https://doi.org/10.1080/14794802.2017.1410215>
- Stanton, J. D., & Dye, K. M. (2017). Investigating the function of a transport protein: Where is ABCB6 located in human cells? *CourseSource*. <https://doi.org/https://doi.org/10.24918/cs.2017.19>
- Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering metacognition to support student learning and performance. *CBE—Life Sciences Education*, 20(2), fe3. <https://doi.org/10.1187/cbe.20-12-0289>
- Tanner, K. D. (2012). Promoting student metacognition. *CBE—Life Sciences Education*, 11(2), 113–120. <https://doi.org/10.1187/cbe.12-03-0033>
- Toulmin, S. E. (2003). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- Van De Bogart, K. L., Dounas-Frazer, D. R., Lewandowski, H., & Stetzer, M. R. (2015). The role of metacognition in troubleshooting: An example from electronics. Paper presented at: Physics education research conference (College Park, MD).
- Van De Bogart, K. L., Dounas-Frazer, D. R., Lewandowski, H. J., & Stetzer, M. R. (2017). Investigating the role of socially mediated metacognition during collaborative troubleshooting of electric circuits. *Physical Review Physics Education Research*, 13(2), 020116. <https://doi.org/10.1103/PhysRevPhysEducRes.13.020116>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research*, 84(1), 30–43. <https://doi.org/10.1080/00220671.1990.10885988>
- Wilson, K. J., Brickman, P., & Brame, C. J. (2018). Group work. *CBE—Life Sciences Education*, 17(1), fe1. <https://doi.org/10.1187/cbe.17-12-0258>
- Zagallo, P., Meddleton, S., & Bolger, M. S. (2016). Teaching real data interpretation with models (TRIM): Analysis of student dialogue in a large-enrollment cell and developmental biology course. *CBE—Life Sciences Education*, 15(2), ar17. <https://doi.org/10.1187/cbe.15-11-0239>
- Zheng, J., Xing, W., & Zhu, G. (2019). Examining sequential patterns of self- and socially shared regulation of STEM learning in a CSCL environment. *Computers & Education*, 136, 34–48. <https://doi.org/10.1016/j.compedu.2019.03.005>