# Accuracy and Idea Consideration: A Study of Small-Group Interaction in Biology

**Andy R. Cavagnetto,[†]* Joshua Premo,[‡] Zachary Coleman,[§] and Kate Juergens[¶]**
[†]Department of Teaching & Learning and School of Biological Sciences, Washington State University, Pullman, WA 99164-2132; [‡]Department of Biology, Utah Valley University, Orem, UT 84058; [§]College of Pharmacy, Washington State University, Pullman, WA 99164; [¶]Department of Teaching & Learning, Washington State University, Pullman, WA 99164

## ABSTRACT
**Small-group discussion is a central component of 21st-century biology classrooms. Many factors shape these discussions and thus influence potential learning gains. This study examined how accuracy and idea consideration shaped small-group discussions in undergraduate biology labs (12 groups, M = 42.8 talk turns). To do this, we asked 1) Is there a relationship between a student's science accuracy and the amount peers consider the student's ideas? 2) To what extent does peer consideration of a student's ideas predict that student's ability to steer the conversation? Building on this second question, we then explored 3) Does general group academic ability or immediate conversational accuracy better predict group learning? To answer these questions, we coded aspects of discourse (science accuracy, idea consideration, etc.) before quantitative analysis. Strong correlation was found between students' science accuracy and idea consideration ($r$ = 0.70). Both accuracy and idea building predicted one's ability to steer the conversation. Subsequent analysis highlighted the critical role of immediate discourse in group learning. Group-level analysis revealed that group performance was not related to the group's overall ability in the classroom, but rather the immediate accuracy of their group conversations. Implications and limitations are discussed.**

## INTRODUCTION
The educational research community generally accepts that student-to-student dialogue supports increased content learning and reasoning skills. Several studies have supported this belief by demonstrating the power of small-group collaborative work (Johnson and Johnson, 2009; Pai *et al.*, 2015; Larrain *et al.*, 2018; Ruiz-Gallardo and Reavey, 2019; Hand *et al.*, 2021). Yet not all studies of student-to-student dialogue uniformly show positive outcomes (Kuhn, 2015; Sfard, 2015), with significant variation in benefit having been noted (g = –0.95 to 1.83; Pai *et al.*, 2015). These findings suggest that more nuanced factors in small groups should be examined to determine what drives productivity in small-group dialogue.

Many potential processes have been identified that can support and inhibit small-group learning. Nokes-Malach *et al.* (2015) report on cognitive and social factors of influence. For example, tasks that can be accomplished individually do not require investment in others for success, so in these settings, the cognitive cost of coordinating between group members outweighs the benefit of the group. Also, simply having to wait one's turn to speak or attending to others' ideas can cause one to miss opportunities to retrieve information or ideas that are critical to making sense of a concept. Social factors such as uncommitted group members and fear that others will judge one's ideas can also serve as inhibitors of productive discourse. Conversely, hearing others' ideas can bring new insights that are unavailable to an individual, cue one's inert knowledge, and correct one's inaccurate ideas. Fundamentally, many of these processes described by Nokes-Malach *et al.* (2015) rest on the extent of energy invested

**Rebecca Price,** *Monitoring Editor*
Submitted Mar 19, 2021; Revised Oct 27, 2021; Accepted Nov 5, 2021
CBE Life Sci Educ March 1, 2022 21:ar5
DOI:10.1187/cbe.21-03-0067
*Address correspondence to: Andy Cavagnetto (andy.cavagnetto@wsu.edu). ORCID: 0000-0001-8057-1660.

in collaboration. As such, one's perception of benefit relative to cost of collaboration is important to consider when designing and carrying out group activities. While benefit to cost is ultimately determined by the individual, we argue that many of the factors identified as influential to small-group collaboration indirectly or directly influence benefit relative to cost. In the following paragraphs, we highlight factors from previous studies at the individual and task level that influence perceptions of benefit and thus potential for committing to group work.

Slavin (1996) highlights the role that personal goals and perceptions of others play in one's motivation to collaborate. In his review of research on cooperative learning, Slavin argues that if cooperation increases the potential to meet personal goals, students' motivation for cooperation will be high. Our own work has shown that the potential benefit of cooperating is central to one's willingness to invest in peers during science learning (Premo *et al.*, 2018a). Thus, perceptions of peers matter. Perceived expectations, even when based on poor information or bias (e.g., gender or racial stereotypes) have been found to be influential in groups (Grunspan *et al.*, 2016). Friendship among group members and the degree to which one cares for group members can increase the extent and intent of cooperation (Slavin, 1996; Berndt, 2002; Majolo *et al.*, 2006). Similarly, as shown in studies related to game theory, potential for reciprocity has also been found to influence the amount of investment a person is willing to make in group members (Nowak and Highfield, 2011). Personality traits also add to the benefit to cost ratio. Characteristics such as agreeableness, conscientiousness, and extraversion can make group work feel easier and are often valued in groups (Curşeu *et al.*, 2019). Similarly, related actions by group members such as speculative (as opposed to definitive) talk and creation of space for others to offer ideas have been found to support group productivity (Boyd *et al.*, 2019). Such contexts can contribute to psychological safety (Theobald *et al.*, 2017; Van den Bossche *et al.*, 2006) and subsequent risk taking with regard to sharing and challenging ideas in groups. Academic ability can also be viewed through a benefit to cost lens. The ideas available within a conversation can influence the amount of effort needed for cooperation. Therefore, being knowledgeable about the topic or having a knowledgeable group member can decrease cost and increase potential benefit for a student when choosing the extent to engage in science discussions.

Characteristics of the learning task can influence cooperation. In general, it is important to seed interdependency among group members (i.e., reliance on others to accomplish one's goals). This can be accomplished with a jigsaw structure (Aronson, 1978), by assigning discrete roles for students (Johnson and Johnson, 2009), introducing reward structures (e.g., friendly competition between groups; Webb, 1982; Slavin, 1996; Sober and Wilson, 1999), and ensuring that the task is too complex for any individual to complete alone (Saleh and Large, 2011; Sears and Reagin, 2013). Other factors related to the task, such as clearly defined and shared goals (Sniezek and Henry, 1989; Scribner *et al.*, 2007; Van den Bossche *et al.*, 2006), can increase the chances of group members moving in the same direction and reduces the potential to talk past one another. Ultimately, these individual and task characteristics affect motivation to cooperate, which importantly, can have downstream effects on the quality of the group interactions (Asterhan, 2018). For us, this is the fundamental pathway by which task structures and individual characteristics influence learning in small groups.

### Epistemic Vigilance and Considering Others' Ideas

We have outlined factors that have been shown to influence the quality of cooperation. We see these factors as indirectly (in the case of task-related factors) or directly (in the case of personal factors) influencing group members' benefit to cost ratios. Factors influencing potential benefit are important, because thinking deeply about ideas is hard work; therefore, rich discussions related to target concepts require a considerable investment in both time and energy. This is particularly true in instances when group members have limited understanding of target concepts or when students are asked to synthesize across multiple data sources, as can be the case in science classes.

However, even when students' benefit to cost ratios are favorable and they are committed to collaboration, there is potential for misinformation and miscommunication to reduce the effectiveness of collaboration. Misinformation may be due to the competence level of an individual, yet it is important to recognize that there is nearly always some level of misalignment (in background knowledge, values, etc.) among members that can decrease the clarity of communication between them. Mercier and Sperber (2011) argue that we naturally defend against misinformation and miscommunication through the practice of epistemic vigilance. Epistemic vigilance includes the consideration of who is providing the idea (trust calibration) and how well the idea aligns with prior experiences and knowledge (coherence checking). If an idea or information aligns with the receiver's prior knowledge, acceptance offers little risk, so the idea is likely to be considered. However, the greater the incoherence with prior knowledge and experience, the greater the potential risk that these ideas are incorrect or false. For Mercier and Sperber (2011), if there is incoherence with prior knowledge, then either trust in the sender must be sufficiently high or the incoherence of the message with prior knowledge needs to be reduced before an idea will be fully considered. Reduction in incoherence is what the sender attempts to do via well-reasoned and logical arguments.

In educational settings where students are asked to learn dialogically with peers, the consideration of ideas is paramount. This is because consideration of ideas is a precursor for conceptual change. One's ideas cannot change without the initial consideration of the new idea in relation to one's current schema. Thus, idea consideration is where learning occurs. Cognitive processes such as cross-cueing, error correction, and increased availability of ideas and cognitive resources (Nokes-Malach *et al.*, 2015) all occur as ideas are considered. Idea consideration not only influences what and how knowledge is constructed by group members, but it also can feed back into the group by influencing the foci or direction of the group conversation. This means that understanding student consideration of ideas, as communicated via the epistemic vigilance framework, may provide insights into optimizing group discourse during learning. The current study focuses on idea consideration to explore dynamics of group talk. First, the positive relationship between a student's science accuracy during conversation and displays of idea consideration was tested. Second, both accuracy of a student and the idea consideration directed toward the student were used to predict one's ability to influence the topic of group conversation. It is hypothesized that higher levels of both science accuracy and indicators of idea consideration may privilege certain students in steering group discourse. Finally, the impacts of accuracy were examined at the whole-group level. Of central focus was whether learning is better predicted by the academic ability of a group (a longer-term group characteristic) or by the immediate scientific accuracy during group conversations. The answers to these questions provide insights into both factors that shape how group conversations occur as well as why some groups learn more than others in biology classrooms.

### Context

Recently, we have worked to develop a strategy to support increased collaboration and subsequent learning of course target concepts that could be implemented by graduate teaching assistants (TAs) with little to no educational training. The context of our work has been a modified jigsaw activity within a large-enrollment course-based undergraduate research experience (CURE) in introductory cell biology and genetics. The strategy requires students to work together using a jigsaw structure (Aronson, 1978) to respond to three questions targeting key aspects of the CURE near the conclusion of each lab session. The questions are examined during three phases of the task. In phase 1, each group of three to four students receives one question to discuss and reach consensus. In phase 2, one person from each group forms a new group in which each member worked on a different question in phase 1. Each student leads the discussion of the question that student worked on in phase 1. Group members are tasked with reaching consensus on the three questions. In phase 3, each phase 2 group writes their solutions to each of the three questions on the whiteboard. The TA then moderates a whole-class discussion to examine the similarities and differences across groups to reach a scientifically acceptable solution to each question (for a more detailed treatment of this design, see Premo *et al.*, 2018b).

These phases were designed to increase the benefit of cooperation. The goals of the task were clearly defined, and the jigsaw structure (Aronson, 1978) was designed to create interdependence and consensus seeking among group members—particularly in phase 2 of the activity. Additionally, we created questions that were complex enough to have high likelihood of requiring collaboration for success. Complexity not only creates a need for collaboration, but also increases the likelihood of contrasting ideas being raised in small-group discussion—an established characteristic of academically productive talk (Howe, 2014). Further, the jigsaw activity was in the context of a CURE. There is some evidence that CUREs can increase students' ownership of their learning and shared goals (Corwin *et al.*, 2015), which we expected might also support greater benefit to cost for students. We expected that students would put more energy into advancing their own projects—including the cooperative aspects of their projects.

Despite trying to capture many influential task factors, our initial study (Premo *et al.*, 2018b) found no clear and compelling performance gains on 1-week delayed quizzes when comparing the intervention group with class sections that did not collaborate. In other words, we were unable to detect a positive group effect. We suspect that this could be a function of personal factors that were not strongly influenced by the task structures accounted for in the instructional strategy. For example, while CUREs allow for greater student ownership and decision making, increased ownership has been documented in class sections with small student enrollment (relative to our prior study), allowing for more individual attention and ability for personalization. Participants in our prior study and current work were students enrolled in a large-enrollment course (more than 500 students every semester) with higher levels of standardization than other CURE programs due to this large-enrollment context. Due to this larger-scale structure with less personalization, students may continue to see the class as simply a requirement to be completed for their majors, thereby limiting their feelings of true ownership. Further, the lab sections of the large-enrollment introductory course are taught by TAs with limited pedagogical training. As with many novice teachers, some of the lab TAs are heavily concerned with classroom management, which can lead to a reduction in the autonomy and therefore ownership of groups and individual students. So, while many of the structures of the course and task hold potential to support positive group effects, the structures do not guarantee gains.

In the initial study, we were unable to directly control for personal factors while maintaining an appropriate level of ecological validity for the given context (e.g., effective tracking of student motivation during the task, perceived reciprocity, or goal orientation during the task). While it is challenging to measure personal factors in real time, analyzing student-to-student talk could highlight the extent to which students considered one another's ideas (i.e., coherence checking) and the nature of group dynamics, which are influenced by personal factors. Thus, analyzing idea consideration and group dynamics could provide insights into how learning is negotiated among group members in collaborative activities and, in doing so, inform considerations for successful group collaboration in the future. The current study sought to better understand how individuals engage with one another by using video and audio recordings of student groups from an introductory undergraduate cell biology laboratory course. In particular, we were interested in understanding the extent to which student idea consideration aligned with accuracy (as predicted by epistemic vigilance) and the potential role of these constructs in determining differential impact that some students might have in shaping group discussions. These goals shaped the following research questions:

RQ1. Is there a relationship between a student's science accuracy and the amount peers consider the student's ideas?

RQ2. Does peer consideration of a student's ideas predict the student's ability to influence the direction of the conversation?

As a follow-up to these questions, we broadened our focus to the group level. If results of RQ1 and RQ2 aligned with our hypothesis that students dynamically adjust the extent of idea consideration based on accuracy, this would result in differential ability of some students to shape how conversations unfolded. Yet does this difference in ability to steer the conversation really matter in terms of student performance? A common perception is that group gains are often a function of a knowledgeable individual simply dispensing knowledge to group mates. Yet our own previous work (Cavagnetto *et al.*, 2020; Premo *et al.*, 2018b, 2021) as well as other studies (Barron, 2003; Kuhn, 2015; Howe and Zachariou, 2019) point to a more complex mechanism of action. These studies suggest that group members do not simply learn more because they have greater academic ability, but rather members build their understanding through negotiation (via cross-cueing, error correction, re-exposure, exposure to diverse ways of thinking about the problem, and/or increased engagement; for a comprehensive summary, see Nokes-Malach *et al.*, 2015). Our third research question was designed to provide some insight into which of these two broad mechanisms (ability or conversational dynamics) was at work in this context. Thus, the third research question was:

RQ3. Does general group academic ability or immediate conversational accuracy better predict group learning?

## METHODS

### Participants

Eighteen students enrolled in a university-level introductory biology course consented to have their group work audio- and video-recorded at two time points as part of the study. Of these 18, 14 participants identified as female and 14 were Caucasian. All but two participants were science majors, but student progress in their programs of study varied. Seven participants were sophomores and seven were juniors. The remainder included one freshman, one senior, and two postbaccalaureate students. When taken together, most participants were Caucasian females in the middle years of their degrees. The sample was convenient—that is, reliant on volunteers who attended the lab sessions and completed all questions sets and subsequent quizzes. As such, the sample was not fully representative of the class. While demographics across semesters vary, our sample had slightly inflated representation of female (77%) and Caucasian (77%) students (~65% female and 65% Caucasian per semester). Postbaccalaureates also were elevated in the sample.

### Setting and Data Collection

The current study was part of a larger project that integrated question sets designed to support student understanding of the theoretical underpinnings of a semester-long course-based research experience, the SEAPhages program (Staub *et al.*, 2016). Students have historically completed the technical aspects of the lab without taking much time to reflect on the theory behind their actions in the lab. Therefore, reflective question sets were integrated into the lab with the goal of getting students to think more deeply about concepts and practices that both the course instructor and lab TAs had identified as needing further support. The integrated question sets began the fifth week of the semester. Students completed a total of 9 weeks of question sets addressing the following themes: evaluating a scientific paper (week 5), phage–bacteria interactions (week 6), experimental design (week 7), experimental troubleshooting (week 8), phages as biological tools (week 9), experimental replication (week 10), serial dilutions (week 11), plagiarism in scientific writing (week 12), and DNA isolation from phages (week 13). Video and audio recordings of students working together in groups of three to four were collected during 2 of the 9 weeks and are the focus of this study. Recordings were collected week 6 (Phage and Bacterial Interactions) and week 9 of the semester (Phages as Biological Tools). It is important to note that group members did vary between these weeks based on where students chose to sit for the activity.

Redundant recording devices were used to capture student talk due to the high volume and overlap of conversations. Each group was recorded by their own video camera and an audio recording device that was placed between the two middle seats of each lab table. Each audio recorder (one per group) had two directional recording microphones and an additional two directional microphones that were wired and placed facing inward from the outside of the outermost group members at each table to best capture group conversation. After collection, each audio file was merged with the corresponding video file to allow identification of individuals speaking . Then each file was cut down to only include conversation from the first to the last talk turn of science discussion related to the question sets to allow us to consistently track time for various characteristics of the talk. This was important, as the audio/video recorders needed to be started at different times based on the needs of the instructor. After cutting excess recording, the remaining files were professionally transcribed. Once files were transcribed, an author (J.P.) went through each transcript to do the following: 1) correct any terms incorrectly transcribed (many terms were science specific and not always picked up correctly by transcribers), 2) watch the video to remove any section of conversation picked up by a recording that came from another group, and 3) make sure that each talk turn in the transcript was an individual speaker, because at times during the transcription process, contributions from multiple students could be inadvertently fused into a single talk turn.

Each student completed a delayed postdiscussion quiz covering the material from the prior week's question set (1-week delay for week 6 and a 2-week delay for week 9 due to a university holiday). Quizzes were a minimal part of the overall course grade. Each quiz was worth 2 points and included three questions that needed to be answered (corresponding to at least two of the reflective prompts from the previous week's discussion). Answers were scored in 0.25-point or 0.5-point increments (depending on the question) by the lab instructor (see Table 1 for example materials). The quiz scores assessing both week 6 and week 9 question sets were used as performance outcomes in the current study.

### Coding and Analyses

Coding of student discourse focused on the talk turn level. A talk turn was defined as an uninterrupted instance of talk by an individual student in the group (similar to Barron, 2003). Video recorded groups were analyzed using an a priori designed coding scheme that sought to identify talk turns that indicate idea consideration toward a peer in a group, as well as accuracy of science ideas, use of evidence/justification, and on-task orientation of each student. Additionally, we examined the relative impact that each peer had in the conversation (via contributions that shifted conversation). Table 2 illustrates the codes used in this study and Table 3 provides an example of code application. A total of 514 talk turns were identified during all the group interactions ($M_{group} = 42.8$) and lengths of effective group work (defined as a group's first instance of science discussion to their last) ranged between 5 and 15 minutes.

**TABLE 1. Example of materials**

| Intervention question set | Quiz prompt | Scoring rubric |
|---|---|---|
| 1. One application of phages is to quickly detect the presence of food-borne pathogenic bacteria in food. As an example, a microbiologist working for Chobani takes two batches of yogurt off of the production line. She suspects that one yogurt sample is contaminated with *Listeria*, a common food-borne pathogen, while the other one may or may not be contaminated.<br>A. The scientist takes an extract from the contaminated yogurt sample and adds 1000 *Listeria*-specific virus particles. After shaking and incubation for a day, what will happen to the number of phage particles in the sample? Why? Is this experiment selective for *Listeria* or could other bacteria lead to false-positive results?<br>B. Based on your answer to Part A, how could the scientist tell if the second yogurt sample was contaminated with either *Listeria* or with *E. coli*, another potentially pathogenic bacterial species? | A food scientist is working for a dairy farm, and she suspects that a dangerous strain of *E. coli* may have contaminated a batch of milk. Describe an experiment that would allow the scientist to quickly and selectively screen for the presence of *E. coli*. | Set up a plaque assay (or spot test) with bacteria from the milk as the host and a phage known to infect *E. coli*. Plaques would indicate presence of the *E. coli* in the milk [0.5 points] |
| 2. Phage therapy is an alternative to antibiotic treatment for bacterial infections. In phage therapy, phages specific to pathogenic bacteria are delivered to the site of an infection (e.g., on wound dressings, by oral ingestion, through an IV).<br>A. One side effect of antibiotic treatment is that these chemicals kill many beneficial bacteria in the human gut. Why might phage therapy, for example, to combat a *Listeria* infection in the gut, be superior?<br>B. One problem with antibiotics is that they are typically unstable and quickly degrade in the body, resulting in the need for frequent, high doses during treatment. What advantage would phage therapy have over antibiotic treatments in this respect? | 2. Phage therapy has been shown to have advantages over traditional chemical antibiotics.<br>A. What is one reason that *Salmonella* phages may be preferred over chemical antibiotics to treat a *Salmonella* infection in the human gut?<br>B. How many doses of *Salmonella* phages would a doctor need to supply, in theory, if a patient came in with a *Salmonella* infection? Compare this to the standard, 5- to 7-day course of antibiotics that are normally used. | A. Phage treatment is specific for the species of bacteria causing the problem. [0.25 points]<br>Antibiotic treatment will kill other bacteria that are useful to the patient. [0.25 points]<br>B. In theory, a single phage treatment is sufficient (assuming the phage population maintains itself long enough to eliminate most of the infectious bacteria). A variety of answers allowed as provided by instructor to TAs. |
| 3. *Leuconostoc* is a type of bacteria that is used widely in food fermentation, including during the production of wine. Many reports have surfaced of phages negatively influencing wine making by killing *Leuconostoc* bacteria.<br>A. Discuss with your partners if you think either lytic, lysogenic, or both types of phages would have a major, negative impact on wine making.<br>B. How could wine makers reduce their chances of losing *Leuconostoc* bacteria during wine making? | N/A. Quiz did not target this prompt. | N/A |

After group conversations were transcribed and talk turns identified, both the third and fourth authors coded every talk turn independently for each category. There was strong agreement between both coders for *science accuracy* (Cohen's kappa [$k$] = 0.87), *idea consideration* ($k$ = 0.84), and *on/off task* ($k$ = 0.97) and moderate agreement for *conversational flow* ($k$ = 0.61). The individual toward whom indicators of idea consideration was directed was established by the two coders using both the transcripts and visual examination of video data. These were confirmed by the third coder (J.P.). Initially codes for *evidence of mistrust* and *evidence/justification provided* were included in the scheme, but both were dropped due to low frequency across student groups making them unusable without a more substantial sample size. All coding disagreements were resolved through discussion to generate the final data set for quantitative analysis.

The level of analysis differed across the research questions. Addressing both RQ1 and RQ2 required analysis at the individual level. Therefore, code frequencies were summed for all participants across both recorded sessions to answer RQ1 and RQ2. This served to provide an overall view of a student's science accuracy, idea consideration, and conversational flow. In contrast, RQ3 examined properties of student groups to provide insight into factors during group conversations that we hypothesized would be related to learning. In other words, RQ3 was interested in group-level idea accuracy versus academic ability in predicting student learning. The accuracy code was summed at the group level for the analysis and correlated to both group average academic ability and average quiz score. Relationships between codes were calculated using Pearson's correlations (RQ1), while multiple regression was used in instances where statistical control of additional variables was necessary (i.e., total

TABLE 2. Code descriptions

| Code | Description[a] |
|---|---|
| Science accuracy | Code designed to track the accuracy of statements made by group members. It includes talk turns that are scientifically *Inaccurate*, a mix of *Both* accurate and inaccurate information, and scientifically *Accurate*. |
| Idea consideration | Code designed to track the extent to which students' ideas are considered by their group members. It includes talk turns reflecting *Agreement* (also includes praise and affirmation), *Soliciting Ideas* (direct solicitation of an idea from another student, and *Idea Building* (expanding on a previous talk turn). |
| Direction of idea consideration | Code designed to track which students' ideas were considered. The student ID number of the idea is denoted. |
| Conversational flow | Code designed to identify points in the dialogue in which the direction of the conversation changes to a new topic. These points occurred when a *New Idea* was raised that shifted the conversation and when a student made *Reference to Materials* that resulted in a change in the topic of conversation. |
| On/off task | Code designed to track the amount of conversation that was dedicated to the task (*On*) and the amount of conversation that did not focus on the task (*Off*). |

[a]Italicized words in the description are those used in the coding (see Table 3 for an example).

number of talk turns or academic ability; RQ2). Finally, due to the presence of a smaller sample size at the group level ($n = 12$) nonparametric correction (Spearman's rho) was used for examining group-level correlations in RQ3.

## RESULTS

Results to each research question are provided here. We begin with a summary table (including research questions, variables, analyses, results, and conclusions; see Table 4) and then highlight the results of each research question in sequence.

### RQ1: Is There a Relationship between a Student's Science Accuracy and the Amount Peers Consider the Student's Ideas?

To determine whether there was a relationship between a student's science accuracy and the idea consideration the student receives from peers, we examined the relationship between the amount of scientifically accurate statements made by a student and the number of indications of idea consideration directed toward that student (sum of agreement, peers soliciting ideas from the student, and peers building off the student's ideas). Pearson's correlation results showed a strong correlation between these codes ($r = 0.70$, $p < 0.01$), indicating that the amount of idea consideration a student receives is highly associated with that student's science accuracy (Figure 1).

There are two alternative factors that may account for this relationship. First, the total amount of talk turns expressed by the student may be influential. Talking more during a group conversation may increase the likelihood that one receives idea consideration (i.e., you are more likely to have someone agree with you the more ideas you offer) and one's science accuracy (i.e., the more ideas you put forth, the more likely you will provide an accurate idea). To account for this potential, we used multiple regression to examine whether science accuracy predicts idea consideration when statistically controlling for the total amount of talk turns by the student. Regression results showed that both science accuracy, $b = 0.18$, $t(17) = 3.45$, $p < 0.01$, and total talk turns, $b = 0.37$, $t(17) = 2.52$, $p < 0.05$, simultaneously predicted significant variance in idea consideration received by a student ($R^2 = 0.68$). These results suggest that science accuracy is predictive of the idea consideration one received even when controlling for total amount of talk.

Second, students may have been using external cues about a peer's academic ability to judge the extent to which they should consider that peer's ideas. Participating students had been in the same classroom for several weeks, and grades were a

TABLE 3. Example of coding scheme application

| ID | Talk turn | Accuracy | Idea consideration | Direction of consideration | Conversation flow | On/off task |
|---|---|---|---|---|---|---|
| 15 | Um, centrifuge. Wait isn't centrifuge like a type of filter? | Accurate | | | New Idea | On |
| 16 | No it's the one where they spin it around to get all the solids distilled. | Accurate | Idea Building | 15 | | On |
| 15 | Oh, okay. Oh so it separates the… | Accurate | Idea Building | 16 | | On |
| 16 | Separates the liquid and the solid particles? | Accurate | Idea Building | 15 | | On |
| 6 | Yeah, so why do they do it? | | Agreement & Soliciting Ideas | 16 | New Idea | On |
| 16 | Uh, the, phage is less dense than water. | Accurate | | | | On |
| 6 | So the phage will be solid. | Both | Idea Building | 16 | | On |
| 16 | Or separate at least. They'll separate the solids, the bacteria, and the water into layers, the phage, centrifuges all of them. | Accurate | | | | On |
| 6 | And then they can filter it? | | Soliciting Ideas & Idea Building | 16 | New Idea | On |
| 16 | Then they can filter it. | | | | | On |

TABLE 4. Summary of research findings

| Research question | Variables | Analyses | Results | Conclusions |
|---|---|---|---|---|
| Is there a relationship between a student's science accuracy and the amount peers consider the student's ideas? | Science accuracy code (accurate), idea consideration codes (sum of agreement codes, soliciting ideas, and idea building), students' final semester percentage grades, total talk turns | Pearson correlation<br>Multiple regression | $r = 0.70$, $p < 0.01$<br>Academic Ability ($p > 0.05$) was removed from the model. Science accuracy, $b = 0.18$, $t(17) = 3.45$, $p < 0.01$, and total talk turns, $b = 0.37$, $t(17) = 2.52$, $p < 0.05$, predicted idea consideration ($R^2 = 0.68$). | A student's science accuracy is highly correlated with peer consideration of the student's ideas.<br>A student's science accuracy is predictive of the idea consideration the student receives, even when controlling for total amount of talk. |
| Does peer consideration of a student's ideas predict the student's ability to influence the direction of the conversation? | Science accuracy code (accurate), idea consideration codes (agreement, soliciting ideas, & idea building), and conversational flow codes (new idea & reference to materials) | Multiple regression | Science accuracy predicted conversation flow. The sum of idea consideration codes did not. Optimal model included science accuracy, $b = 0.27$, $t(17) = 2.46$, $p < 0.05$, $R^2 = 0.40$, and idea building, $b = 1.13$, $t(17) = 2.88$, $p < 0.05$, $R^2 = 0.20$. | Accuracy was the best predictor of one's ability to influence the direction of the conversation. Providing an idea that peers could be built upon predicted additional ability to direct the conversation. |
| Does general group academic ability or immediate conversational accuracy better predict group learning? | Group average academic ability (average semester grade), number of accurate statements per group (accurate code), average quiz score per group | Spearman's rho correlation | Average academic ability of a student group ($\rho = -0.25$, $p = 0.44$) did not predict quiz performance. Science accuracy ($\rho = 0.58$, $p < 0.05$) predicted quiz performance. | Accuracy within the conversation predicts performance, while academic ability does not determine success. |

common topic of student discussion both before and following their completion of the study's activity. To investigate this potential, student academic ability in the course (indicated by final semester percentage grade) was added into the regression model predicting idea consideration (before adding total talk turns and accuracy into the model). Academic ability did not significantly predict idea consideration displayed toward a student either by itself or in the more comprehensive model. Thus, academic ability was removed, allowing the most parsimonious model (reported earlier) to be adopted. Together, these results support that a student's science accuracy and number of talk turns together were able to predict most of the variance (68%) in the idea consideration that student received from peers.

### RQ2: Does Peer Consideration of a Student's Ideas Predict the Student's Ability to Influence the Direction of the Conversation?

Having established that there was a significant relationship between a student's accuracy and the idea consideration the student received, our next question was whether idea consideration and accuracy were able to predict a student's ability to influence the direction of group discussion. To address this question, we examined the relationships between science accuracy, idea consideration, and conversational flow codes per student. Results showed that all three factors were highly correlated. Multiple regression was then used to assess the extent to which science accuracy and idea consideration could predict one's ability to steer the conversation (conversational flow). When regressed separately, peer consideration of a student's ideas predicted only slightly more variance in a student's ability to impact conversational flow, $b = 0.46$, $t(17) = 3.63$, $p < 0.01$, $R^2 = 0.42$, than did science accuracy, $b = 0.42$, $t(17) = 3.60$, $p < 0.01$, $R^2 = 0.41$. Yet when simultaneously regressed on conversational flow, only science accuracy, $b = 0.15$, $t(17) = 2.27$, $p < 0.05$, but not idea consideration, $b = 0.08$, $t(17) = 0.39$, $p = 0.70$, significantly predicted variance in a student's ability to direct the conversation (conversational flow).

To further examine the potential of peer idea consideration to predict one's ability to influence the direction of the conversation, we then examined whether any of the subcategories of the idea consideration code (explicit agreement, question asking, or idea building) were uniquely predictive of conversational flow alongside science accuracy. When each of the subcategories were separately entered into the model, it was found that only idea building (the number of times a student had ideas built upon) predicted significant additional variance in a student's impact on conversational flow. Thus, the optimal model included both science accuracy, $b = 0.27$, $t(17) = 2.46$, $p < 0.05$, $R^2 = 0.40$, and idea building, $b = 1.13$, $t(17) = 2.88$, $p < 0.05$, $R^2 = 0.20$, which together predicted 60% of the variance in conversational flow among the students. These results indicate that, while science accuracy is highly related to idea

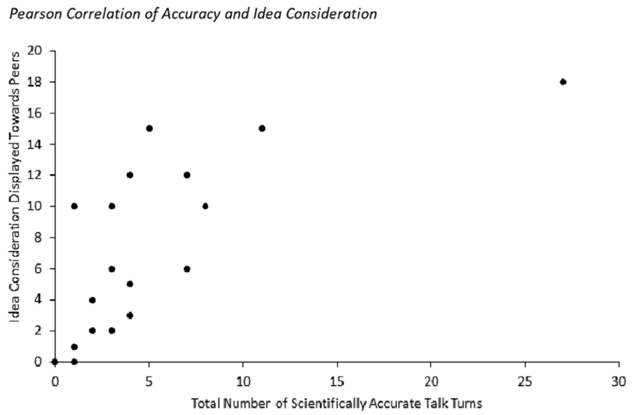


**FIGURE 1. The relationship between the number of scientifically accurate talk (accuracy) turns provided by a student and the idea consideration displayed toward that student ($r = 0.70$, $p < 0.01$). Note that the strength of this relationship was only slightly decreased ($r = 0.69$, $p < 01$) with the removal of the more extreme data point on the far right of the graph.**

consideration, one's accuracy appeared to be the best predictor of ability to influence the direction of the conversation. Yet providing an idea that peers could meaningfully build upon to further the conversation (i.e., idea building) predicted additional ability to direct the conversation. Importantly, the relationship between science accuracy and influence on conversation direction was not because the student dominated the conversation. As seen in Figure 2, all groups had multiple actively contributing members. Only one group had a single individual contribute more than half of the talk turns in the discussion.

### RQ3: Does General Group Academic Ability or Immediate Conversational Accuracy Better Predict Group Learning?

A common perception among instructors is that a single individual can dramatically enhance or reduce the effectiveness of all group members in small-group learning activities. We have heard from many science faculty members who have expressed the concern that if one group member has a misconception, the misconception may be propagated throughout the group. In other words, group members may be "infected" with the misconception. Results from RQ1 and RQ2 suggest that this concern may not be a common occurrence. Idea consideration is related to science accuracy, and both idea consideration and accuracy relate to an individual student's ability to impact group discussion. Yet does this ability to direct the conversation have positive implications for overall group learning? For example, if a student leader differentially impacts the direction of the conversation by supplying more accurate ideas, does this help the group as a whole? To answer this question, we focused on whether groups' overall academic ability in the course or conversation-specific accuracy (their total number of accurate statements during the small-group activity) correlated most strongly with group learning. Due to a small sample size at the group level ($n = 12$), data nonnormality was present, supporting the use of Spearman's rho to examine these relationships.

Results showed that there was no significant relationship between group performance on the quizzes and the average academic ability of a student group ($\rho = -0.25$, $p = 0.44$). In contrast, there was a significant relationship between group performance on the quizzes and science accuracy within their conversations ($\rho = 0.58$, $p < 0.05$). This difference aligns with past findings that group ability does not always determine group success (Barron, 2003) and suggests that how group discussion occurs may be the critical factor.

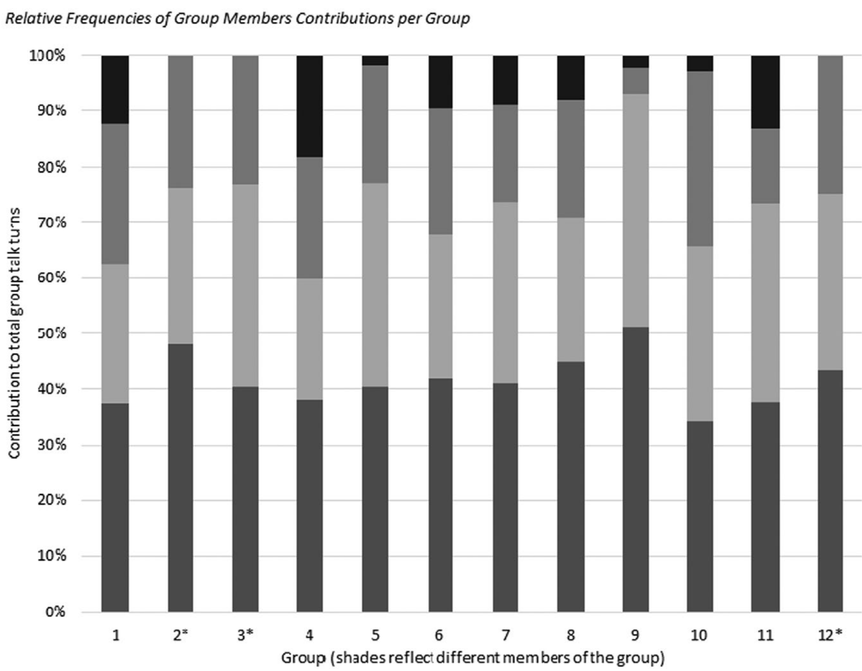


**FIGURE 2. The distribution of talk turns per group. Groups had four members, except for those groups marked with an asterisk (*), which had three members.**

## DISCUSSION

Classroom environments centered on peer–peer discourse have the potential to promote greater student learning in science, but this potential does not guarantee learning gains. This study sought to investigate the intersection between factors that may play a role in how conversations unfold in student groups and the relationships these factors have to assessment performance. We specifically investigated student small-group interactions to understand how peer consideration of students' ideas and students' science accuracy predicted their ability to influence the direction of conversation. Given results showing relationships between these factors, we then examined how both within-conversation accuracy and academic ability influenced group performance on delayed content-related quizzes.

Results of the study suggest that coherence checking was an active part of how students chose to engage in ideas within group discussions and that this process is related to the ability of students to influence the direction of the conversation. Specifically, results for RQ1 showed that the accuracy of statements provided by students was highly related to the extent to which students' ideas were considered by their peers. This finding contrasts the common concern that misconceptions held by a student in a group are likely to be propagated across group members. If students were basing their decisions upon student reputation or alternative factors outside the conversation, we would not see a relationship between accuracy and consideration of ideas. Given that the correlation is strong, we believe this suggests that consideration of ideas was primarily based upon students' checking the coherence of an idea with their own background knowledge and other information available in the task materials (i.e., coherence checking). This conclusion is further supported by the finding that science accuracy remained a significant indicator of idea consideration even when controlling for potential confounding variables (total talk and academic ability of the student).

Student accuracy was also found to predict a student's ability to influence the direction or topic of conversation. Students who provided more accurate statements had a greater influence on the topics discussed by their groups. Interestingly, this was not because they dominated the conversation. All groups had multiple actively contributing members. Only one group had a single individual contribute more than half of the talk turns in the discussion. The idea that one student is not dominating the conversation is further supported by the finding that the ability of peers to build off their ideas was an additional factor that accounted for students' ability to direct conversation. Thus, in part, a student's ability to present ideas that teammates could build upon supported a student's ability to direct conversations. From RQ2, we can also conclude that students do not appear to be blindly following the highest-achieving student(s) in the group, as student academic ability was not predictive of ability to direct group conversation.

In RQ3, we investigated whether group academic ability or within-discussion accuracy predicted learning gains as measured by delayed (1-week and 2-week postdiscussion) quizzes. Accuracy of within-discussion statements predicted group performance, while group academic ability did not. This finding supports previous studies that have found relationships between the number of accurate contributions and performance (Chiu, 2008a,b). One might conclude that, when group members are subject to more accurate statements, they are more likely to recall those ideas on the related quiz. We raise some skepticism concerning this simplistic explanation for two reasons. First, simple exposure to an idea does not readily lead to successful transfer on a quiz delayed 1 or 2 weeks (Haskell, 2001). Second, we would argue that the explanation does not capture the complexity involved in student sense-making. Instead, it seems to rely on a tabula rasa version of learning that has long been debunked (National Research Council, 1999). While we acknowledge that exposure may help with recall, students interpret new ideas considering their prior knowledge and experiences. As such, we see overreliance on exposure diverting attention from other mechanisms, such as cueing inert knowledge and raising alternatives that challenge misconceptions (Howe, 2014; Nokes-Malach *et al.*, 2015).

Taken together, the findings of the current study show evidence that what happens during group discussions is key to student learning gains. Students are evaluating the merit of ideas put forth by group members. Students are not commonly making decisions based on the perception of others, but rather on the coherence of the ideas with their prior knowledge and the information available within the task. We see this finding as important, because it offers solid evidence that the risk of misconceptions being propagated within a group is low.

These findings offer important implications for university science instructors. First, considering the strong relationship among frequency of accurate statements and overall group performance, group constituency is important. Demographic variables such as race and gender have been shown to influence group dynamics, including whose ideas are given priority (Grunspan *et al.*, 2016; Schnittka and Schnittka, 2016; Graham and Echols, 2018). This study extends the findings of group constituency to include optimizing scientifically accurate statements within groups. Yet how can accuracy be optimized within groups? If one views that group accuracy arises from the sum of accuracy of each individual group member when they enter the group, then methods that predispose groups to be accurate before the conversation (e.g., manipulating group composition or preparation) are most appropriate. But results from the current study indicate that group science accuracy was not simply driven by student accuracy coming into the conversation. Rather, groups were able to functionally filter through contributions of members toward those that were more accurate and thus more likely to further the group's ability to effectively complete the assigned task. This in-the-moment filtering of ideas (coherence checking) resulted in more-accurate individuals being able to differentially shift the conversational focus. The results for RQ2 suggest that the ability to shift the direction of conversation was at least in part a function of peers' abilities to build upon the accurate ideas. This finding broadens instructor consideration of group constitution to also include consideration of resources available for coherence checking.

The results also suggest that it is important to support the processes by which groups are considering the ideas put forth by their members. As previously indicated, consideration of ideas may be influenced by demographic variables. Yet the task-related resources available for idea consideration are also very important. Focusing on the resources available before or during the task to support coherence checking of ideas may be particularly effective due to the limited time faculty have to work with students in class. For example, an important question to consider is "Do students have sufficient resources available to check the coherence of the likely ideas put forth by group members?" Inherent in this question are other valuable questions to consider, such as "What exactly is the learning target for the activity?," "What are commonly held misconceptions?," "How do novices tend to think about the science topic?," and "What information should be supplied either before class or during the task that will allow students to evaluate various ideas?" Similarly, instructors might consider whether groups understand how to systematically solve problems. Brief cues that allow the group to focus on coherence checking may be needed. For example, "What information related to the task is known?," "What is unknown?," "What are possible solutions to the unknown?," "What known information is relevant to evaluating the possible solutions to the unknown?," and "How does the provided information support or refute our various ideas about the unknown?" can serve as prompts to organize the discussion toward coherence checking.

Examination of resources for coherence checking should also consider what information is not required. Providing students with too much information can unnecessarily influence cognitive load, thereby reducing the potential for effective coherence checking (Chandler and Sweller, 1991). We recognize that the ability to filter through denser conceptual material and multiple ideas is a skill we want to foster in students. Yet this is a more expert-oriented skill that may not be most appropriate for students early in their undergraduate careers. Instead, allowing students to work through ideas without this additional cognitive load may allow more cognitive resources to be dedicated to the coherence-checking process. Furthermore, students (and instructors) may benefit from making the coherence-checking process more visible. Encouraging students to think aloud or on paper about how different ideas align with their current understanding and the understanding of those around them may be pertinent to this effort, given past results supporting the importance of student metacognition (Donker *et al.*, 2014).

Finally, we believe the findings serve as a reminder that small-group work requires time. Time is always an issue in science instruction as instructors strive to cover large amounts of science content, but coherence checking requires sufficient time for students to leverage their background knowledge and task resources provided. Therefore, instructors should not overlook the amount of time required for any small-group activity. When instructional planning conflicts arise due to time constraints, we suggest limiting the learning targets to those manageable within less time over maintaining the same learning targets and expecting students to move through them at a faster pace. The latter path not only increases student cognitive load but also may interfere with the coherence-checking process, because students feel pressure to accept ideas and move to the next section of the task.

## LIMITATIONS AND IMPLICATIONS FOR FUTURE RESEARCH

Studies of student-to-student dialogue are challenging undertakings. Quantitative studies of dialogue are particularly challenging because of sample size issues. This study, most notably RQ3, suffers from limited sample size at the group level. This is a limitation that could be addressed in future studies by drawing on participants with greater diversity in background experiences and cultures. Second, as demonstrated in the literature reviewed earlier, there are myriad factors that influence dialogic interactions. We attempted to account for some of these factors (e.g., academic performance, total amount of student talk), but were unable to systematically account for others (e.g., conceptions of others). In this study, we were interested in coherence checking—the extent to which ideas were assessed for accuracy. We did not measure this directly, but rather did so by looking at indicators of idea consideration such as agreement, idea building, and direct solicitation of ideas. We acknowledge that codes such as agreement and direct solicitation of ideas could also be influenced by social factors such as friendship (Majolo *et al.*, 2006). Similarly, we do not know for certain whether students recognize the accuracy of the statements in the moment or if they are cued by other factors embedded within the talk utterance (e.g., confidence). Future studies might develop methods to rule out some of these extraneous factors or attempt to look at the interaction effects of multiple factors. For example, it would be appropriate to examine the role of accuracy and idea consideration among groups with targeted group constitution (e.g., various gender constituencies). Finally, in the current study, we did not consider the role of justification in peer idea consideration. In some of our recent work, we found that justification of ideas is a critical characteristic of discourse that differentiates higher- and lower-performing groups (Premo *et al.*, 2021). It may be that justification of one's ideas is more likely when they are accurate and that justification may be mediating the relationship between accuracy and idea consideration. That is, one is more likely to consider and successfully evaluate ideas that are justified, because the reasoning is more explicit. This should be addressed in future work seeking to examine additional mechanisms that may cue students to either consider or reject an idea presented by a peer.

## ACKNOWLEDGMENTS

This study was supported by an educational research grant from the Washington State University College of Veterinary Medicine.

## REFERENCES

Aronson, E. (1978). *The jigsaw classroom*. Beverly Hills, CA: Sage.

Asterhan, C. S. C. (2018). Exploring enablers and inhibitors of productive peer argumentation: The role of individual achievement goals and of gender. *Contemporary Educational Psychology*, *54*, 66–78. https://doi.org/10.1016/j.cedpsych.2018.05.002

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Berndt, T. J. (2002). Friendship quality and social development. *Current Directions in Psychological Science*, *11*(1), 7–10. https://doi.org/10.1111/1467-8721.00157

Boyd, M. P., Chiu, M. M., & Kong, Y. (2019). Signaling a language of possibility space: Management of a dialogic discourse modality through speculation and reasoning word usage. *Linguistics and Education*, *50*, 25–35. https://doi.org/10.1016/j.linged.2019.03.002

Cavagnetto, A. R., Hand, B., & Premo, J. (2020). Supporting student agency in science. *Theory into Practice*, *59*(2), 128–138. doi=10.1080/00405841.2019.1702392

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2

Chiu, M. M. (2008a). Flowing toward correct contributions during group problem solving: A statistical discourse analysis. *Journal of the Learning Sciences*, *17*(3), 415–463. https://doi.org/10.1080/10508400802224830

Chiu, M. M. (2008b). Effects of argumentation on group micro-creativity: Statistical discourse analyses of algebra students' collaborative problem solving. *Contemporary Educational Psychology*, *33*(3), 382–402. https://doi.org/10.1016/j.cedpsych.2008.05.001

Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. *CBE—Life Sciences Education*, *14*, 1–13. https://doi.org/10.1187/cbe.14-10-0167

Curşeu, P. L., Ilies, R., Vîrgă, D., Maricuţoiu, L., & Sava, F. A. (2019). Personality characteristics that are valued in teams: Not always "more is better"? *International Journal of Psychology*, *54*(5), 638–649. https://doi.org/10.1002/ijop.12511

Donker, A. S., De Boer, H., Kostons, D., Van Ewijk, C. D., & van der Werf, M. P. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, *11*, 1–26. https://doi.org/10.1016/j.edurev.2013.11.002

Graham, S., & Echols, L. (2018). Race and ethnicity in peer relations research. In Bukowski, W. M., Laursen, B., & Rubin, K. H. (Eds.), *Handbook of peer interactions, relationships, and groups (2nd edition)* (pp. 590–614). New York, NY: Guilford.

Grunspan, D. Z., Eddy, S. L., Brownell, S. E., Wiggins, B. L., Crowe, A. J., & Goodreau, S. M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PLoS ONE*, *11*(2), e0148405. https://doi.org/10.1371/journal.pone.0148405

Hand, B., Chen, Y. C., & Suh, J. K. (2021). Does a knowledge generation approach to learning benefit students? A systematic review of research on the science writing heuristic approach. *Educational Psychology Review*, *33*, 535–577. https://doi.org/10.1007/s10648-020-09550-0

Haskell, R. E. (2001). *Transfer of learning: Cognition, instruction, and reasoning*. San Diego, CA: Academic Press.

Howe, C. (2014). Optimizing small group discourse in classrooms: Effective practices and theoretical constraints. *International Journal of Educational Research*, *63*, 107–115. https://doi.org/10.1016/j.ijer.2013.03.011

Howe, C., & Zachariou, A. (2019). Small-group collaboration and individual knowledge acquisition: The processes of growth during adolescence and early adulthood. *Learning and Instruction*, *60*, 263–274. https://doi.org/10.1016/j.learninstruc.2017.10.007

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365–379. https://doi.org/10.3102%2F0013189X09339057

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, *44*(1), 46–53. https://doi.org/10.3102%2F0013189X15569530

Larrain, A., Freire, P., Grau, V., L´opez, P., Salvat, I., Silva, M., & Gastellu, V. (2018). The effect of peer-group argumentative dialogue on delayed gains in scientific content knowledge. In Grau, V., & Whitebread, D. (Eds.), *Relationships between classroom dialogue and support for metacognitive, self- regulatory development in educational contexts* (New Directions for Child and Adolescent Development 162, pp. 67–87).

Majolo, B., Ames, K., Brumpton, R., Garratt, R., Hall, K., & Wilson, N. (2006). Human friendship favours cooperation in the iterated prisoner's dilemma. *Behaviour*, *143*(11), 1383–1395. Retrieved July 15, 2021, from www.jstor.org/stable/4536413

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral & Brain Sciences*, *34*(2), 57–74. https://dx.doi.org/10.1017/S0140525X10000968

National Research Council. (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academies Press.

Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review*, *27*, 645–656. doi 10.1007/s10648-015-9312-8

Nowak, M., & Highfield, R. (2011). *SuperCooperators: Altruism, evolution, and why we need each other to succeed*. New York, NY: Simon and Schuster.

Pai, H. H., Sears, D. A., & Maeda, Y. (2015). Effects of small-group learning on transfer: A meta-analysis. *Educational Psychology Review*, *27*, 79–102. doi 10.1007/s10648-014-9260-8

Premo, J., Cavagnetto, A., Collins, L., Davis, W. B., & Offerdahl, E. (2021). Discourse remixed: Using interdependency to shift student learning through talk. *Journal of Experimental Education*. Retrieved December 2, 2021, from https://doi.org/10.1080/00220973.2021.1993771

Premo., J., Cavagnetto, A., & Davis, W. (2018b). Promoting collaborative classrooms: The impacts of interdependent cooperative learning on undergraduate interactions and achievement. *CBE—Life Sciences Education*, *17*, ar32. Retrieved from www.lifescied.org/doi/10.1187/cbe.17-08-0176

Premo, J., Lamb, R., & Cavagnetto, A. (2018a). Conditional cooperators: Student prosocial dispositions and their perception of the classroom social environment. *Learning Environments Research*, *21*(2), 229–244. https://doi.org/10.1007/s10984-017-9251-z

Ruiz-Gallardo, J. R., & Reavey, D. (2019). Learning science concepts by teaching peers in a cooperative environment: A longitudinal study of preservice teachers. *Journal of the Learning Sciences*, *28*(1), 73–107. https://doi.org/10.1080/10508406.2018.1506988

Saleh, N., & Large, A. (2011). Collaborative information seeking behavior of engineering students. In *Proceedings of the Annual Conference of CAIS at New Orleans, LA.* https://doi.org/10.29173/cais662

Schnittka, J., & Schnittka, C. (2016). "Can I drop it this time?" Gender and collaborative group dynamics in an engineering design-based afterschool program. *Journal of Pre-College Engineering Education Research*, *6*(2), 1–24. https://doi.org/10.7771/2157-9288.1120

Scribner, J. P., Sawyer, R. K., Watson, S. T., & Myers, V. L. (2007). Teacher teams and distributed leadership: A study of group discourse and collaboration. *Educational Administration Quarterly*, *43*(1), 67–100. https://doi.org/10.1177/0013161X06293631

Sears, D. A., & Reagin, J. M. (2013). Individual versus collaborative problem solving: Divergent outcomes depending on task complexity. *Instructional Science*, *41*, 1153–1172. doi 10.1007/s11251-013-9271-8

Sfard, A. (2015). Why all this talk about talking classrooms? Theorizing the relation between talking and learning. In Resnick, L. B., Asterhan, C., & Clarke, S. N. (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 245–253). Washington, DC: AERA.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, *21*, 43–69.

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, *43*(1), 1–28. https://doi.org/10.1016/0749-5978(89)90055-1

Sober, E., & Wilson, D. S. (1999). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Staub, N. L., Poxleitner, M., Braley, A., Smith-Flores, H., Pribbenow, C. M., Jaworski, L., … Anders, K. R. (2016). Scaling up: Adapting a phage-hunting course to increase participation of first-year students in research. *CBE—Life Sciences Education*, *15*(2), ar13. https://doi.org/10.1187/cbe.15-10-0211

Theobald, E. J., Eddy, S. L., Grunspan, D. Z., Wiggins, B. L., & Crowe, A. J. (2017). Student perception of group dynamics predicts individual performance: Comfort and equity matter. *PLoS ONE*, *12*(7), e0181336. https://doi.org/10.1371/journal.pone.0181336

Van den Bossche, P., Gijselaers, W. H., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments: Team learning beliefs and behaviors. *Small Group Research*, *37*(5), 490–521. https://doi.org/10.1177/1046496406292938

Webb, N. M. (1982). Student interaction and learning in small groups. *Review of Educational Research*, *52*(3), 421–445. https://doi.org/10.3102/00346543052003421