# Which Group Dynamics Matter: Social Predictors of Student Achievement in Team-Based Undergraduate Science Classrooms

**Joshua Premo,[†*] Brittney N. Wyatt,[†] Matthew Horn,[‡] and Heather Wilson-Ashworth[†]**
[†]Department of Biology and [‡]Department of Chemistry, Utah Valley University, Orem, UT 84058

## ABSTRACT
**While group work in undergraduate science education tends to have overall benefit, less is known about the specific peer–peer dynamics that optimize learning during group interaction. The current study used peer ratings and self-reported data from 436 students enrolled in team-based undergraduate science courses (biology or chemistry) to determine group dynamics that predicted both willingness to work with peers in the future and individual achievement in the course. Results show that greater personal connection and contributions predicted willingness to work with a group member ($R^2_{biology} = 0.75$; $R^2_{chemistry} = 0.59$). While active contribution to a group predicted greater achievement, more non-content interactions (e.g., encouragement, listening to feedback, being polite) predicted lower achievement, despite these being on-task and relevant. Additionally, having group members who were willing to continue working with a student was a positive predictor of that student's achievement regardless of course. Strikingly, students in chemistry were significantly less willing to work with women in their groups compared with men. Finally, not all forms of group conflict predict decreased achievement. These findings highlight group factors such as student behavior within the group, aspects of the group social environment, and peer support that can be targeted for optimizing undergraduate science learning.**

## INTRODUCTION
Strides have been made over the last decade to shift the nature of undergraduate science instruction in the United States. While there is still room for growth, adoption of more "active" styles of engaging students in learning have increased relative to traditional lecture (American Association for the Advancement of Science [AAAS], 2019). One often underappreciated benefit of using more active instructional strategies is increased opportunities for students to experience the social aspects of science (Chinn and Malhotra, 2002; Tanner *et al.*, 2003; Ford, 2008; Manz, 2015). Teamwork is one specific social aspect of the sciences and science, technology, engineering, and mathematics (STEM) careers that is crucial for scientific productivity (Wang, 2016). However, not all groups perform equally (Salas *et al.*, 2018), and a host of negative repercussions can result from formally requiring students to work in groups (Blumen *et al.*, 2014; Eddy *et al.*, 2015; Nokes-Malach *et al.*, 2015; Chang and Brickman, 2018). This means that understanding how group dynamics can be optimized (i.e., maximizing positive and minimizing negative effects) will become increasingly important as active instructional practices become more common in university science programs.

Guided inquiry (GI) is an instructional method that emphasizes student-centered group learning while the instructor serves as a facilitator (Lewis and Lewis, 2005; Hein, 2012; Walker and Warfa, 2017). GI makes use of cooperative learning theory in higher education (Johnson *et al.*, 2014) by adopting a cooperative goal-orienting structure to engage students in group work. This cooperative structure generates positive

**Kristy Jean Wilson,** *Monitoring Editor*
Submitted Jun 30, 2021; Revised May 16, 2022; Accepted Jun 15, 2022
CBE Life Sci Educ September 1, 2022 21:ar51
DOI:10.1187/cbe.21-06-0164
*Address correspondence to: Joshua Premo (JPremo@uvu.edu).

interdependence wherein peers must rely on one another to be successful. During GI, groups are encouraged to *construct* knowledge by asking and defining questions, interpreting data, arguing from evidence, and practicing oral/written communication of scientific ideas (Lawson, 1995) as opposed to *receiving* knowledge in a traditional lecture approach. While GI draws from well-researched science education techniques, it has been shown to only have minimal impact on student learning (Walker and Warfa, 2017). The reasons for this have not been well explored, but studies have focused on characteristics of GI that may impact method efficacy, including student attitudes (Schroeder and Greenbowe, 2008) and implementation (Simonson and Shadle, 2013; Balasubramaniam, 2015). Little attention has been given to the contribution of group dynamics to learning through GI. Thus, the current study seeks to explore how differential student experiences and behaviors in groups predict end-of-course achievement. This will not only provide insight into the social side of GI learning but also more broadly inform our understanding of group work in higher education.

The current study examined the extent to which a multitude of factors (group behaviors, self-perceptions of group-work ability, and group experiences) predicted students' performance in two introductory undergraduate science courses (biology for non-science majors and chemistry for science majors) that extensively used GI pedagogy (present in more than 90% of class meetings). To better understand how group factors impact student learning through GI, we sought to answer the overarching question "What group dynamics positively predict both student engagement with peers and achievement in undergraduate science?" This was answered by examining data from 44 groups in biology and 65 groups in chemistry (436 students total) collected in a pretest–posttest design.

We began by investigating how group social factors predicted student willingness to work with fellow group members. Research question 1 (RQ1) asked: "What qualities do students use to determine if they are willing to work with a group member in the future?" We hypothesized that student would be more willing to work with their groups in the future if they felt that group members contributed directly to their learning. This is because individuals invest based on the likelihood of a return (i.e., conditional cooperation; Keser and van Winden, 2000; Fehr and Fischbacher, 2004; Premo *et al*., 2018c), which is most likely to come from a group member directly contributing to knowledge construction within the group.

Following initial support for this hypothesis, we then explored the contributions that different group-work qualities had on a student's overall course achievement in research question 2 (RQ2): "Which group-work qualities predict student achievement (final grade point average in the course)?" We generated three hypotheses regarding RQ2. First, we hypothesized that having one student willing to work with another would be fundamentally important to the first student's achievement in a team-based course setting. Second, we hypothesized that the quality of student engagement with group members and with content ideas (i.e., contributions) would predict greater course achievement (Summers and Volet, 2010; Jurkowski and Hänze, 2015; Curşeu *et al.*, 2018; Premo *et al.*, 2021; Cavagnetto *et al.*, 2022). Third, we hypothesized that relationship conflict in the group would predict lower achievement, based on previous literature that described negative effects of this form of group conflict (Darnon *et al.*, 2002; Buchs *et al.*, 2010; Sohr *et al.*, 2018).

## LITERATURE REVIEW

### Group Work in Undergraduate Biology and Chemistry Education

Over the last decade, the landscape of undergraduate science education has begun to shift toward greater reliance on peer–peer dynamics during learning. A recent report by the AAAS noted a real, but slow, shift toward integrating group work in undergraduate science education (AAAS, 2019). Self-reported integration of group work in biology increased from 29% to 50.4% during 2004–2014. In the physical sciences (including chemistry), this shift was smaller, moving from 20% to 33% during the same period (Eagan, 2016). Large-scale observations of more than 2000 undergraduate STEM classes found similar rates of group work, with around 50% of biology and 36.5% of chemistry courses integrating some form of group work into instruction (Stains *et al.*, 2018). While we know that group work in undergraduate science courses has increased, we know less about what type of specific group-work strategies are being used (Stains *et al.*, 2018; AAAS, 2019). It has been reported that general strategies in many biology and chemistry classrooms include: 1) informal peer–peer discussions around instructor prompts, 2) assigned groups working on a task (e.g., worksheet), and 3) group projects. However, instructor adherence to specific group-work models (e.g., cooperative learning; Johnson *et al.*, 2014; and peer instruction; Lasry *et al.*, 2008) is less clear. One specific group-work strategy noted as being present by both Stains *et al.* (2018) and AAAS (2019) is process-oriented guided-inquiry learning (POGIL; Moog *et al.*, 2006). At present, at least 300 undergraduate instructors in the United States are implementing POGIL in their classrooms, with most doing so two to three times per week (AAAS, 2019). This means that understanding the dynamics of what makes groups effective for learning is critical to undergraduate science education.

Understanding group dynamics is not only important for increasing student learning, but also for the prevention of a host of negative social and learning consequences (Blumen *et al.*, 2014; Eddy *et al.*, 2015; Nokes-Malach *et al.*, 2015; Chang and Brickman, 2018). From a social perspective, putting students into groups can encourage a diffusion of responsibility that can result in disengagement (Pieterse and Thompson, 2010; Maiden and Perry, 2011), increased anxiety and exclusion (Eddy *et al.*, 2015), and negative impacts through relationship conflict (Darnon *et al.*, 2002; Buchs *et al.*, 2010; Sohr *et al.*, 2018). It has even been found that some negative social consequences can remain regardless of active efforts by instructors (e.g., use of roles, peer evaluation, group contracts) to reduce them (Chang and Brickman, 2018).

Negative social consequences of group work coincide with potential negative impacts on student cognitive processes (Nokes-Malach *et al.*, 2015). For example, it has been suggested that the requirement of social interactions in student groups reduces cognitive resources available for cognitive processing (Kirschner *et al.*, 2009; Nokes-Malach *et al.*, 2015). Additionally, the process of interacting can result in retrieval disruption, making student recall of prior knowledge more difficult (Zhang *et al.*, 2017). These factors may contribute to the consistent finding in cognitive psychology that groups underperform in comparison to what one might expect given the combined knowledge of each individual (Marion and Thorley, 2016). Thus, group interactions introduce variables that must be accounted for to optimize learning.

The unintended consequences of grouping students in an environment that introduces new social variables put instructors in a position where they must maximize beneficial collaborative interactions between students and minimize negative consequences (Andrews and Rapp, 2015). Yet this cannot take place without a firm understanding of how different group dynamics impact engagement and thus the potential for learning. At a basic level, students recognize that the quality of learning with peers can be inconsistent, and thus students are not always willing to engage cooperatively (Eddy *et al.*, 2015; Chang and Brickman, 2018; Premo *et al*., 2018c). Interdependency theory (Balliet *et al.*, 2017) argues that individuals invest energy and resources into helping others in situations where they will receive a return (i.e., conditional cooperation). For example, the more students perceive a learning benefit that comes from peer interactions, the more they are willing to invest time and effort into working with their peers (Premo *et al.*, 2018a, b, c). This finding has also been supported by observing student cooperative behavior in classrooms. The interaction between likelihood of return perceived by students (i.e., reciprocity) and investment in peers has been found to predict observed whole-class prosocial behavior (i.e., disposition toward working with peers) in undergraduate science laboratory classrooms (Premo *et al.*, 2018a).

The fact that students will invest less in working with peers under circumstances when doing so is less valuable to their own learning has the potential to undermine group work. Thus, knowing which group dynamics preferentially impact willingness to work with a group member is a foundational step toward ensuring effective group dynamics. This goal is targeted by the first research question of the current study: "What qualities do students use to determine if they are willing to work with a group member in the future?"

## The GI Approach

One early effort to shift chemistry instruction away from a purely didactic lecture to more student-centered in nature was through the use of GI (Farrell *et al.*, 1999). This technique uses many elements of cooperative learning (e.g., instructor as a facilitator, use of groups, and specific member roles; Johnson *et al.*, 2014) to support students in completing activities that take them through a three-phase learning cycle: 1) exploration, 2) concept invention, and 3) application (Lawson, 1995). Students begin with the exploration phase, in which they interpret a model or data, then move to a concept invention phase, in which they answer questions to reveal important relationships between concepts, and finally apply these relationships in an application phase at the end of the session. This cycle is used in GI at least once in each class to allow students to construct understandings of major principles throughout the course.

Despite the significant use of GI methodologies (such as POGIL), and in contrast to the widespread success of cooperative learning (Kyndt *et al.*, 2013) from which much of the technique is founded, a systematic review of the technique found that only minor learning gains result from GI (Walker and Warfa, 2017). In contrast, Hein (2012) reported some differences between traditional lectures and POGIL classrooms, but critically, this study does not demonstrate a significant difference in learning when a standardized outcome (same examination type) is used for comparison. In a biochemistry POGIL course, Bailey *et al.* (2012) found that student learning increased from pretest to posttest, but lacked a comparison group to contextualize whether these gains were more than those found via traditional approaches (Bailey *et al.*, 2012).

However, not all studies agree that only minor learning gains occur through GI methodology. Significant impacts of peer-led team learning (PLTL, an approach building from GI that uses peer facilitation of six to eight students per group; Eberlein *et al.*, 2008) have been found in general chemistry classes. Lewis and Lewis (2005) found that, after statistically controlling for student academic ability, the amount of PLTL sessions a student attended positively predicted the student's score on three out of five course examinations. In addition, when overall examination performance was compared between PLTL and a control group, PLTL students scored significantly higher in all five examinations. Similar success was reported by Farrell *et al.* (1999) in their implementation of GI within a general chemistry course. In comparison to traditional lecture semesters, students who participate in GI have lower ratings of both course failure and withdrawal. More recent work by Bokosmaty *et al.* (2019) also reported significantly increased percentages of high-performing students in three chemistry courses following the adoption of a variety of instructional changes, including significant online pre-instructional supports (i.e., a partially flipped classroom in which students engaged with material, including discussion boards, concept videos, or pre-assessments, before class), as well as in-class activities that are significantly influenced by POGIL approaches (Bokosmaty *et al.*, 2019). Yet caution should be taken in the attribution of increased student performance to POGIL, as this is not the only experimental manipulation. When taken together, these studies suggest that GI approaches have the potential to increase student learning, yet impacts are highly inconsistent across studies.

We suggest that one factor contributing to this inconsistency in student learning is the variable quality of student–student engagement in groups. Recent qualitative work in chemistry supports the idea that group dynamics are critical to the GI learning processes (Warfa *et al.*, 2018). By examining group discourse, it was found that specific norms of understanding emerge at the group level during GI and shape how groups developed an understanding of chemical phenomena. This result shows that variances exist in how teams engage with content during GI and suggests that these variances may impact learning processes. Thus, a more quantitative examination of which specific forms of variance predict greater student learning, and to what degree, is needed for optimization of GI approaches. RQ2, "Which group-work qualities predict student achievement?," provided this quantitative examination.

## Sources of Variance in Group Learning

Current research examining the relationship between group work and learning has centered on interaction behaviors such as contributions of group members (Summers and Volet, 2010), the general group social environment (e.g., negative aspects; Monson, 2019), social comfort (Theobald *et al.*, 2017), and discourse between group members (Jurkowski and Hänze, 2015; Curşeu *et al.*, 2018; Premo *et al.*, 2021; Cavagnetto *et al.*, 2022). These factors have the potential to impact the efficacy of GI learning.

How undergraduate students choose to interact in groups can account for some differences in student learning. One investigation by Summers and Volet (2010) found that the number of high-level contributions provided by a student related to the student's performance (greater student contribution to knowledge in the group related to a higher grade). This finding suggests that student contributions and engagement in the most productive sections of group conversation may be particularly impactful to student learning. The idea that differential engagement may account for learning differences is mirrored in Jurkowski and Hänze's (2015) investigation into the role of transactional communication (i.e., building on a peer's idea) in undergraduate psychology student groups. They not only found that transactional communication can be increased via specific training, but engagement in this practice had positive effects on knowledge acquisition in a group. Both studies suggest that students may learn more in groups when they are explicitly and actively engaging with ideas and contributing during group work.

More general social experiences in teams can also impact learning. For example, Monson (2019) found that undergraduate sociology students' experiences during a group research project could account for their performance on a related final paper. After controlling for several factors statistically (e.g., individual characteristics, group composition, task, group grades), the author found that individuals who reported negative experiences did not score significantly lower on the outcome measure. Yet if their groups reported a negative experience, and thus did poorly on the group outcome, individual students' performances were worse. In other words, students in negative group climates learn less from the group and subsequently perform worse. In alignment, Theobald *et al.* (2017) found that greater social comfort in active-learning groups predicts greater student content mastery (score increase of 27.5%). Taken together, these results suggest that the general social environment of a group can account for some of variance in group learning.

Studies that account for more specific peer–peer dynamics in groups are less common. One example is Curşeu *et al.*'s (2018) investigation of how discussion quality mediated the relationship between motivation to engage in cognitive activities and performance on a written group report. Their results showed that the quality of a group's discussion completely mediates the motivation to engage in cognitive activities and group performance. This suggests that students can be motivated to engage with cognitive activities in a group, but without high-quality discussion, motivation cannot drive performance. In total, contemporary research on factors impacting student learning in teams suggests both the quality of student engagement in groups and the general character of the group's social environment have the potential to impact individual student achievement.

## METHODS

This study seeks to investigate the overall question "What group dynamics positively predict both student engagement with peers and achievement in undergraduate science?" The following section reports on the data collected and participants, instructional modes of the courses, methods of measurement, and methods of statistical analysis.

### Data Collection and Participants

Data were collected from 109 student groups with a total of 436 students either in a chemistry course for science majors (261 students total) or in a biology course for non-science majors (175 students total). These classes were chosen due to the highly team-based nature of both courses and the similarity in instructional approach. Eighty-nine percent of student groups were four members in size for the duration of the course, with some being reduced as a result of course withdrawal. All data were collected in a pretest–posttest format, with the pretest being completed in the first 2 weeks of the semester, and the posttest being completed in the last 2 weeks of the semester.

All students completed ratings of group behavior using the Comprehensive Assessment of Team Member Effectiveness (CATME) peer evaluation assessment (Ohland *et al.*, 2012; see Table 1). All measures were completed by students on the CATME online platform (catme.org). Almost all students who consented to the study and completed the CATME peer evaluation assessment were included in the analysis for RQ1 (What qualities do students use to determine if they are willing to work with a group member in the future?). A minority of students who did consent ($n = 27$) did not complete the measures and could not be included. During the initial semester of data collection, the online CATME peer evaluation only contained questions related to seven peer-rating group qualities (Ohland *et al.*, 2012): 1) contributing to group's work; 2) interacting with group mates; 3) keeping the group on track; 4) expecting quality; 5) having relevant knowledge, skills, and abilities; 6) willing to work the group member in the future; and 7) personal connection to the group member. As more data were collected in later semesters, the online CATME peer evaluation was updated with an additional seven group qualities: 1) relationship conflict, 2) process conflict, 3) task conflict (Jehn and Mannix, 2001), 4) group interdependence (Vegt *et al.*, 2001), 5) group commitment (Carless and De Paola, 2000), 6) interpersonal cohesiveness (Loughry and Tosi, 2008), and 7) group attraction (Loughry and Tosi, 2008; see Table 1). As these additional questions were not originally available to students on the online CATME platform, groups from earlier semesters could not be included in the analysis for RQ2 (Which group-work qualities predict student achievement?). This reduced the sample size to 116 biology students (29 groups) and 185 (42 groups) chemistry students in the analysis for a total of 14 group qualities.

Demographics for participants included in the analysis of both RQ1 and RQ2 can be seen in Table 2. Despite a significant reduction in participants from RQ1 to RQ2, demographic compositions of participants and average course performance remained relatively consistent between the research questions. There are several important demographic differences between the chemistry and biology courses that are important to note: 1) the chemistry course had 65–71% men compared with the more even distribution of 53–54% men in the biology course (nonbinary and/or gender queer were not options for students

**TABLE 1. Online CATME peer evaluation assessment**

| Section | Quality | Example items | Type[a] |
|---|---|---|---|
| CATME peer ratings (Ohland *et al.*, 2012) | Contributing to the group's work | Does more or higher-quality work than expected./Makes important contributions that improve the team's work. | PR & SR |
| | Interacting with group mates | Asks for and shows an interest in teammates' ideas and contributions./Provides encouragement or enthusiasm to the team. | PR & SR |
| | Keeping the group on track | Makes sure that teammates are making appropriate progress./ Gives teammates specific, timely, and constructive feedback | PR & SR |
| | Expecting quality | Motivates the team to do excellent work./Believes that the team can do excellent work. | PR & SR |
| | Having relevant knowledge, skills, and abilities | Demonstrates the knowledge, skills, and abilities to do excellent work./Able to perform the role of any team member if necessary | PR & SR |
| | Willing to work with them in the future | I would gladly work with this individual in the future./If I were selecting members for a future work team, I would pick this person. | PR |
| | Personal connection | I like this person as an individual./I consider this person to be a friend. | PR |
| Group conflict (Jehn and Mannix, 2001) | Relationship conflict | How often do people get angry while working in your group?/ How much emotional conflict is there in your work group? | SR |
| | Process conflict | How often are there disagreements about who should do what in your work group?/How much conflict is there in your group about task responsibilities? | SR |
| | Task conflict | How much conflict of ideas is there in your work group?/How often do people in your work group have conflicting opinions about the project you are working on? | SR |
| Group interdependence (Vegt *et al.*, 2001)[b] | | I depend on my teammates for the completion of my work./I have to work closely with my teammates to do my work properly | SR |
| Group commitment (Carless and De Paola, 2000)[b] | | Our team is united in trying to reach its goals for performance./I'm unhappy with my team's level of commitment to the task. (scale reversed) | SR |
| Group work (Loughry and Tosi, 2008) | Interpersonal cohesiveness | Team members like each other./Team members get along well | SR |
| | Group attraction | Team members get to participate in enjoyable activities./Team members like the work that the group does. | SR |

[a]PR, peer rating; SR, self-report.
[b]Some sections of online CATME Peer Evaluation Assessment just measured one group quality.

to select due to a limitation of inclusive demographic selections). 2) The chemistry course had a more ethnically diverse student population, with ~22% of students identifying as non-Caucasian compared with ~14% in biology. 3) Students' final grade point averages in chemistry were lower than in biology. 4) In the biology course, a majority of students were in either their first or second year, while students in the chemistry course were distributed across the four years.

### Context

*Instructor Competency.* The instructors of both the biology (H.W.-A.) and chemistry (M.H.) courses have received significant training in POGIL. Both have trained at POGIL regional conferences and have obtained certification to train other practitioners. In addition, both professors have facilitated six POGIL training conferences, instructing more than 300 high school and university teachers in the POGIL pedagogy during the period of data collection. H.W.-A. holds a PhD in cellular biology and physiology and has 20 years of experience teaching biology at the undergraduate level. M.H. holds a PhD in chemistry and has 20 years of experience teaching chemistry at the undergraduate level.

*Course Descriptions.* Participants in the study were either enrolled in introductory biology for nonmajors or introductory chemistry for majors. See Table 3 for a comparison of the two courses. While the demographics of students and enrollment size varied between the two courses (described earlier), many similarities in instructional approach were present. One critical difference to note is that the biology course was structured to encourage higher dependence on group members. Specifically, one biology assignment was randomly gathered from each group to represent the efforts of the entire group. In the chemistry course, students knew whose work would be collected for the group. Additionally, students in the biology course completed part of their final course examination as a group, which constituted 30% of the final exam score. More details as to the curriculum used in each course can be found in the Supplemental Material.

### Measures

*Instrument Description.* Students completed the CATME peer evaluation assessment at two time points (within the first 2 weeks of the semester and during the last 2 weeks) in both the biology and chemistry courses via the CATME online platform (catme.org). The CATME platform first trained students by

**TABLE 2. Demographics**

| | Biology | | Chemistry | |
|---|---|---|---|---|
| Variable | RQ1 | RQ2 | RQ1 | RQ2 |
| 1. Number | 175 | 116 | 261 | 185 |
| 2. *Gender (women, men) | 47%, 53% | 46%, 54% | 29%, 71% | 35%, 65% |
| 3. Academic standing | | | | |
| Year 1 | 58% | 58% | 16% | 10% |
| Year 2 | 33% | 35% | 46% | 46% |
| Year 3 | 6% | 7% | 23% | 24% |
| Year 4 | 3% | 0% | 15% | 19% |
| 4. Average age | 20.7 | 20.5 | 22.3 | 22.59 |
| 5. *Ethnicity | | | | |
| Asian/Asian American | 2% | 2% | 1% | 1% |
| Hispanic/Latino/a | 8% | 6% | 13% | 12% |
| White/Caucasian | 85% | 89% | 76% | 79% |
| Pacific Islander | <1% | <1% | 1% | <1% |
| Multiracial/other | 4% | 2% | 5% | 5% |
| Black/African American | 0% | 0% | 1% | <1% |
| Native American | 0% | 0% | 1% | 0% |
| 6. Average course grade | 3.10/4 | 3.08/4 | 2.62/4 | 2.66/4 |

having them rate hypothetical students' behaviors before asking students to rate their own group members. For RQ1, students responded to 11 questions (five peer rating and six Likert type), while for RQ2, students answered an additional 23 Likert survey questions assessing their views of, and experiences with, their groups (where 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, and 5 = strongly agree). As a reminder, the CATME peer evaluation assessment did not include the additional questions on their online platform until later in the study, thus RQ2 has a smaller sample size. See Table 1 for more details on these measures and example descriptors/items used in the CATME peer evaluation assessment. The full CATME peer evaluation assessment and more details regarding its implementation can be found in the Supplemental Material.

**TABLE 3. Comparison of courses**

| Course element | Biology course | Chemistry course |
|---|---|---|
| Population served | Non-science majors | Science majors |
| Size and demographics | 50–65 students per section, majority first year students | 100–150 students per section, majority second year students |
| Use of GI | >85% of course material is taught via GI and is present in almost every class during the semester (method highly informed by Moog *et al.*, 2006) | |
| Permanence of groups | Groups are established at the start of the semester and largely remain the same throughout | |
| Composition of groups | 4 members per group<br>Based on study time availability Avoidance of minority women presence in a group (e.g., 3 men and 1 woman) | At least 4 members per group, larger groups were uncommon<br>Based on study time availability |
| Roles assigned to team members (rotated per class session) | Manager (keeps the team on task and asks questions for the group)<br>Spokesperson (presents the team's answers)<br>Reader (reads questions out loud to the group) | |
| | Scribe (ensures team consensus on answers as assignment collection for grading is random) | Recorder (turns in the assignment for the team's grade) |
| Grading | 10% GI activities* (group grade)<br>10% study modules<br>10% research study<br>20% homework<br>30% individual exams<br>20% individual and group final exam<br>GI activities are group grades and students do not know which student's assignment will be collected.<br>Final examination is partially completed as a group (70% of the exam was based on individual score. while 30% came from the group score). | 10% reflections<br>20% GI activities** (group grade)<br>20% homework<br>20% individual exams<br>30% individual final exam<br>GI activities are group grades, and students know which student's assignment will be collected. |

*Measurement Quality.* Recent research in undergraduate science education has argued that structural validity of an instrument cannot be assumed across different educational contexts and that factor analysis should be conducted and reported on when an instrument is used in a new context (Knekta *et al.*, 2019). Thus, we assessed the quality of measurement models for all aspects of group work being measured via confirmatory factor analysis (CFA) and Cronbach's alpha reliability. The Cronbach's alpha values were largely in the range of acceptability ($\alpha$ = 0.70 to 0.89), with one subscale being mediocre at 0.66 but sufficient for exploring new areas of education research (Hair, 2011). CFA was then used to test a hypothesized latent structure of the CATME peer evaluation assessment and provided evidence of structural validity (Brown, 2015). Evaluation of the CFA measurement model quality was conducted using fit indices. Essentially, fit indices indicate the extent to which individual survey questions/ratings align with their hypothesized structure (i.e., we would expect all questions that measure relationship conflict to group together, as they are designed to measure the same factor, and not group with those designed to assess a different factor like interdependency).

We used three common fit indices to assess the quality of the CATME peer evaluation assessment for RQ1 and the additional group-work questions added to the CATME peer evaluation assessment for RQ2 survey subscales independently. For the root mean-square error of approximation (RMSEA), <0.06 is considered to be a threshold value indicating a good fit for the model (Hu and Bentler, 1999), with values up to 0.08 being considered acceptable (MacCallum *et al.*, 1996). A comparative fit index (CFI) of 0.90 or above indicates adequate model fit (van de Schoot *et al.*, 2012). The last fit index reported was the standardized root mean-square residual (SRMR). Values for the SRMR below 0.05 indicate good model fit (Diamantopoulos *et al.*, 2000), but up to 0.08 is considered acceptable (Hu and Bentler, 1999). All models were run using a robust maximum-likelihood estimator to correct for variations in normality. While the above cutoff values were used, we recognize that no single cutoff value is appropriate to be used across all contexts (Hancock and McNeish, 2017) and that acceptable overall fit can occur with a weaker index if the remaining indices support acceptance of the model (Brown, 2015). Thus, a majority rule was used where at least two out of three fit indices had to indicate at least adequate fit to be used in the results.

CFA results for the RQ1 CATME peer evaluation assessment included RMSEA = 0.10, CFI = 0.96, and SRMR = 0.03. Given the cutoff values chosen, the RMSEA value indicates poor fit, the CFI indicates excellent fit, and the SRMR indicates good fit. Thus, the latent factors from Ohland *et al.* (2012) are used in the analyses, given that 2/3 of the fit indices had at least acceptable fit. CFA results for the RQ2 CATME peer evaluation assessment survey subscales included RMSEA = 0.03, CFI = 0.97, and SRMR = 0.04. All these values indicate that the hypothesized latent subscale structure for items had good fit and support the analysis using the specified latent factors discussed in *Measures*.

Assessments contributing to the measure of course achievement include both GI teamwork activities and examinations. These are based on published curricular materials developed by experts in their respective scientific fields, which support both content and construct validity. For more information on these, see the Supplemental Material.

### Analyses

*Regression Analysis and Model Fitting.* We used multiple regression to examine factors that predicted dispositions toward working with a group member and end of course achievement. A large number of potential predictors were collected, and regression was used to examine the relative impact of these factors simultaneously (Field, 2013). A separate model selection process was completed for each research question and subject area (four models total). The process of arriving at the best-fitting model was the same for each model and followed the steps outlined in Jordt *et al.* (2017).

After we generated a highly inclusive initial model for predicting each research question by course (see the left column of Table 4 for all variables initially included), backwards stepwise regression was used to remove variables from the model until we had the best-supported model based on Akaike information criterion (AIC) validation. AIC is a relative fit index. This means that AIC values only indicate goodness of fit when compared with the AIC values of other potential models; lower values indicate a more optimal model (Burnham and Anderson, 1998). We adopted the model with the lowest AIC, with the simpler model being accepted if $\Delta$AIC was $2 > |x| > 0$ (Jordt *et al.*, 2017). Variables are retained in the preferred model if they meet the criteria, even if they are nonsignificant predictors, as these still have explanatory power based on AIC (Burnham and Anderson, 1998). One caveat holds: To control for variance based on gender and underserved student status, these characteristics were retained in the model regardless of significance.

## RESULTS

### Biology Had More Dramatic Changes in Group Qualities throughout the Semester Than Chemistry

To begin investigating our overall research question, "What group dynamics positively predict both student engagement with peers and achievement in undergraduate science?," we first examined within each class how group-work qualities changed throughout the semester. With the updated version of the CATME peer evaluation assessment, 14 group qualities were assessed in a pretest–posttest manner (see Table 1 for all measures collected). These included qualities for which group members rated each of their peers (e.g., contributions, interactions, quality of work, personal connection) and assessed their group as a whole (e.g., conflict in the group). Due to data being collected at both the beginning and end of the semester, this allowed us to examine any change in groups during the semester.

Before assessing the changes in group qualities during the semester, we first compared pretest scores for groups in biology with those in chemistry to assess any initial differences. Analysis showed no significant differences between groups in each class at the start of the semester, indicating that any significant changes at the end of the semester were not due to any inherent qualities. Of the 14 group qualities assessed, eight categories demonstrated a statistically significant change in the biology course, while six were identified in the chemistry course (see Figure 1). Overall, individuals in both biology and chemistry rated their group members higher in group-work abilities (i.e., contribution; interaction; keeping the group on track; expecting quality group work; and having relevant knowledge, skills, and abilities) at the end of the semester than at the start of the

**TABLE 4. Summary of research questions, predicting variables, and results by class (B = biology, C = chemistry)[a]**

| RQ1: What qualities do students use to determine if they are willing to work with a group member in the future? | | | |
|---|---|---|---|
| **Qualities assessed for each individual group member** | **Qualities that determined willingness to work in the future** | **B** | **C** |
| Student's contributions | More willing to work with the student if student contributed more | X | X |
| Personal connection to the student[b] | More willing to work with the student if they felt connected to the student | X | X |
| Student's interactions with group[b] | n.s. | | |
| Student's efforts to keep the group on track[b] | n.s. | | |
| Student's expectations of quality work[b] | n.s. | | |
| Student's knowledge, skills, and abilities[b] | n.s. | | |
| Demographic factors for RQ1 | | | |
| Gender of a student | Less willing to work with the student if the student was a woman | | X |
| Student is from an underserved population | n.s. | | |
| **RQ2: Which group-work qualities predict student achievement (final grade point average in the course)?** | | | |
| **Qualities assessed for each individual/group** | **Group-work qualities that predicted student achievement** | **B** | **C** |
| Student's contributions[b,c] | Higher grade if student contributed more | X | X |
| Group members' willingness to work with a student[b] | Higher grade if group members were willing to work with the student in the future | X | X |
| Students' self-reported interactions with their group | Lower grade if students reported that they interacted more with their groups | X | X |
| Student's interactions with group[b] | Lower grade if student interacted more with the group | | X |
| Process conflict in group | Higher grade if more process conflict was present in the student's group | | X |
| Relationship conflict in group | Lower grade if more relationship conflict was present in the student's group | X | |
| Task conflict in group | Higher grade if more task conflict was present in the student's group | X | |
| Student's expectations of quality work[b,c] | Higher grade if the student expected quality group work | X | |
| Student perception of group interdependence | n.s. | | |
| Student perception of group interpersonal cohesiveness | n.s. | | |
| Student perception of group commitment to tasks | n.s. | | |
| Student perception of group attraction to tasks | n.s. | | |
| Student's efforts to keep the group on track[b,c] | n.s. | | |
| Student's knowledge, skills, and abilities[b,c] | n.s. | | |
| Demographic factors for RQ2 | | | |
| Student is from an underserved population | Lower grade if the student was from an underserved population | | X |
| Gender of a student | n.s. | | |

[a]All variables were collected in a pretest–posttest format. Detailed model results including coefficients and R-squared values can be in Tables 5, 6, 7, and 8. Rows with the darkest shading show variables that were significant for both classes, rows with lighter shading were only significant for one class, and unshaded rows were not significant for either class.
[b]These variables were measured by each member in a group rating other group members.
[c]Student self-report data for these subscales were also assessed but were not significant predictors.

semester. This highlights that groups in both the biology and chemistry classes noticed an increase in their group members' skills throughout the semester when the class was centered around GI.

While there were similarities between the group-work qualities that changed between biology and chemistry, the magnitude of this change was different between courses (see Figure 1). Students in the biology course rated the growth of their peers around five times higher in the areas of contribution (biology $d = 0.64$, chemistry $d = 0.12$), interactions (biology $d = 0.67$, chemistry $d = 0.13$), and having relevant knowledge, skills, and abilities (biology $d = 0.62$, chemistry $d = 0.11$) than chemistry groups. Biology groups had four times the growth in expecting quality work (motivates the team to do excellent work, believes that the team can do excellent work; biology $d = 0.84$, chemistry $d = 0.2$) and around 2.5 times growth in evaluating their peers' ability to stay on track (biology $d = 0.59$, chemistry $d = 0.23$) when compared with chemistry groups. Biology groups also gained personal connections to their group members that were more than three times larger than those seen in chemistry groups (biology $d = 0.43$, chemistry $d = 0.14$). Finally, biology groups experienced a significant decrease in task conflict ($d = 0.35$) during the semester. This was paired with an increase in interpersonal cohesion in their groups ($d = 0.28$). Neither of these changes were seen in chemistry groups.

When taken together, these changes in group qualities suggest a fundamental difference in group experiences in the biology course which were not experienced in the chemistry course. This may have direct implications for the role of group work in the courses including student learning and the development of

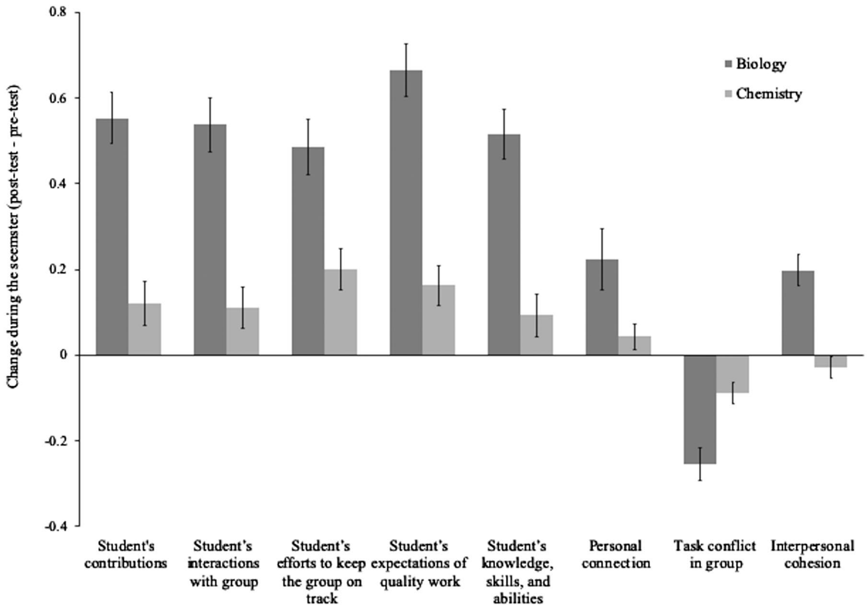


**FIGURE 1. Changes in group qualities during the semester for both the biology and chemistry courses. All differences shown were statistically significant following a Bonferroni correction ($p < 0.003$).**

teamwork skills. Potential reasons for these differences will be explored in the discussion below.

### RQ1: What Qualities Do Students Use to Determine If They Are Willing to Work with a Group Member in the Future?

Engaging students in collaborative work is a foundational step toward understanding how to promote student-centered learning. We hypothesized that student behaviors that explicitly signal greater likelihood of benefit (e.g., a student's contributions) would predict group members' willingness to work with a student in the future. For example, if a student feels that all group members are contributing, and they will not be left with the burden of all the group work, then that student will be more willing to continue to work in a group. Alternatively, students will be more willing to continually engage in group work if they feel an increase in collaboration will benefit their learning (Premo *et al.*, 2018a). While we are not able to get at the exact mechanism of the type of benefit students might be perceiving, this hypothesis is supported in both the biology and chemistry courses as described in the following sections.

**TABLE 5. Final regression model predicting willingness to work with a group member in the future (biology)[a]**

| Factor[b] | B | SE |
|---|---|---|
| Intercept | –0.16 | 0.23 |
| Student is a woman | 0.03 | 0.03 |
| Student is from an underserved population | 0.04 | 0.05 |
| Student's initial contributions to the group* | 0.13 | 0.05 |
| Student's overall contributions to the group* | 0.46 | 0.05 |
| Personal connection to the student* | 0.51 | 0.07 |

[a]$n = 175$; Adjusted $R^2 = 0.75$.
[b]Initial = beginning of semester; overall = end of semester.
*$p < 0.05$.

*Biology.* A detailed look at the final biology model (Table 5) indicates that out of all the variance in a group member's continued willingness to work with a student, the model can account for 75% of that variance (adjusted $R^2 = 0.75$). This indicates that the final regression model was a very strong model for predicting a group's continued willingness to work with a student. Additional nuances in the model results show that, while both a student's contributions at the start (initial) and end of the semester (overall) are predictive, they differ in magnitude. A student's contributions overall for the semester had a much higher impact (B = 0.46) than a student's contributions at the start of the semester (B = 0.13). Additionally, group members' personal connections to a student were important (B = 0.51), similar in magnitude to a student's contribution rating.

*Chemistry.* The final chemistry model can be seen in Table 6 and is able to account for a majority of the variance (adjusted $R^2 = 0.59$) in group members' continued willingness to work with a student. This model accounts for less variance in willingness to work than the final biology model, but still can account for 59% of this disposition, with only three predictors indicating that this is a good model overall. Despite model consistency between biology and chemistry (in both contributions and personal connection predicting continued willingness to work), two differences are present in chemistry. First, gender is predictive of a group's continued willingness to work with a student. Group members were significantly less willing to work with a woman than a man, showing significant gender bias in chemistry groups. Second, only a student's overall semester contributions to the group predicted the group's continued willingness to work, and not initial contributions at the start of the semester.

*Overall.* When comparing biology and chemistry, a group member's contributions to the group had the largest impacts on group members' continued willingness to work with a student ($B_{biology} = 0.59$, $B_{chemistry} = 0.48$). Group members' feelings of personal connection to the student were also predictive of willingness to work in both models. While being willing to work with those you have a personal connection to is not a particularly surprising factor, it was neither the sole nor the largest factor in predicting students' willingness to engage with group members in this study. Interestingly, a number of group-work qualities that have been previously tied to group effectiveness (Ohland *et al.*, 2012) did not predict willingness to work with a student in both the biology and chemistry models ($p < 0.05$). These included the student's interactions with the group; efforts to keep the group on track; expectations for quality group work; and the student's knowledge, skills, and abilities. While a student's knowledge, skills, and abilities signal the implicit

**TABLE 6. Final regression model predicting willingness to work with a group member in the future (chemistry)[a]**

| Factor[b] | B | SE |
|---|---|---|
| Intercept | 0.67 | –0.21 |
| Student is a woman* | –0.09 | –0.03 |
| Student is from an underserved population | –0.04 | –0.04 |
| Student's overall contributions to the group* | 0.48 | 0.04 |
| Personal connection to the student* | 0.43 | 0.06 |

[a]$n = 261$; Adjusted $R^2 = 0.59$.
[b]Overall = end of semester.
*$p < 0.05$.

potential for future benefit, this factor was not predictive ($p < 0.05$). This may be because what a student contributes to the group is more tangible and impactful in comparison to a student's potential knowledge, skills, and abilities. This difference has implications for how instructors can encourage more continued willingness to work within their classroom groups, which will be explored in the *Discussion*.

### RQ2: Which Group-Work Qualities Predict Student Achievement (Final Grade Point Average in the Course)?

Given that both courses are highly reliant on students working in groups (>85% of instructional time), we hypothesized that group experiences would be related to course achievement. This is supported by research suggesting that both the quality of student engagement with ideas (e.g., contributions) in groups and the group's social environment (e.g., conflict, interpersonal cohesion) are linked to learning (Jurkowski and Hänze, 2015; Theobald *et al.*, 2017; Curşeu *et al.*, 2018; Monson, 2019; Premo *et al.*, 2021; Cavagnetto *et al.*, 2022).

*Biology.* Six group qualities predicted student achievement in biology (see Table 7). Specifically, if students said that they expected quality group work, this predicted higher achievement. Yet if students said that they interacted more with their groups (e.g., encouragement, listening to feedback, being polite), this predicted lower achievement. The finding that self-reported interactions predicted decreased achievement contrasts with the finding that group members' ratings of a student's contributions predicted the opposite (i.e., increased achievement). This suggests a core difference between the cognitive functions of just interacting with a group versus contributing to the group in terms of learning. In addition, group members' desire to work with a student predicted higher achievement. This shows the importance of having supportive group members—increased ability of a student to learn in a group. Finally, greater presence of task conflict in their group predicted increased achievement, while more relationship conflict predicted decreased achievement. Of these predictors of achievement in biology, task conflict (B = 0.39) and group members' ratings of a student's contributions (B = 0.37) predicted the largest increases in achievement. Relationship conflict in the group predicted the largest decrease in achievement (B = –0.39).

*Chemistry.* Five group qualities and one demographic attribute predicted student achievement in chemistry (Table 8). As in biology, if students said that they interacted more with their groups (e.g., encouragement, listening to feedback, being polite), this predicted lower achievement. In terms of group members' ratings, the more group members said a student contributed and the more group members wanted to work with the student, the higher the student's grade. In addition to self-reported interactions predicting lower student achievement in chemistry, if group members rated a student's interactions as higher, this predicted additional decrease in the student's final grade. The contrast between contribution as a positive predictor of achievement and more general interactions (e.g., encouragement, listening to feedback, being polite) as a negative predictor can be seen as a theme across both biology and chemistry. A predictor that was present in chemistry but not biology included when a student was from an underserved population. Students who were non-Caucasian received lower grades in chemistry. Finally, group process conflict predicted positive increases in one's grade, which was a form of conflict that was not predictive in biology.

**TABLE 7. Final regression model predicting student achievement in biology[a]**

| Factor[b] | B | SE |
|---|---|---|
| Intercept | 0.55 | 0.57 |
| Student is a woman | 0.04 | 0.06 |
| Student is from an underserved population | 0.03 | 0.10 |
| Student's initial interactions with group* | –0.27 | 0.09 |
| Student's expectations of quality work* | 0.25 | 0.10 |
| Initial task conflict in group* | 0.39 | 0.10 |
| Overall relationship conflict in group* | –0.39 | 0.13 |
| Student's overall contributions* | 0.37 | 0.10 |
| Overall group member willingness to work with a student in the future* | 0.29 | 0.09 |

[a]$n = 116$; Adjusted $R^2 = 0.41$.
[b]Initial = beginning of semester; overall = end of semester.
*$p < 0.05$.

*Overall.* Results in both biology (Table 7) and chemistry (Table 8) only partially align with our hypotheses. Group qualities were able to predict a much higher percentage of a student's grade in biology (41% of a student's final grade point average in the course) than in chemistry (15%). This suggests that dynamics of group learning were less impactful on student

**TABLE 8. Final regression model predicting student achievement in chemistry[a]**

| Factor[b] | B | SE |
|---|---|---|
| Intercept | 1.58 | 0.55 |
| Student is a woman | 0.01 | 0.07 |
| Student is from an undeserved population* | –0.29 | 0.08 |
| Student's initial interactions with group* | –0.19 | 0.08 |
| Overall process conflict in group* | 0.26 | 0.13 |
| Student's overall contributions* | 0.31 | 0.09 |
| Student's overall interactions with group* | –0.25 | 0.10 |
| Overall group member willingness to with student in the future* | 0.23 | 0.10 |

[a]$n = 185$; Adjusted $R^2 = 0.15$.
[b]Initial = beginning of semester; overall = end of semester.
*$p < 0.05$.

achievement in chemistry than in biology (though 15% of a student's achievement can easily determine the student passing or failing a university course). Despite this difference, there were common group qualities that predicted achievement in both courses. These included a students' contributions and their group members' continued willingness to work with them. As will be explored in the *Discussion*, the presence of different forms of conflict increased/decreased achievement. Interestingly, more interactions (e.g., encouragement, listening to feedback, being polite) consistently predict decreased achievement across courses, despite being on-task and relevant to working in a group (see Table 4 for an overview of results).

## DISCUSSION

As strategies that rely on peer-to-peer dynamics (in contrast to lecture) become increasingly adopted in undergraduate science, the efficacy of these approaches will rely on both: 1) the ability of the instructor to promote collaborative engagement between students and 2) the encouragement of dynamics that increase group learning. Thus, we sought to determine what predicts a group member's continued willingness to work with a student before exploring the extent to which group qualities predict student achievement.

### Which Group Dynamics Matter?

#### *Continued Willingness to Work with a Student Depends on Conditional Cooperation and Personal Connection*

*Conditional Cooperation.* Students who are willing to work with one another are foundational to group work in the classroom. Additionally, some experiences, such as interdependence structures and more exposure to group work can result in students who are more willing to work with peers in the future (Premo *et al.*, 2018a; Van Ryzin *et al.*, 2020). Thus, in this study, we hypothesized that students would be more willing to continue to work with their peers in the future if they perceived doing so would improve their learning (conditional cooperation; Keser and van Winden, 2000; Premo *et al.* 2018a, b). Specifically, students would be more willing work with peers in the future who were directly contributing to group work. In this study, there was strong evidence that despite the difference in class content (biology vs. chemistry), students who completed group work in a GI-centered class were more willing to continue to work with their peers when they felt that there was direct contribution to the group.

The importance of a student's contributions is striking in that neither a student's ability (having relevant knowledge, skills, and abilities) nor other group-work dynamics (interacting with the group, keeping the group on track, and expecting quality) significantly predicted group members' future willingness to engage with a student. While these factors have been found to be critical for group effectiveness in other studies (Ohland *et al.*, 2012), students are not using these behaviors to judge who they are willing to work with in the future.

*Personal Connection.* Personal connection was the additional quality group members used to determine continued willingness to work with a student in both biology and chemistry. This suggests that allowing students to self-select groups may have the benefit of generating groups with increased willingness to work together if students have friends in the classroom whom they have worked with in the past. It must be recognized that the students without these social connections would not benefit from grouping in this way. Personal connection may also be increased through greater student sense of belonging. Prior research has found connections between a student's sense of belonging and friendship (Pittman and Richmond, 2008). Additionally, there are a number of instructor qualities that have been found to be associated with classroom belonging, such as instructor warmth and structuring the classroom to support belonging, engagement, and motivation (Dewsbury and Brame, 2019; Freeman *et al.*, 2007; Zumbrunn *et al.*, 2014). An informative area of further study would be examining how a sense of belonging and inclusive classroom environment impact how students work and learn from one another.

Finally, it is important to recognize that continued willingness to work with a student is not just about the level of collaboration within a group. As it will be discussed further, the more a group is willing to work with a student in the future, the higher the student's final course grade (regardless of whether the student was in biology or chemistry). This means that students who contribute more tend to have more collaborative groups (due to higher willingness to work), which ultimately leads to an increase in their final grades.

*Gender Bias.* One critical difference between group members' continued willingness to work with students between biology and chemistry was that chemistry students held a significant bias against future group work with women. Prior research has shown that gender differences do exist in chemistry classrooms, with women having significantly lower self-efficacy and higher test anxiety despite equal achievement in the subject (Sunny *et al.*, 2017). This phenomenon is likely connected to the larger issue of gender bias in chemistry, which can be seen in popular chemistry textbooks (Murray *et al.*, 2021) and among practicing chemists (Miller-Friedmann *et al.*, 2018).

Additionally, women have historically been underrepresented in science (Beede *et al.*, 2011), with this trend being especially alarming in the state where this study occurred (only 23.5% of science jobs are held by women; Hess *et al.*, 2015), indicating that there may be a cultural component to the gender bias seen in chemistry. While this does not explain why there is no significant gender bias in biology, it may be due to the increased number of men in the chemistry course. Previous studies have found that gendered stereotypes (science is for men) negatively impact women's science identity (Starr, 2018). Despite women having significantly higher achievement in high school STEM classes, they are significantly less likely to pursue a STEM major in college (Jiang *et al.*, 2020).

Thus, promoting student success relies on the ability of science environments to support the development of student science identity and sense of belonging in science. Identifying more with science increases student science motivation and desire to pursue science-related careers (Estrada *et al.*, 2011). Investigating how one should holistically define, measure, and support science identity and sense of belonging for students, especially women, is critical. Further work is needed to identify how gender bias is manifesting in chemistry classrooms and which interventions would be most appropriate for reducing the impacts of this on women in chemistry classrooms.

### Dynamics That Predict Achievement: Contributions, Conflict, and Support

*"Contributing to" versus "Interacting with" a Group.* As was mentioned earlier, contribution was an important predictor of student achievement for both classrooms. Yet surprisingly, interaction-based behaviors in groups had the opposite impact and predicted lower achievement. This may seem counterintuitive. However, a core difference between contribution and interactions, as defined by the CATME instrument items, is that contributions require active cognitive processing of ideas, while interactions do not. CATME defines interaction as involving more surface-level engagement with the group through "showing interest," "being encouraging," and "asking for feedback"; none of which require active cognitive processing of ideas. For example, a student can encourage peers to generate ideas without having to think deeply about those same ideas. The contrast between noncontent interactions and content contributions is aligned with research highlighting that greater learning results from more active cognitive processing of ideas. Students learn more when interactions include high discussion quality (Curşeu *et al.*, 2018), productive discourse around contrasting opinions (Sampson and Clark, 2011; Howe, 2014; Premo *et al.*, 2021), and the justification of one's ideas (Premo *et al.*, 2021).

The contrast between contributions and interactions provides insights into how to structure group work that promotes contributions, while being mindful of inclusive interactions. For example, if a student shows interest in another student's ideas and is encouraging, the student being encouraged is likely to find the group environment more welcoming and inclusive. Yet the results of this study show that these noncontributing interactions during learning activities themselves (i.e., GI in this case) detract from achievement. We would argue that interactive behaviors during group work are not detrimental in and of themselves but may be functioning to decrease the amount of time a student has to actively process ideas and contribute. Students always have a finite amount of time to learn from another during group work. Thus, to maximize learning during these periods, students should be primarily focused on contribution and related content conversations.

*Group Conflict.* Conflict in student groups has been conceptualized as having a negative influence on student learning (Buchs *et al.*, 2004; Sohr *et al.*, 2018), but prior research has shown that some form of conflict can also be useful in supporting group performance (Jehn and Mannix, 2001; Jehn and Shah, 1996). In direct alignment with the results of this study, Curşeu *et al.* (2012) found that task conflict can be useful in promoting group performance, while relationship conflict has the opposite effect. Thus, conflict in groups is not a unidimensional construct. Instead, conflict involving ideas (task conflict in this study) and the processes by which a team engages with ideas (process conflict in this study) both appear to have the potential to promote learning. This likely arises from the potential for "cognitive conflict" (Piaget, 1985) around ideas that are likely to arise in these scenarios.

Cognitive conflict occurs when a student's current cognitive structures are disrupted based on the presence of discrepant ideas from team members—a factor likely involved in both forms of conflict. This difference between cognitive conflict of ideas versus conflict in interpersonal relationships was also recently noted by (Butera *et al.*, 2019) and emphasizes the positive impact of some form of conflict on cognition. We hypothesize that task and process conflict may be promoting learning through changing the types of interactions occurring between students, so they are more likely to involve higher cognitive processing (just like the contributing behaviors mentioned earlier). Both contrasting opinions (Howe, 2014; Premo *et al.*, 2021; Sampson and Clark, 2011) and having to justify one's ideas (Premo *et al.*, 2021) have been shown to promote learning, and future work should explore how best to promote these in undergraduate science.

*Peer Support.* Knowing the factors that promote student–student engagement is critical, because many active-learning strategies rely on students being willing to actively work with peers. Results from the current study highlight this importance. Having group members who were continually willing to work with a student was a positive predictor of achievement regardless of class. Being part of a group of students who are willing to continually engage with one another is not merely a convenience but a core necessity for student achievement in courses that use group work.

*Underserved Students.* In this study, underserved students (non-Caucasian) received significantly lower grades in chemistry than their Caucasian peers. This was not seen in the biology class. Similar to women, underserved students are minoritized in chemistry with nonequitable representation in the classroom and in chemistry-related careers (Menon, 2021). Previous studies have demonstrated that a decreased sense of belonging is predictive of lower achievement for non-Caucasian students (Fink *et al.*, 2020; Edwards *et al.*, 2021). As such, underserved students may not feel a sense of belonging in their groups and/or classrooms, and this might impact their contribution to a group. In particular, underserved students may be battling stereotype threat (Spencer *et al.*, 2016) and may engage in code switching (Morton, 2014) in these situations. As mentioned earlier, this study demonstrates that active contribution to a group predicts achievement. If students are allocating cognitive effort to avoid stereotypes and implement code switching in their courses, that energy is not being spent on cognitive processing (Johnson *et al.*, 2021; Taylor and Walton, 2011). While more research is needed to investigate this further, inclusive pedagogy within the classroom can help promote the success of all students.

## Accounting for Differences between Biology and Chemistry

*Majors versus Nonmajors.* Several factors predicting group members' willingness to work with a student in the future and student achievement were common between the biology (nonmajors) and chemistry (majors) courses. This suggests that these same factors may be important group dynamics in other areas of science instruction. Yet several factors were unique to just one course, and there was significantly more growth in group members' ratings of their group members' abilities in biology over chemistry. Both warrant a discussion of the extent to which the nature of the courses (nonmajors vs. majors) and the students in them, may be contributing to this difference.

Prior research has shown science major students have some differences from non-science majors. For example, science major students have significantly higher science motivation (Glynn *et al.*, 2011), content knowledge (Partin *et al.*, 2013), and science self-efficacy (Johnstone *et al.*, 2001). For non-science majors, motivation to learn science tends to be connected to how relevant science is to their own (non-science) career trajectories (Glynn *et al.*, 2007), which is more explicitly present for science majors. Yet differences between these groups are not always clear. For example, it has been found that non-science majors can have higher attitudes toward science than science majors (Partin *et al.*, 2013) and that both groups can hold similar views about the nature of science (Miller *et al.*, 2010).

Based on these findings, one might hypothesize that science majors (i.e., those in the chemistry course) would be more motivated to engage in and develop group-work skills due to having higher science motivation. As this study did not collect information on science motivation, we cannot conclude science majors in this study had a higher science motivation than the nonmajors. However, the finding that students in the biology course (nonmajors) grew in their group-work skills two to five times more than students in chemistry (majors) contrasts this hypothesis. Based on the increased growth, one could predict that the biology students would be more motivated in class if motivation was indeed the reason for these differences. Thus, several alternative factors may account for the differences in both group-work skill development and model differences between biology and chemistry.

*Alternative Factors between Classes.* Results of this study highlight the importance of students being able to actively contribute to groups during learning. One method for encouraging student contribution is through increased interdependency between students (i.e., conditions under which students can only succeed through working with others). Higher levels of student interdependence are related to increased student learning (Tomcho and Foels, 2012), more interaction (Johnson and Johnson, 2002), and increased engagement in justifying one's ideas (Premo *et al.*, 2021). This can be done through a variety of methods, such as 1) group grades, 2) structured group activities, and 3) obtaining skills and/or knowledge from another person. While this study is limited in its ability to draw conclusions about the overall differences in interdependency between courses, there are key differences between the structure of the biology course and the structure of the chemistry course. In the biology course, random collections of one group assignment (i.e., a student cannot choose the best group member's work to hand in) and a group final examination were incorporated to encourage group members to work together. The high-value positions of group work as part of their final exam in the biology course could be increasing students' commitment to group work during the semester. These key differences might give some insights into why students in biology grew significantly more in their group-work qualities when compared with students in chemistry.

Finally, other factors that may be differentiating the classes include the difference in gender ratios and diversity between classes (Table 1). Both biases against women and underserved students were only seen in the chemistry course, which suggests that these dynamics may have had larger impact on groups, which could account for some of the differences between the courses.

### Recommendations for Instructors

Given the key results that highlight the importance of meaningful contributions in groups, instructors should structure group work that will promote intentional engagement within groups. We suggest that instructors go through the following steps to help promote greater collaboration in groups. First, reflect on whether students will benefit from working in a group for a given activity. Some activities are more appropriate for individuals to complete. Organizing students in groups for activities that are not appropriate for a group may place students in positions where they are unlikely to meaningfully contribute. This can ultimately detract from future willingness to work in a group if students feel like they do not need to be in a group to complete a task but are put into groups regardless. Second, reflect on what group-work structures are needed to help ensure all students can contribute. Diversity in student preparation and prior knowledge means the ability to contribute will vary. Strategies like different resource allocation (readings, figures, etc.) among group members (Buchs *et al.*, 2004; Buchs and Butera, 2009; Bertucci *et al.*, 2011) or assigning different roles to group members (De Wever *et al.*, 2009; Shimazoe and Aldrich, 2010) have the potential to increase the likelihood that a student will be able to contribute to the group as a whole. We suspect that not all types of resource allocation/roles will be equally effective in promoting contribution, and reflection about this strategy is important. For example, a common group role is a "recorder" who writes down the group's ideas. This role may be less likely to contribute to discussion than other roles due to time taken to write down ideas instead of discussing them.

Finally, students may not contribute to a group if they feel like they do not belong in a group (Eddy *et al.*, 2015). Strategies that may improve the sense of belonging for students in the group include: 1) icebreakers for groups (Chlup and Collins, 2010), 2) metacognitive reflections on group member contributions (Kwon *et al.*, 2013; Smith and Mancy, 2018), and 3) emphasis on inclusive learning environments (Gale *et al.*, 2017; Dewsbury and Brame, 2019). Use of these tactics is likely to reduce anxiety, increase a sense of belonging, and promote more collaboration among students in their classrooms and groups.

As mentioned previously, to maximize learning, students should be primarily focused on contribution and related content conversations. However, this generates a conundrum, because interactions and behaviors that support inclusivity are also critical for a positive social environment and can promote later contributions from group members (see above). We suggest a model of group work wherein different types of group discussions are localized to certain time points in class. Some activities are more appropriate for groups to have interacting-like behaviors to connect with each other at the start of the semester or as a check-in at midsemester. Alternatively, other group activities should be focused on contribution-like behaviors, which are critical for students to understand concepts. Instructors should identify what type of group work is needed, when, and how to best support students in their group behavior. This will help to maximize the potential for contributions of students, and thus cognitive processing, while still generating spaces for the building of inclusive environments in class.

## CONCLUSION

Results of the current study reveal that, while many factors have been found to be critical for group work, only a subset of group qualities appear to be directly predictive of achievement in undergraduate science. Some group qualities that have be assumed to be connected to willingness to work with a student and achievement were not. For example, both group cohesiveness and commitment to the group's task are neither positively nor negatively connected to achievement. In terms of group behavior, a student's display of relevant knowledge, skills, and abilities does not predict either group members' continued willingness to work with the student or the student's own course achievement. These findings suggest that instructors may be able to promote more optimal group learning through preferentially encouraging and designing curricular structures that allow for equitable contributions in teams rather than attempting to manage a greater variety of group qualities. By focusing on promoting group qualities that were predictive of both group members' willingness to work with peers in the future and achievement, we can work to optimize how groups are learning in our classrooms.

### Limitations and Future Work

The goals of the current study were to identify which group qualities are critical to both student engagement with, and learning from, group work in undergraduate science. Yet the conclusions that can be drawn are limited in a variety of respects. First, peer ratings of behaviors are only a proxy of the actual behaviors that could be observed through audio/video means and thus are more prone to error and bias. Second, the data sources are only from two undergraduate courses collected over multiple semesters, which limits their generalizability to a variety of other undergraduate science contexts. Despite this, we hypothesize that "in common" factors found to predict achievement in both courses (i.e., contributions, personal connection, interactions, willingness to work, and conflict) in the current student are likely at play in a variety of other science classrooms and thus should bear further scrutiny in understanding how the social environment promotes or detracts from undergraduate science learning. Future work should focus on the intersection between inclusivity and achievement in promoting group dynamics in the classroom. Results of this study suggest that both inclusivity (which promotes content contributions) and cognitive processing of ideas (which promotes achievement) are critical and intertwined. Thus, promoting both in a mutually supporting way is a goal that can elevate all students in undergraduate science education.

## ACKNOWLEDGMENTS

This work was funded by a Utah Valley University High Impact Engaged Learning Grant from the Office of Engaged Learning.

## REFERENCES

American Association for the Advancement of Science. (2019). *Levers for change: An assessment of progress on changing STEM*. Retrieved March 1, 2021, from www.aaas.org/resources/levers-change-assessment-progress-changing-stem-instruction

Andrews, J. J., & Rapp, D. N. (2015). Benefits, costs, and challenges of collaboration for learning and memory. *Translational Issues in Psychological Science*, *1*(2), 182–191. https://doi.org/10.1037/tps0000025

Bailey, C. P., Minderhout, V., & Loertscher, J. (2012). Learning transferable skills in large lecture halls: Implementing a POGIL approach in biochemistry. *Biochemistry and Molecular Biology Education*, *40*(1), 1–7. https://doi.org/10.1002/bmb.20556

Balasubramaniam, G. (2015). Process oriented guided inquiry learning as an innovative teaching method in clinical biochemistry: Piloting of the process. *Education in Medicine Journal*, *7*(2), e60–e63. https://doi.org/10.5959/eimj.v7i2.337

Balliet, D., Tybur, J. M., & Van Lange, P. A. M. (2017). Functional interdependence theory: An evolutionary account of social situations. *Personality and Social Psychology Review*, *21*(4), 361–388. https://doi.org/10.1177/1088868316657965

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011). *Women in STEM: A* gender gap to innovation (ESA Issue Brief #04-11). . Washington, DC: Economics and Statistics Administration, U.S. Department of Commerce. https://doi.org/10.2139/ssrn.1964782

Bertucci, A., Johnson, D. W., Johnson, R. T., & Conte, S. (2011). The effects of task and resource interdependence on achievement and social support: An exploratory study of Italian children. *Journal of Psychology*, *145*(4), 343–360. https://doi.org/10.1080/00223980.2011.574167

Blumen, H. M., Young, K. E., & Rajaram, S. (2014). Optimizing group collaboration to improve later retention. *Journal of Applied Research in Memory and Cognition*, *3*(4), 244–251. https://doi.org/10.1016/j.jarmac.2014.05.002

Bokosmaty, R., Bridgeman, A., & Muir, M. (2019). Using a partially flipped learning model to teach first year undergraduate chemistry. *Journal of Chemical Education*, *96*(4), 629–639. https://doi.org/10.1021/acs.jchemed.8b00414

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New york, NY: Guilford.

Buchs, C., & Butera, F. (2009). Is a partner's competence threatening during dyadic cooperative work? It depends on resource interdependence. *European Journal of Psychology of Education*, *24*(2), 145–154. https://doi.org/10.1007/BF03173007

Buchs, C., Butera, F., & Mugny, G. (2004). Resource interdependence, student interactions and performance in cooperative learning. *Educational Psychology*, *24*(3), 291–314. https://doi.org/10.1080/0144341042000211661

Buchs, C., Pulfrey, C., Gabarrot, F., & Butera, F. (2010). Competitive conflict regulation and informational dependence in peer learning. *European Journal of Social Psychology*, *40*(3), 418–435. https://doi.org/10.1002/ejsp.631

Burnham, K. P., & Anderson, D. R. (1998). Practical use of the information-theoretic approach. In Burnham, K. P., & Anderson, D. R. *Model selection and inference: A practical information-theoretic approach* (pp. 75–117). New York, NY: Springer. https://doi.org/10.1007/978-1-4757-2917-7_3

Butera, F., Sommet, N., & Darnon, C. (2019). Sociocognitive conflict regulation: How to make sense of diverging ideas. *Current Directions in Psychological Science*, *28*(2), 145–151. https://doi.org/10.1177/0963721418813986

Carless, S. A., & De Paola, C. (2000). The measurement of cohesion in work teams. *Small Group Research*, *31*(1), 71–88. https://doi.org/10.1177/104649640003100104

Cavagnetto, A. R., Premo, J., Coleman, Z., & Juergens, K. (2022). Accuracy and idea consideration: A study of small-group interaction in biology. *CBE—Life Sciences Education*, *21*(1), ar5. https://doi.org/10.1187/cbe.21-03-0067

Chang, Y., & Brickman, P. (2018). When group work doesn't work: Insights from students. *CBE—Life Sciences Education*, *17*(3), ar52. https://doi.org/10.1187/cbe.17-09-0199

Chinn, C. A., & Malhotra, B. A. (2002). Epistemology authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 171–174.

Chlup, D. T., & Collins, T. E. (2010). Breaking the ice: Using ice-breakers and re-energizers with adult learners. *Adult Learning*, *21*(3–4), 34–39. https://doi.org/10.1177/104515951002100305

Curşeu, P. L., Boroş, S., & Oerlemans, L. A. (2012). Task and relationship conflict in short-term and long-term groups: The critical role of emotion regulation. *International Journal of Conflict Management*, *23*(1), 97–107. https://doi.org/10.1108/10444061211199331

Curşeu, P. L., Chappin, M. M. H., & Jansen, R. J. G. (2018). Gender diversity and motivation in collaborative learning groups: The mediating role of group discussion quality. *Social Psychology of Education*, *21*(2), 289–302. https://doi.org/10.1007/s11218-017-9419-5

Darnon, C., Buchs, C., & Butera, F. (2002). Epistemic and relational conflicts in sharing identical vs complementary information during cooperative learning. *Swiss Journal of Psychology*, *61*(3), 139–151. https://doi.org/10.1024/1421-0185.61.3.139

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2009). Structuring asynchronous discussion groups: The impact of role assignment and self-assessment on students' levels of knowledge construction through social negotiation. *Journal of Computer Assisted Learning*, *25*(2), 177–188. https://doi.org/10.1111/j.1365-2729.2008.00292.x

Dewsbury, B., & Brame, C. J. (2019). Inclusive teaching. *CBE—Life Sciences Education*, *18*(2), fe2. https://doi.org/10.1187/cbe.19-01-0021

Diamantopoulos, A., Siguaw, J. A., & Siguaw, J. A. (2000). *Introducing LISREL: A guide for the uninitiated*. London, UK: Sage.

Eagan, K. (2016). Becoming more student-centered? *An* examination of faculty teaching practices *across STEM and non-STEM* disciplines *between 2004 and 2014* (A report prepared for the Alfred P. Sloan Foundation). Los Angeles, CA: Higher Education Research Institute.

Eberlein, T., Kampmeier, J., Minderhout, V., Moog, R. S., Platt, T., Varma-Nelson, P., & White, H. B. (2008). Pedagogies of engagement in science. *Biochemistry and Molecular Biology Education*, *36*(4), 262–273. https://doi.org/10.1002/bmb.20204

Eddy, S. L., Brownell, S. E., Thummaphan, P., Lan, M.-C., & Wenderoth, M. P. (2015). Caution, student experience may vary: Social identities impact a student's experience in peer discussions. *CBE—Life Sciences Education*, *14*(4), ar45. https://doi.org/10.1187/cbe.15-05-0108

Edwards, J. D., Barthelemy, R. S., & Frey, R. F. (2021). Relationship between course-level social belonging (sense of belonging and belonging uncertainty) and academic performance in general chemistry. *Journal of Chemical Education*, *99*(1), 71–82. https://doi.org/10.1021/acs.jchemed.1c00405

Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, P. W. (2011). Toward a model of social influence that explains minority student integration into the scientific community. *Journal of Educational Psychology*, *103*(1), 206–222. https://doi.org/10.1037/a0020743

Farrell, J. J., Moog, R. S., & Spencer, J. N. (1999). A guided-inquiry general chemistry course. *Journal of Chemical Education*, *76*(4), 570–574. https://doi.org/10.1021/ed076p570

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London, UK: Sage.

Fink, A. F., Frey, R., & Solomon, D. E. (2020). Belonging in general chemistry predicts first-year undergraduates' performance and attrition. *Chemistry Education Research and Practice*, *21*(4), 1042–1062. https://doi.org/10.1039/D0RP00053A

Ford, M. (2008). "Grasp of practice" as a reasoning resource for inquiry and nature of science understanding. *Science & Education*, *17*(2), 147–177. https://doi.org/10.1007/s11191-006-9045-7

Freeman, T. M., Anderman, L. H., & Jensen, J. M. (2007). Sense of belonging in college freshmen at the classroom and campus levels. *Journal of Experimental Education*, *75*(3), 203–220. https://doi.org/10.3200/JEXE.75.3.203-220

Gale, T., Mills, C., & Cross, R. (2017). Socially inclusive teaching: Belief, design, action as pedagogic work. *Journal of Teacher Education*, *68*(3), 345–356. https://doi.org/10.1177/0022487116685754

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48*(10), 1159–1176. https://doi.org/10.1002/tea.20442

Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2007). Nonscience majors learning science: A theoretical model of motivation. *Journal of Research in Science Teaching*, *44*(8), 1088–1107. https://doi.org/10.1002/tea.20181

Hair, J. F. (2011). Multivariate data analysis: An overview. In Lovric, M. (Ed.), *International encyclopedia of statistical science* (pp. 904–907). Berlin, Heidelberg, Germany: Springer. https://doi.org/10.1007/978-3-642-04898-2_395

Hancock, G. R., & McNeish, D. M. (2017). More powerful tests of simple interaction contrasts in the two-way factorial design. *Journal of Experimental Education*, *85*(1), 24–35. https://doi.org/10.1080/00220973.2015.1065220

Hein, S. M. (2012). Positive impacts using POGIL in organic chemistry. *Journal of Chemical Education*, *89*(7), 860–864. https://doi.org/10.1021/ed100217v

Hess, C., Milli, J., Hayes, J., & Hegewisch, A. (2015). *The status of women in the states: 2015* (IWPR Report #C400). Washington, DC: Institute for Women's Policy Research. Retrieved December 1, 2020, from http://statusofwomendata.org/wp-content/uploads/2015/09/PDF-of-final-Work-Family-chapter-9-4-2015.pdf

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

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jehn, K. A., & Mannix, E. A. (2001). The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal*, *44*(2), 238–251. https://doi.org/10.2307/3069453

Jehn, K. A., & Shah, P. P. (1996). Interpersonal relationships and task performance: An examination of mediating processes in friendship and acquaintance groups. *Academy of Management Proceedings*, *1996*(1), 279–283. https://doi.org/10.5465/ambpp.1996.4980697

Jiang, S., Simpkins, S. D., & Eccles, J. S. (2020). Individuals' math and science motivation and their subsequent STEM choices and achievement in high school and college: A longitudinal study of gender and college generation status differences. *Developmental Psychology*, *56*(11), 2137–2151. https://doi.org/10.1037/dev0001110

Johnson, D. G., Mattan, B. D., Flores, N., Lauharatanahirun, N., & Falk, E. B. (2021). Social-cognitive and affective antecedents of code switching and the consequences of linguistic racism for Black people and people of color. *Affective Science*, *3*, 5–13. https://doi.org/10.1007/s42761-021-00072-8

Johnson, D. W., & Johnson, R. T. (2002). Cooperative learning and social interdependence theory. In Tindale, R. S., Heath, L., Edwards, J., Posavac, E. J., Bryant, F. B., Suarez-Balcazar, Y., Henderson-King, E., & Myers, J. (Eds.), *Theory and research on small groups* (pp. 9–35). New York, NY: Springer US. https://doi.org/10.1007/0-306-47144-2_2

Johnson, D. W., Johnson, R. T., & Smith, K. A. (2014). Cooperative learning: Improving university instruction by basing practice on validated theory. *Journal on Excellence in College Teaching*, *25*(4), 85–118.

Johnstone, R., Haines, V. A., & Wallace, J. (2001). Do factors that differentiate science and nonscience majors predict majoring in science? *Alberta Journal of Educational Research*, *47*(3), 280–284.

Jordt, H., Eddy, S. L., Brazil, R., Lau, I., Mann, C., Brownell, S. E., … & Freeman, S. (2017). Values affirmation intervention reduces achievement gap between underrepresented minority and white students in introductory biology classes. *CBE—Life Sciences Education*, *16*(3), ar41. https://doi.org/10.1187/cbe.16-12-0351

Jurkowski, S., & Hänze, M. (2015). How to increase the benefits of cooperation: Effects of training in transactive communication on cooperative learning. *British Journal of Educational Psychology*, *85*(3), 357–371. https://doi.org/10.1111/bjep.12077

Keser, C., & van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, *102*(1), 23–39.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior*, *25*(2), 306–314. https://doi.org/10.1016/j.chb.2008.12.008

Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*, *18*(1), rm1. https://doi.org/10.1187/cbe.18-04-0064

Kwon, K., Hong, R.-Y., & Laffey, J. M. (2013). The educational impact of metacognitive group coordination in computer-supported collaborative learning. *Computers in Human Behavior*, *29*(4), 1271–1281. https://doi.org/10.1016/j.chb.2013.01.003

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review*, *10*, 133–149. https://doi.org/10.1016/j.edurev.2013.02.002

Lasry, N., Mazur, E., & Watkins, J. (2008). Peer instruction: From Harvard to the two-year college. *American Journal of Physics*, *76*(11), 1066–1069. https://doi.org/10.1119/1.2978182

Lawson, A. E. (1995). *Science teaching and the development of thinking*. Belmont, CA: Wadsworth.

Lewis, S. E., & Lewis, J. E. (2005). Departing from lectures: An evaluation of a peer-led guided inquiry alternative. *Journal of Chemical Education*, *82*(1), 135–139. https://doi.org/10.1021/ed082p135

Loughry, M. L., & Tosi, H. L. (2008). Performance implications of peer monitoring. *Organization Science*, *19*(6), 876–890.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Maiden, B., & Perry, B. (2011). Dealing with free-riders in assessed group work: Results from a study at a UK university. *Assessment & Evaluation in Higher Education*, *36*(4), 451–464. https://doi.org/10.1080/02602930903429302

Manz, E. (2015). Representing student argumentation as functionally emergent from scientific activity. *Review of Educational Research*, *85*(4), 553–590. https://doi.org/10.3102/0034654314558490

Marion, S. B., & Thorley, C. (2016). A meta-analytic review of collaborative inhibition and post collaborative memory: Testing the predictions of the retrieval strategy disruption hypothesis. *Psychological Bulletin*, *142*(11), 1141–1164. https://doi.org/10.1037/bul0000071

Menon, B. R. K. (2021). The missing colours of chemistry. *Nature Chemistry*, *13*(2), 101–106. https://doi.org/10.1038/s41557-020-00632-8

Miller, M. C. D., Montplaisir, L. M., Offerdahl, E. G., Cheng, F.-C., & Ketterling, G. L. (2010). Comparison of views of the nature of science between natural science and nonscience majors. *CBE—Life Sciences Education*, *9*(1), 45–54. https://doi.org/10.1187/cbe.09-05-0029

Miller-Friedmann, J., Childs, A., & Hillier, J. (2018). Approaching gender equity in academic chemistry: Lessons learned from successful female chemists in the UK. *Chemistry Education Research and Practice*, *19*(1), 24–41. https://doi.org/10.1039/C6RP00252H

Monson, R. A. (2019). Do they have to like it to learn from it? Students' experiences, group dynamics, and learning outcomes in group research projects. *Teaching Sociology*, *47*(2), 116–134. https://doi.org/10.1177/0092055X18812549

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.

Morton, J. M. (2014). Cultural code-switching: Straddling the achievement gap. *Journal of Political Philosophy*, *22*(3), 259–281. https://doi.org/10.1111/jopp.12019

Murray, C., Anderson, Y., Simms, C. H., & Seery, M. K. (2021). Representations of women and men in popular chemistry textbooks in the United Kingdom and Republic of Ireland. *Chemistry Education Research and Practice*, *23*, 373–384. https://doi.org/10.1039/D1RP00187F

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*(4), 645–656. https://doi.org/10.1007/s10648-015-9312-8

Ohland, M., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R., Finelli, C., … & Schmucker, D. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self and peer evaluation. *Academy of Management Learning & Education*, *11*(4), 609–630. https://doi.org/10.5465/AMLE.2010.0177

Partin, M. L., Underwood, E. M., & Worch, E. A. (2013). Factors related to college students' understanding of the nature of science: Comparison of science majors and nonscience majors. *Journal of College Science Teaching*, *42*(6), 89–99.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago: University of Chicago Press.

Pieterse, V., & Thompson, L. (2010). Academic alignment to reduce the presence of "social loafers" and "diligent isolates" in student teams. *Teaching in Higher Education*, *15*(4), 355–367. https://doi.org/10.1080/13562517.2010.493346

Pittman, L. D., & Richmond, A. (2008). University belonging, friendship quality, and psychological adjustment during the transition to college. *Journal of Experimental Education*, *76*(4), 343–362. https://doi.org/10.3200/JEXE.76.4.343-362

Premo, J., Cavagnetto, A., Collins, L., Davis, W. B., & Offerdahl, E. (2021). Discourse remixed: Shifting science learning through talk. *Journal of Experimental Education*. https://doi.org/10.1080/00220973.2021.1993771

Premo, J., Cavagnetto, A., & Davis, W. B. (2018a). Promoting collaborative classrooms: The impacts of interdependent cooperative learning on undergraduate interactions and achievement. *CBE—Life Sciences Education*, *17*(2), ar32. https://doi.org/10.1187/cbe.17-08-0176

Premo, J., Cavagnetto, A., & Lamb, R. (2018b). The Cooperative Classroom Environment Measure (CCEM): Refining a measure that assesses factors motivating student prosociality. *International Journal of Science and Mathematics Education*, *16*(4), 677–697. https://doi.org/10.1007/s10763-017-9804-8

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

Salas, E., Reyes, D., & McDaniel, S. (2018). The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist*, *73*(4), 593–600. https://doi.org/10.1037/amp0000334

Sampson, V., & Clark, D. B. (2011). A comparison of the collaborative scientific argumentation practices of two high and two low performing groups. *Research in Science Education*, *41*(1), 63–97. https://doi.org/10.1007/s11165-009-9146-9

Shimazoe, J., & Aldrich, H. (2010). Group work can be gratifying: Understanding & overcoming resistance to cooperative learning. *College Teaching*, *58*(2), 52–57. https://doi.org/10.1080/87567550903418594

Schroeder, J., & Greenbowe, J. T. (2008). Implementing POGIL in the lecture and the Science Writing Heuristic in the laboratory—student perceptions and performance in undergraduate organic chemistry. *Chemistry Education Research and Practice*, *9*(2), 149–156. https://doi.org/10.1039/B806231P

Simonson, S. R., & Shadle, S. (2013). Implemented process oriented guided inquiry learning (POGIL) in undergraduate biomechanics: Lessons learned by a novice. *Journal of STEM Education: Innovations and Research*, *14*(1), 56–63. Retrieved December 15, 2020, from www.jstem.org/jstem/index.php/JSTEM/article/view/1545

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

Sohr, E. R., Gupta, A., & Elby, A. (2018). Taking an escape hatch: Managing tension in group discourse. *Science Education*, *102*(5), 883–916. https://doi.org/10.1002/sce.21448

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, *67*(1), 415–437. https://doi.org/10.1146/annurev-psych-073115-103235

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., … & Young, A. M. (2018). Anatomy of STEM teaching in American universities: A snapshot from a large-scale observation study. *Science*, *359*(6383), 1468–1470. https://doi.org/10.1126/science.aap8892

Starr, C. R. (2018). "I'm not a science nerd!": STEM stereotypes, identity, and motivation among undergraduate women. *Psychology of Women Quarterly*, *42*(4), 489–503. https://doi.org/10.1177/0361684318793848

Summers, M., & Volet, S. (2010). Group work does not necessarily equal collaborative learning: Evidence from observations and self-reports. *European Journal of Psychology of Education*, *25*(4), 473–492. https://doi.org/10.1007/s10212-010-0026-5

Sunny, C. E., Taasoobshirazi, G., Clark, L., & Marchand, G. (2017). Stereotype threat and gender differences in chemistry. *Instructional Science*, *45*(2), 157–175. https://doi.org/10.1007/s11251-016-9395-8

Tanner, K., Chatman, L. S., & Allen, D. (2003). Approaches to cell biology teaching: Cooperative learning in the science classroom—beyond students working in groups. *Cell Biology Education*, *2*(1), 1–5. https://doi.org/10.1187/cbe.03-03-0010

Taylor, V. J., & Walton, G. M. (2011). Stereotype threat undermines academic learning. *Personality and Social Psychology Bulletin*, *37*(8), 1055–1067. https://doi.org/10.1177/0146167211406506

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

Tomcho, T. J., & Foels, R. (2012). Meta-analysis of group learning activities: Empirically based teaching recommendations. *Teaching of Psychology*, *39*(3), 159–169. https://doi.org/10.1177/0098628312450414

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

van Ryzin, M. J., Roseth, C. J., & Biglan, A. (2020). Mediators of effects of cooperative learning on prosocial behavior in middle school. *International Journal of Applied Positive Psychology*, *5*(1–2), 37–52. https://doi.org/10.1007/s41042-020-00026-8

Vegt, G. S. V. D., Emans, B. J. M., Van, E., & Vuert, D. (2001). Patterns of interdependence in work teams: A two-level investigation of the relations with job and team satisfaction. *Personnel Psychology*, *54*, 51–69.

Walker, L., & Warfa, A.-R. M. (2017). Process oriented guided inquiry learning (POGIL®) marginally effects student achievement measures but substantially increases the odds of passing a course. *PLoS ONE*, *12*(10), e0186203. https://doi.org/10.1371/journal.pone.0186203

Wang, J. (2016). Knowledge creation in collaboration networks: Effects of tie configuration. *Research Policy*, *45*(1), 68–80. https://doi.org/10.1016/j.respol.2015.09.003

Warfa, A.-R. M., Nyachwaya, J., & Roehrig, G. (2018). The influences of group dialog on individual student understanding of science concepts. *International Journal of STEM Education*, *5*(46), 1–14. https://doi.org/10.1186/s40594-018-0142-3

Zhang, H., Fu, Y., Zhang, X., & Shi, J. (2017). The effect of item similarity and response competition manipulations on collaborative inhibition in group recall. *Scientific Reports*, *7*(1), 11946. https://doi.org/10.1038/s41598-017-12177-x

Zumbrunn, S., McKim, C., Buhs, E., & Hawley, L. R. (2014). Support, belonging, motivation, and engagement in the college classroom: A mixed method study. *Instructional Science*, *42*(5), 661–684. https://doi.org/10.1007/s11251-014-9310-0